

RESEARCH ARTICLE

Comparison of Cox Regression and Parametric Models: Application for Assessment of Survival of Pediatric Cases of Acute Leukemia in Southern Iran

Saeed Hosseini Teshnizi¹, Seyyed Mohammad Taghi Ayatollahi^{2*}

Abstract

Background: Finding the most appropriate regression model for survival data in cancer cases in order to determine prognosis is an important issue in medical research. Here we compare Cox and parametric regression models regarding survival of children with acute leukemia in southern Iran. **Methods:** In a retrospective cohort study, information for 197 children with acute leukemia over 6 years was collected through observation and interviews. In order to identify factors affecting their survival, the Cox and parametric (exponential, Weibull, log-logistic, log-normal, Gompertz and generalized gamma) models were fitted to the data. To find the best predictor model, the Akaike's information criterion (AIC) and the Coxsnell residual were employed. **Results:** Out of 197 children, 164 (83.3%) had ALL and 33 (16.7%) AML; the mean (\pm standard deviation) survival time was 52.1 \pm 8.10 months. According to both the AIC and the Coxsnell residual, the Cox regression model was the weakest and the log-normal and Weibull models were the best for fitting to data. Based on the log-normal model, age (HR=1.01, p=0.004), residence area (HR=1.60, p=0.038) and WBC (White Blood Cell) (HR=1.57, p=0.014) had significant effects on patient survival. **Conclusion:** Parametric regression models demonstrate better performance as compared to the Cox model for identifying risk factors for prognosis with acute leukemia data. Just because the assumption of PH (Proportional Hazards) is held for the Cox regression model, we should not ignore parameter models.

Keyword: Cox regression- parametric models- acute leukemia- pediatric cases

Asian Pac J Cancer Prev, **18** (4), 981-985

Introduction

One of the most prevalent cancers among children is leukemia (Coustan-Smith et al., 2003). It's the second cause of death in the first 15 years of life and contains more than 30 percent of cancers in children (Almasi-Hashiani, Zareifar, Karimi, Khedmati, and Mohammadbeigi, 2013; Jemal et al., 2008; Lau et al., 2004; Leukemia, 2008). Some reports indicate an increase in the prevalence of this disease in Iran during recent years (Neyestani et al., 2007). Acute leukemia can be further subdivided in acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). In children, approximately 80% of cases are ALL, and 15-20% AML (Zwaan et al., 2015) as observed in most studies around the world, but not in Africa (Rego, Pinheiro, Metze, & Lorand-Metze, 2003). Risk factors such as age at diagnosis, smoking, bone marrow transplants, disease history, lodging and resistance to treatment, and laboratory factors as the number of white blood cells (WBC), hemoglobin, sodium, platelet, potassium and calcium are effect on survival time of acute leukemia (Kurnaz et al., 2016).

Although the survival of children with acute leukemia

has increased in many developed countries, about 20% of these patients have a survival rate of less than 5 years (Lightfoot et al., 2012). Today, due to the increasing use of survival analysis in medical research, the need for efficient and flexible model for survival data is increasingly felt. In most medical research which had been designed to determine the survival of cancer patients, Cox regression was used because it's more flexible and there is no need to estimate the baseline hazard function. Although in Cox regression the distribution of observed survival time is not clear, proportional hazards (PH) assumption should be investigated (Ravangard et al., 2011). Since the violation of this assumption jeopardizes the validity of the results of the Cox regression, in this case it would be better to use the parametric models such as Exponential, Weibull, Log-Logistic, Log-Normal, Gompertz and Gamma than the Cox model (Almasi-Hashiani et al., 2013; Lau et al., 2004; Neyestani et al., 2007). Compared to Cox model, parametric models have both more validity and higher accuracy in parameter estimates and also do not require checking the assumption of proportional hazards. Parametric models, in contrast, require the distribution of observed survival time, and reducing the sample size

¹Clinical Research Development Center of Children Hospital, Hormozgan University of Medical Sciences, Bandar Abbas,

²Department of Biostatistics, Shiraz University of Medical Sciences, Shiraz, Iran. *For Correspondence: ayatolahim@sums.ac.ir

considerably affects the relative performance of the estimates (Zare et al., 2013). In the proportional hazard model, modeling is done based on hazard function. In this case, if baseline hazard is considered parametric, one of the exponential, Weibull or Gompertz models will be obtained. If the baseline hazard is considered indefinite, the semi-parametric Cox model will be achieved. In the accelerated failure-time model, modeling is done on time logarithm to the occurrence of next state. The obtained models in this case include: Exponential, Weibull, Lognormal, Log-logistic, Gompertz and Generalized gamma (Hougaard, 1999).

Although several studies on the comparison of Cox model with parametric models have been performed, few studies have compared the semi-parametric Cox model with all parametric models. In order to identify influential factors of survival in pediatric cases with acute leukemia, we used both Cox proportional hazard model and parametric models and compared the result.

In general, the aim of this study was to compare the parametric methods and Cox's regression model to determine the independent factors in the survival of children with acute leukemia among the children's hospitals in Bandar Abbas city, southern Iran.

Materials and Methods

Patients and Methods

In a retrospective cohort study during March 25, 2010 and March 20, 2015, totally 218 records with leukemia (ALL and AML) were extracted which 197 patients (90.3%) of them had eligible to enter this study. These patients with acute leukemia (All and AML) who referred to Bandar Abbas city children's hospital for treatment of leukemia. Also, demographic information including age at diagnosis, sex, blood group, laboratory factors such as platelets, hemoglobin, hematocrit, white blood cell (WBC) count at the time of diagnosis, the date of disease diagnosis, hospitalization period (duration time which patient hospitalized in hospital), and the final status (alive or dead) were collected considering the ethical issues. For patients whose status (dead or alive) was not known, the necessary information was obtained through phone call and if their final status was not determined, they were excluded. Finally, the sample size was reduced to 197. In order to compare the parametric and semi-parametric models, Akaike information criterion (AIC) was applied. AIC is a measure of the goodness of fit and each model have a Smaller value, it indicates that the proposed model fits the data well. The following formula was used to calculate AIC:

$$AIC = -2 \text{ Log likelihood} + 2p, \quad (1)$$

In the above formula, p is the number of parameters (D. Kleinbaum & M. Klein, 2005). In additional AIC criteria method, a graphical method which use to evaluate for overall goodness of fit after fitting models to survival data. When the survival function has been estimated, the Cox-Snell residuals defined as:

$$r_{Ci} = H(t_i) = -\log(S(t_i)), \quad (2)$$

which $H(t_i)$ and $S(t_i)$ are empirical cumulative hazard function and survival empirical hazard function,

respectively. A perfectly fitting model would generate a plot that lies directly on top of the diagonal line (Ansin, 2015). For quantitative variables, the mean \pm standard deviation (SD) and for qualitative variables numbers and percentages in each category were reported. Proportional hazards test, semi-parametric Cox regression and parametric models (exponential, Weibull, Log-Logistic, Log-Normal, Gompertz, Generalized Gamma) were performed. For assessing the PH assumption, we calculated the correlation between ranking of individual failure times and the Schoenfeld residuals. If the PH assumption is met, then the correlation should be near zero (D. G. Kleinbaum & M. Klein, 2005). For assessing effect of each variable, Hazard Ratio (HR) was estimated for each model.

All statistical analyses were performed using Stata (Intercooled, version 11, Stata Corp, College Station, TX) and R 3.0.1. p-value less than 0.05 was considered as statistically significant.

Results

Out of 197 patients, 104 (47%) died due to leukemia and 73 (53%) were censored (right censored). 121 (61.4%) were boys; 111 (56.3%) lived in rural and 86 (43.7%) in the cities. The mean \pm SD age of the children was 5.95 ± 3.81 (ranged 1-16) years. 109 (55.4%) patients were urban and the rest rural. Blood group for 55 (27.9%) patients was A, 45 (22.8%), 89 (45.2%) and 8 (4.1%) were B, AB and O, respectively. 164 of them (83.3%) were ALL (52% L1, 45.6% L2 and 2.4% L3) and 33 (16.7%) were AML (25% M0, 12.5% M1, 25% M2, 12.5% M6 and 25% M7) (Table 1).

Before using the Cox regression model, PH assumption for each of the variables were investigated. The result of Schoenfeld test showed that none of the variables violated the PH assumption ($p > 0.05$). So we could use the Cox proportional hazard model for this study. AIC criterion

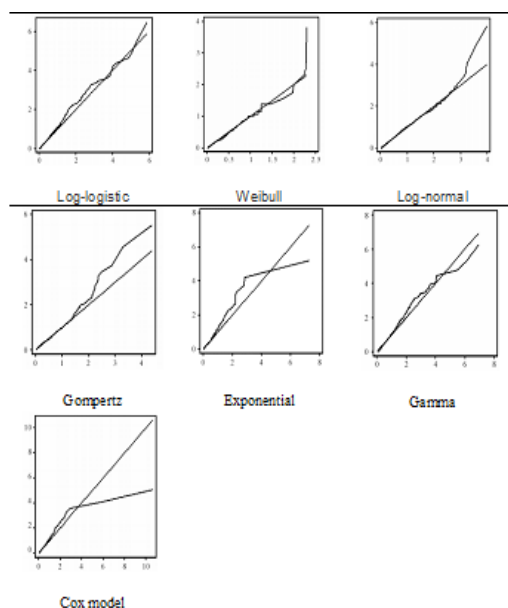


Figure 1. The Cox-Snell Residual Plots for Assessing Goodness of Fit Parametric and Cox Models

Table 1. Descriptive Statistics of Diagnostic Factors of Acute Leukemia Patients

Characteristics	Mean±SD	Min	Max
Age (year)	5.95 ± 3.81	1	16.00
WBC	13595.83±6234.52	1124.5	90,000.00
Hemoglobin	17.45±9.27	3.12	29.00
Platelets	67387.17±135931.53	---	---
Survival time (week)	52.12±8.10	0.3	156.12
hospitalization period (day)	5.50±3.74	1	25.00

SD, Standard deviation

Table 2. The Comparison of the AIC between the Cox Proportional Hazard Model and Parametric Mode

Model	Number of parameters	Log-likelihood	AIC
Cox	11	-412.22	846.44
Exponential	13	-277.55	581.1
Weibull	13	-264.27	554.54
Gompers	13	-274.87	575.74
Log-normal	13	-264.19	554.38
Log-logistic	13	-265.61	557.22
Gamma	14	-264.03	556.06

for Cox model was the highest (846.44) and among parametric models, log-normal (554.38) and Weibull (554.54) with slight differences, respectively, had the lowest AIC. On the other hand, based on plots of Coxsnell residual, in graph for parametric models spatially log-

normal and Weibull is more near to 45-degree diagonal line than cox model. So, infer that parametric models spatially log-normal and Weibull had better fit to this data (Table 2 and Figur1).

Based on both models, log-normal and Weibull, the hazard rate of death due to leukemia when a one year increase in age increases by 1% age (HR=1.01, p=0.004), hazard for patients which resided in village was 60% more than patients residing in cities (HR=1.60, p=0.038) and the hazard rate of death for patient with WBC>5000 was 57% more than the patients with WBC≤5000 (HR=1.57, p=0.014). Based on the results of Cox model, age (HR=1.02, p=0.015), residence area (HR=2.67, p=0.025), WBC (HR=1.65, p=0.013) and type of leukemia (HR=1.68, p=0.039) had a significant effect on survival of patients (Table3).

Table 3. Diagnostic Factors of Acute Leukemia Using Cox and Parametric Multiple Regression Models

Characteristics	Cox	Exponential	Weibull	Log-Logistic	Log-Normal	Gompertz	Gamma
Constant	-	0.01(0.007)	0.05(0.03)	0.05(0.03)	0.05(0.02)	0.02(0.009)	0.01(0.007)
Age	1.02(0.016)*	1.01(0.005)*	1.01(0.005)*	1.01(0.005)*	1.01(0.005)*	1.01(0.006)*	1.01(0.009)*
Sex							
Boy	1.06(0.32)	1.11(0.25)	1.11(0.25)	1.11(0.25)	1.11(0.25)	1.08(0.24)	1.18(0.25)
Girl	-	-	-	-	-	-	-
Residence area							
Village	2.67(0.42)*	1.74(0.39)*	1.60(0.36)*	1.60(0.36)*	1.60(0.37)*	1.64(0.37)*	2.74(0.39)*
City	-	-	-	-	-	-	-
Blood group							
A	0.46(0.47)	0.48(0.24)	0.51(0.25)	0.51(0.25)	0.50(0.22)	0.50(0.26)	0.48(0.24)
B	0.37(0.20)	0.37(0.20)	0.39(0.21)	0.39(0.21)	0.39(0.21)	0.39(0.22)	0.37(0.29)
O	0.62(0.30)	0.64(0.32)	0.65(0.32)	0.65(0.32)	0.66(0.32)	0.65(0.32)	0.64(0.32)
AB	-	-	-	-	-	-	-
Leukemia type							
AML	1.66(0.42)*	1.82(0.46)	1.52(0.37)	1.52(0.37)	1.50(0.36)	1.62(0.41)	1.90(0.36)*
ALL	-	-	-	-	-	-	-
WBC							
≥5000	1.65(0.37)*	1.57 (0.13)*	1.57 (0.13)*	1.57 (0.13)*	1.57 (0.12)*	1.57(0.13)*	1.57 (0.16)*
<5000	-	-	-	-	-	-	-
Hemoglobin	1.01(0.45)	0.68(0.17)	0.68(0.17)	0.68(0.17)	0.69(0.17)	0.69(0.17)	0.68(0.17)
Platelets	0.87(0.75)	1.19(0.30)	1.19(0.30)	1.19(0.30)	1.24(0.32)	1.21(0.30)	1.19(0.30)
Hospitalization	0.97(0.14)	0.97(0.39)*	0.97(0.01)*	0.97(0.01)*	0.97(0.01)*	0.97(0.01)*	0.97(0.39)*

*p-value<0.01; The value presented as HR(SE)

Discussion

For modeling the survival time data, most of researchers are interested in using Cox proportional hazard model rather than parametric models. Although in this model there is no need to a particular distribution for data, the researchers should check an important assumption called the PH assumption (Altman, De Stavola, Love, and Stepniewska, 1995; Paoletti and Asselain, 2010). Although this assumption is important for Cox model, the results of a systematic review revealed that only 5% of the journals which used Cox model assessed PH assumption (Altman et al., 1995). In this case, maybe parametric models have better performance. In our study, PH assumption for each variable and also globally was held but the results of AIC showed that the Cox regression model in comparison of parametric models had the poorest fit to data. Also, among six parametric models, log-normal and Weibull with negligible difference had the lowest AIC, respectively. In other word the results of AIC and also Coxsnell residual showed Log-normal and Weibull with a slight differences were had fitted better than other parametric models and also Cox model. Poorhoseingholi et al. compared the efficiency of the Cox regression and parametric models in the survival analysis on patients with gastric cancer treated in Iran. Their findings showed that the Lognormal model could be replaced with Cox model survival analysis in gastric cancer data (Pourhoseingholi et al., 2007). Shayan et al. concluded that the parametric models such as Generalized Gamma and Log-logistic are more preferable and are a useful tool for fitting to first birth interval than Cox (Shayan et al., 2014). Zare et al., (2013) used AIC and Cox-Snell residuals to compare Cox model and parametric models in modeling transition rates of a multi-state model. They showed that parametric models have often been more reliable and less biased. Also, some studies showed when the data were generated by a log-normal model with an exponential conditional mean function, the Cox model performed poorly (Basu et al., 2004).

In general, ALL is more common than AML among children (Zwaan et al., 2015) and this study revealed that more than 83% of patients had ALL. Based on the results of two models, weibull and Log-normal, age at diagnosis, residence area and WBC had an effective role in survival time or hazard ratio of leukemia. The results of a study showed that age, sex and WBC had a significant effect but residence had no significant effect on survival of ALL in children (Erdmann et al., 2014). In addition, others studies demonstrated that the WBC has a significant effect on the risk of death among childhood leukemia (Hussein et al., 2004).

Many factors affect the survival time of the children with acute leukemia including treatment methods, socio-economic status, family income, parental education, residential area, parental job, etc.; due to lack of access to these factors, they were not studied in the present research (Erdmann et al., 2014).

To fit parametric models, the number of censoring observation is better not to be more than 40% to 50%. In our study, although the percentage of censored observation

was more than 50%, parametric models had fitted better than Cox model (Nardi and Schemper, 1999).

One of the limitations of this study is related to incomplete patients' record. For some of variable such as Blood group and Hemoglobin there was many missing data. Therefore, incomplete of these covariates maybe effect on the results of these prognostic factors in the each of models. Another limitation of this study related to some important diagnostics factors such as which were not recorded in patient records at hospital. Existing more factors effect on survival time of acute leukemia patients can be fitted the models more accurately.

Cox proportional model is the common model which used to modeling the survival data but it is not valid at all; especially when the PH assumption does not hold or survival data following a parametric models. The results of this work showed parametric models were better performance to predict survival time of leukemia patients than Cox model.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

We thank Karen Shashok (Author AID in the Eastern Mediterranean) for improving the use of English in the manuscript.

References

- Almasi-Hashiani A, Zareifar S, Karimi M, Khedmati E, Mohammadbeigi A (2013). Survival rate of childhood leukemia in Shiraz, southern Iran. *Iran J Pediatr*, **23**, 53.
- Altman D, De Stavola B, Love S, Stepniewska K (1995). Review of survival analyses published in cancer journals. *Br J Cancer*, **72**, 511.
- Ansin E (2015). An evaluation of the Cox-Snell residuals. Master Thesis, Uppsala University, 5-6
- Coustan-Smith E, Ribeiro RC, Rubnitz JE, et al (2003). Clinical significance of residual disease during treatment in childhood acute myeloid leukaemia. *Br J Haematol*, **123**, 243-52.
- Erdmann F, Kaatsch P, Zeeb H, et al (2014). Survival from childhood acute lymphoblastic leukaemia in West Germany: Does socio-demographic background matter?. *Eur J Cancer* **50**, 1345-53.
- Hougaard P (1999). Multi-state models: a review. *Lifetime Data Anal*, **5**, 239-64.
- Hussein H, Sidhom I, Naga SA, et al (2004). Outcome and prognostic factors of acute lymphoblastic leukemia in children at the national cancer institute, Egypt. *Pediatr Blood Cancer*, **26**, 507-14.
- Jemal A, Siegel R, Ward E, et al (2008). Cancer statistics. CA: *CA Cancer J Clin*, **58**, 71-96.
- Kleinbaum D, Klein M (2005). Survival analysis. A self-learning approach. Springer, New York, USA. pp 289-343.
- Kleinbaum DG, Klein M (2005). Competing risks survival analysis. survival analysis: A self-learning text, 391-461.
- Kurnaz F, Sahin C, Kaynar L, et al (2016). Factors affecting survival in acute leukemia with donor lymphocyte infusion in the first relapse after allogeneic stem cell transplantation. *J BUON*, **21**, 227-34.
- Lau C, Ho C, Kim C, e al (2004). Cytotoxic activities of

- Coriolus versicolor (Yunzhi) extract on human leukemia and lymphoma cells by induction of apoptosis. *Life Sci*, **75**, 797-808.
- Leukemia CL (2008). The leukemia and lymphoma society. InTerneT SITE, 14-32
- Lightfoot T, Johnston W, Simpson J, et al (2012). Survival from childhood acute lymphoblastic leukaemia: the impact of social inequality in the United Kingdom. *Eur J Cancer*, **48**, 263-269.
- Nardi A, Schemper M (1999). New residuals for Cox regression and their application to outlier screening. *Biometrics*, **55**, 523-29.
- Neyestani TR, Fereydouni Z, Hejazi S, et al (2007). Vitamin C status in Iranian children with acute lymphoblastic leukemia: evidence for increased utilization. *J Pediatr Gastroenterol Nutr*, **45**, 141-44.
- Paoletti X, Asselain B(2010). Survival analysis in clinical trials: Old tools or new techniques. *Surg Oncol*, **19**, 55.
- Pourhoseingholi MA, Hajizadeh E, Moghimi Dehkordi B, et al (2007). Comparing Cox regression and parametric models for survival of patients with gastric carcinoma. *Asian Pac J Cancer Prev*, **8**, 412.
- Ravangard R, Arab M, Rashidian A, et al (2011). Comparison of the results of Cox proportional hazards model and parametric models in the study of length of stay in a tertiary teaching hospital in Tehran, Iran. *Acta Med Iran*, **49**, 650.
- Rego M, Pinheiro G, Metze K, Lorand-Metze I (2003). Acute leukemias in Piauí: comparison with features observed in other regions of Brazil. *Braz J Med Biol Res*, **36**, 331-7.
- Zare A, Mahmoodi M, Mohammad K, et al (2013). Comparison between parametric and semi-parametric cox models in modeling transition rates of a multi-state model: application in patients with gastric cancer undergoing surgery at the Iran cancer institute. *Asian Pac J Cancer Prev*, **14**, 6751-5.
- Zwaan CM, Kolb EA, Reinhardt D, et al (2015). Collaborative efforts driving progress in pediatric acute myeloid leukemia. *J Clin Oncol*, **33**, 2949-62.