Original Article

The Effect of Time-dependent Prognostic Factors on Survival of Non-Small Cell Lung Cancer using Bayesian Extended Cox Model

Vida Pahlevani[1] Hossein Fallahzadeh[2] Nima Pahlevani[3] Abolfazl Nikpour[1] **Morteza Mohammadzadeh[1*]**

1. PhD student, Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
2. Professor, Research Center of Prevention and Epidemiology of Non-communicable Diseases, Department of Biostatistics and Epidemiology, School of Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
3. Student of Medicine, School of Medicine, Kashan University of Medical Sciences, Kashan, Iran

*Correspondence to: Morteza Mohammadzadeh
  m_morteza@modares.ac.ir

Abstract

Background and Purpose: Lung cancer is one of the most common cancers around the world. The aim of this study was to use Extended Cox Model (ECM) with Bayesian approach to survey the behavior of potential time-varying prognostic factors of Non-small cell lung cancer.

Materials and Methods: Survival status of all 190 patients diagnosed with Non-Small Cell lung cancer referring to hospitals in Yazd were recorded from 2009 to 2013 by phone call. We fitted conventional Cox proportional hazards (Cox PH) as well as Bayesian ECM. Inference for estimated risk ratios was based on 90% credible intervals. Log pseudo marginal likelihood criteria (LMPL) was used for model comparison. Statistical computations were based on R language.

Results: In this study, 190 patients with non-small cell lung cancer were followed, of whom 160 died because of the disease (84.2%). Median of survival time was 8 ± 0.076 month. After fitting the Cox PH Model, it was determined that the PH assumption was not satisfied for the type of treatment, the disease stage, and pathology status variables ($p < 0.001$). LPML for Cox PH and Bayesian ECM was -431.593 and -401.01, respectively. Estimated hazard ratio curves based on Bayesian ECM showed that the risk ratio for these variables exhibited significant time varying behavior on hazard of lung cancer through follow up time.

Conclusion: Based on LMPL, Bayesian ECM was found to have a better fit than Cox PH Model which declares, results from Cox PH should be interpreted with care. Especially, from beginning of the study to about 20 month after, very high risk ratio was estimated for variables whose PH was not satisfying for them.

Keywords: Bayesian approach; Chemotherapy and radiotherapy; Extended Cox regression; Time-dependent variables; Non-small cell lung cancer

Citation: Pahlevani V, Fallahzadeh H, Pahlevani N, Nikpour A, **Mohammadzadeh M***. The Effect of Time-dependent Prognostic Factors on Survival of Non-Small Cell Lung Cancer using Bayesian Extended Cox Model. Iran J Health Sci. 2020; 8 (3): 29-36.

1. Introduction

Cancer is a complex disease involving numerous tempo-spatial changes in cell physiology, which ultimately leads to malignant tumors. Abnormal cell growth (neoplasia) is the biological endpoint of the disease. Tumor cell invasion of surrounding tissues and distant organs is the primary cause of morbidity and mortality for most cancer patients (1). Lung cancer (LC) is the leading cause of cancer death among men and the second leading cause of cancer death among women worldwide. Lung cancer rates and trends vary substantially by sex, age, race/ethnicity, socioeconomic status, and geography, because of differences in historical smoking patterns (2). Non-small-cell lung cancer (NSCLC) is the most deadly type of cancer in the United States and worldwide. Although new therapy is available, the survival rate of NSCLC patients remains low (3) . NSCLC mainly consists of two major histological types: adenocarcinoma (AC) and squamous cell carcinoma (SCC). In 2011, there were estimated to be 221,130 new cases of lung cancer in the United States. Over a million people died of lung cancer worldwide this year alone. When possible, surgery to remove the tumor is the best treatment strategy for patients with NSCLC. However, even with adjuvant (postoperative) chemotherapy and radiation, more than 40% of patients will develop recurrences locally or systemically and ultimately succumb to their disease (4) .One of these mechanisms is autophagy. Autophagy is a protected catabolic process in which proteins and organs are omitted through lysosomes. During this, the process of cytoplasm segregation is separated by two specific membrane vesicles called autophagosomes, which are rapidly combined with an endosome or lysosome, creating an autolysozum. The localization of lysosomal hydrolysis causes destruction of cytoplasmic cargo and eventually demolition of products and their release into cytosol for recycling (5). Survival analysis is a collection of statistical procedures for data analysis and the outcome variable of interest is time until an event occurs (6). Perhaps most widely used statistical tool for survival analysis is Cox Proportional Hazards (PH) Model. A key reason for the popularity of the Cox Model is that, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data situations (7). Of course, there are some limitations when PH assumption is not valid – that is, hazard ratio change over time significantly. A comprehensive study by Kalantar-Zadeh showed that PH assumption is violated in a noticeable number of published articles. In this case, several approaches are presented to extending Cox Model to handle time-varying covariates (8). Most often, these methods will increase interpretability of the results, but in cost of adding unknown parameters and complexity of the model. These problems cause critical limitations in simultaneous inference of estimating hazard ratios and their standard errors, specifically in low sample size studies. A good option to deal with such constraints in Cox PH Model with low sample size is the use of Bayesian approaches. Bayesian Cox Models are successfully applied on cancer studies (9,10). Because of the great importance of previous information in natural sciences, and medicine in particular, the use of past patient's information about the effect of prognostic factors on their survival time can help us to collect

sufficient evidence and express diagnosis with a higher probability. Also, relying on the prior information in similar accessible articles minimizes the need for repeating large sampling (11). Most studies use non-Bayesian methods and have limitations, such as the lack of expert knowledge on statistical tests, the sensitivity of the results to sampling errors, and the reduction of statistical power. In this study, we aimed to apply extended Cox Model with Bayesian approach to survey the behavior of potential time-varying prognostic factors of lung cancer and avoid misleading interpretation of the results.

2. Materials and methods

Firstly, a checklist containing the characteristics of the patients and all the factors examined (age, type of pathology, gender, disease stages, history of smoking, type of treatment, duration of smoking, and survival of patients) was prepared . The event in our study was death from the Non-Small Cell lung cancer. The survival status of all 190 patients diagnosed with Non-Small Cell lung cancer referring to Shaheed Ramezanzadeh and Kargar hospitals in Yazd were recorded from 2009 to 2013 by phone call. Kaplan-Meier estimates were used to describe the survival of patients. For inference purpose, we fitted conventional Cox PH as well as Bayesian extended Cox Model.

Time-varying coefficients offer great flexibility in capturing the temporal behavior of covariate effects on event times, which could be hidden from a Cox proportional hazards (Cox PH). Consider a Cox Model with time-varying regression coefficients. Conditional on a p-dimensional vector of covariates, $\mathbf{Z}$, the hazard function is

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp\{\mathbf{Z}^T \boldsymbol{\beta}(t)\} \quad (1)$$

where $\lambda_0(t)$ is the baseline hazard, and $\boldsymbol{\beta}(t)$ is the p-dimensional regression coefficient function of main interest. If $(t) \equiv \boldsymbol{\beta}$, the model (1) will reduce to conventional Cox PH Model. We used a dynamic Cox regression model for right censored data in a Bayesian framework, where the coefficient curves were piecewise constant (12). In the Bayesian approach, we can use previous information for $\boldsymbol{\beta}(t)$ to more accurately estimate the risk ratios over time. The posterior computation is carried out within the Gibbs Sampling Framework. Inference of each coefficient is based on 90% Credible Intervals (CI) instead of confidence interval and p-value. For each model, we generated 55,000 Markov Chain Monte Carlo (MCMC) samples with a burn-in period of 5,000 and thinned the output by 10. The remaining 5,000 MCMC samples were used for convergence checking and summary. Instead of AIC which is used for classical model comparison, we used log pseudo marginal likelihood (LPML) for model comparison (13). Models with higher LPML were preferred to models with lower LPML. Descriptive statistics and plotting were done with "survival" and "ggplot2" R packages. To fit time-varying model in Bayesian framework, we used "dynsurv" R Package (12).

3. Results

In the present study, 190 patients with Non-Small Cell lung cancer were followed from 2009 to 2013, of whom 160 died by the disease (84.2%). Total follow up time was 48 months. Mean and median of survival time was 11.29 ± 0.781 and 8 ± 0.076 month, respectively.

Also, 1, 9 and 18 month cumulative survivals percent for lung cancer patients were 0.947, 0.433, and 0.16, respectively, which showed a dramatic decrease in survival of the patients. Detailed descriptive statistics are reported in Table 1.

Table 1. Characteristics of prognostic factors in Non-small lung cancer patients

Risk Factor	Levels	N(%)	Mean of survival time ±SD
Type of treatment	Chemotherapy and radiotherapy	71(37.3%)	14.4 ± 0.5
	surgery	53(27.8%)	4.5 ± 0.6
	Chemotherapy	66(34.7%)	8.0 ± 0.5
Age (year)	23-6o	77(40.5%)	12.4 ± 1.2
	60-91	113(59.5%)	9.9 ± 0.7
History of smoking	yes	110(57.9%)	10.4 ± 0.8
	no	80(42.1%)	11.9 ± 1.2
Disease Stage	II	58(30.5%)	15.3 ± 1.1
	III	56(29.5%)	11.6 ± 1.0
	IV	76(40%)	6.0 ± 0.6
Gender	Female	54(28.4%)	11.7 ± 1.0
	Male	136(71.6%)	10.5 ± 0.7
Type of pathology	Adenocarcinoma	77(40.5%)	12.2 ± 1.0
	Squamous cell carcinoma	55(28.9%)	10.5 ± 0.9
	Large cell carcinoma	58(30.5%)	8.9 ± 1.1
Duration of smoking	0	80(42.1%)	11.3 ± 0.9
	Less than 1 year	50(26.3%)	10.6± 1.2
	More than 1 year	60(31.5%)	10.0 ± 0.9

The Variance Inflation Factor (VIF) index was used to eliminate prognostic factors with high correlation in the multiple Cox regressions, simultaneously (17). This index had critical values for the smoking status and duration of smoking variables (VIF >5). Therefore, the smoker's status variable was not entered in the multiple regression. After fitting the conventional Cox PH Model, it was determined that the PH assumption was not satisfied for the type of treatment (p-value<0.001), and the disease stage (p-value<0.001) and pathology status (p-value<0.001) variables, which means, the risk ratio (HR) for these variables varied through the study time. Figures 1-3 shows time-varying estimated hazard ratio curves by Bayesian Extended Cox Model based on formula (1) with 90% credible intervals for type of treatment, pathology status, and disease stage variables. Solid red line is fix hazard ratio which is estimated by conventional Cox PH Model. This line is added to Figures 1-3 to increase comparability of the two models. By looking at these figures, we can see noticeable differences in risk ratio estimated by two models through study time, especially from the beginning of study to 20 months later. LPML index for Bayesian Extended Cox Model was -401.01. Hazard ratios based on conventional Cox PH Model for time independent is reported in Table 2. As is shown in Table 2, age, gender, and duration of smoking did not have a significant effect on risk of death in our study. LPML index for Cox PH was found to be -431.593. Thus, Bayesian Extended Cox Model has a better fit than Cox PH Model.

Table 1. Estimated logarithm (log) hazard ratios for patients with time-independent prognostic factors by conventional Cox PH Model

Risk Factor	Levels	Log HR (CI-90%)
Age (year)	23-65	
		Ref
	66-91	0.04(-0.3,0.75)
Gender	Female	Ref
	Male	0.02(-0.36,0.41)
Duration of smoking	0	Ref
	<1 year	-0.2(-0.59,0.17)
	>1 year	0.08(-0.32,0.49)
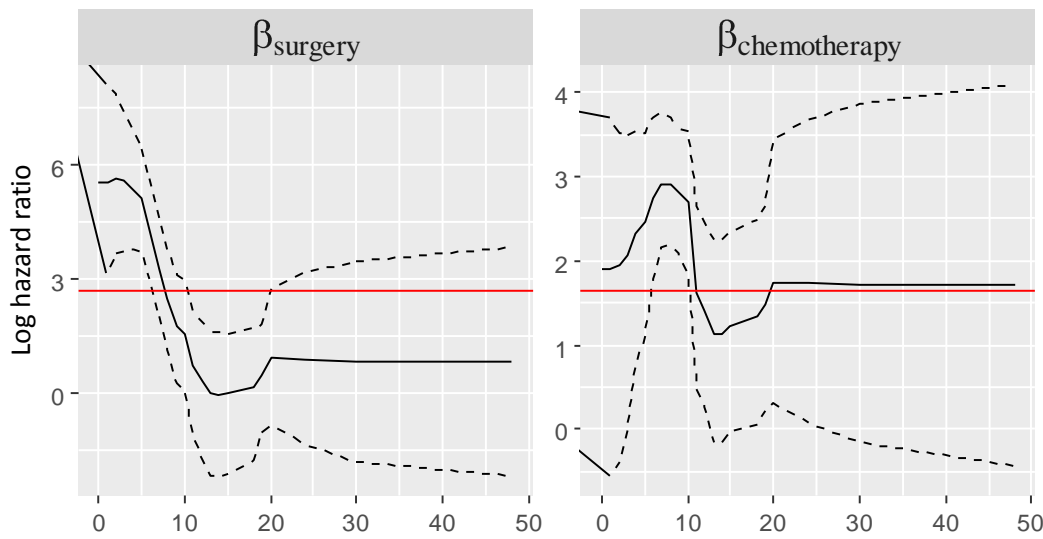
Figure 1. Estimates of coefficient and 90% credible intervals for treatment variable, using Bayesian Extended Cox (black curve) and Conventional Cox PH Models (solid red lines). Chemotherapy and radiotherapy is reference level.
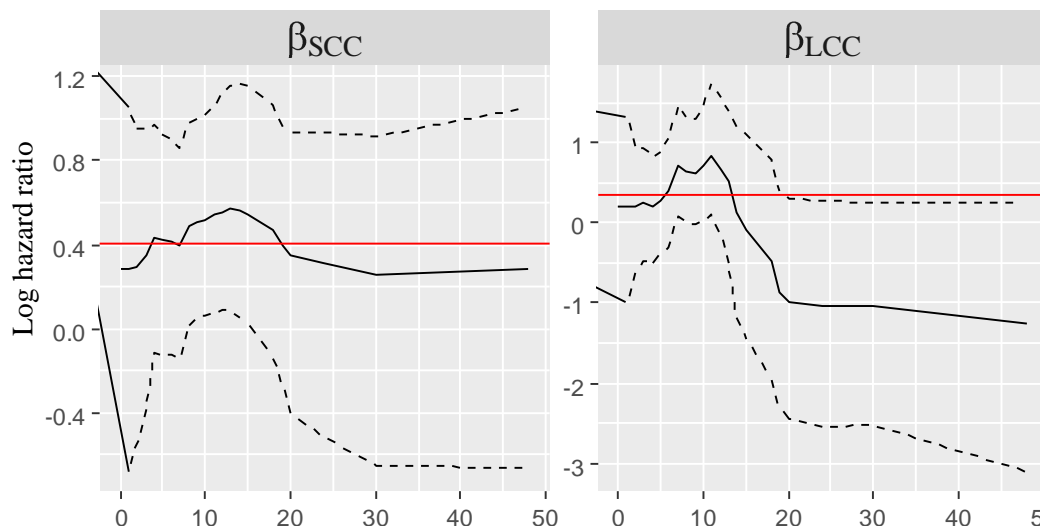

Figure 2. Estimates of coefficient and 90% credible intervals for pathology status variable, using Bayesian extended Cox (black curve) and conventional Cox PH models (solid red lines). SCC and LCC stands for Squamous Cell Carcinoma and Large Cell Carcinoma pathology status, respectively. Adenocarcinoma is reference level.
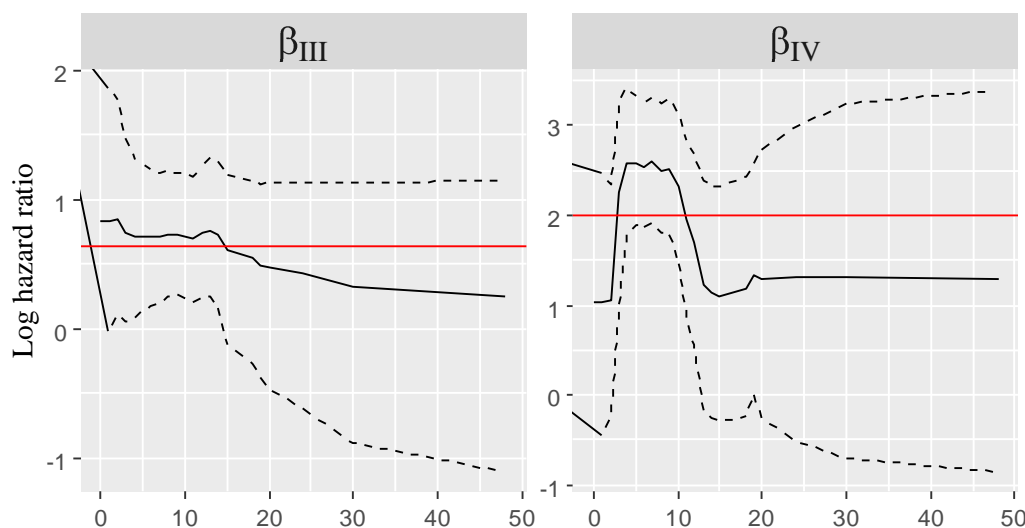
Figure 3. Estimates of coefficient and 90% credible intervals for stage variable, using Bayesian extended Cox (black curve) and conventional Cox PH models (solid red lines). Stage II is reference level.

4. *Discussion*

This study was designed to evaluate prognostic factors of Non-Small Lung Cancer Survival time. Age, gender, and duration of smoking was time independent variables. Since PH assumption violated for the type of treatment, cancer stage and the pathology status variables during the study time, we should not consider a fix estimate for HR through the study. Hence, we used the Extended Cox Model with Bayesian framework introduced to handle these time-dependent variables and to avoid misleading interpretation of results that might have been understood by Conventional Cox PH Model(12) . In opposition to Cox PH Model, the extended Cox Model, estimates a hazard ratio curve for each variable during follow up time. The interpretation of findings from figures 1-3 are as follows: figure 1 shows risk ratios for treatment variable. We set chemotherapy and radiotherapy as reference level. From conventional Cox PH Model (solid red line), $\beta_{surgery}$ is 2.683 (90% CI: 2.274 − 3.098); $\beta_{chemothrapy}$ is 1.653 (90% CI: 1.269 − 2.048). That is, patients with

surgery or chemotherapy have higher risks than those who received both chemotherapy and radiotherapy treatment. These results were consistent with those in the study of Yang et al. in 2017, in which not any significant difference was found between the type of treatments on the lung cancer survival (14). Similar to our study, Fernandez et al., in 2018, used extension Cox because Cox assumption was not satisfied. They found that in 3 until 18 months, hazard ratio increased for patients who had surgery than patients with radiotherapy (15). From Bayesian Extended Cox Model, however, both coefficients exhibited changes over time, suggesting that the inferences treatment based on the classical proportional hazards Cox Model could be misleading. The difference in risk between those with surgery and those receiving both chemotherapy and radiotherapy is substantially different from beginning the study to 16 months, with coefficient magnitude reducing from 5.6 to 0. Then, it remains insignificant after 16 months to the end of the study period. Clearly, we would have over and under estimate of risk ratio

for patients with surgery if we used Cox PH Model (red line). Coefficient magnitude for patients with only chemotherapy treatment against reference treatment reached a peak of about 4.5 around month 10, and then converged to 2.30. Also, Figure 2 shows the estimated risk ratios for pathology status variable. Adenocarcinoma pathology status was as reference level. From solid red line, β_{SCC} was 0.406 (90% CI: 0.069 − 0.753) which was approximately consistent with the estimated curve by Bayesian Extended Cox Model. Qualitatively, similar results reached in Baine et al. in 2018 (16), in which the hazard of Squamous cell carcinoma was higher than non-small cell carcinoma, but it was not significant in Wang in 2018 (17). β_{LCC} is 0.344 (90% CI: -0.030 − 0.719) which is not significant. But its equivalent risk ratio estimated from Bayesian Extended Cox Model (black curve), suggesting that it varied insignificantly between -1 to 1 around 20 months, and then, significantly converged to -1 up to the end of follow up time. That is, after month 20, patients with LCC had 0.63 % less than those with Adenocarcinoma pathology status. Stage II was set to reference level for stage variable. From red line in Figure 3, we see, $\beta_{stage\ III}$ was 0.406 (90% CI: 0.069−0.753), which qualitatively was similar to the study of Pacheco et al. in 2019, in which the high stage of the disease was considered as a negative risk factor (18) . Estimated risk ratio curve for stage III versus stage II confirmed this result just from beginning to 15 months and not for all follow up period. Thus, it can be said that risk of death for patients in stage III was 1.5 times more than stage II, just to month 15. Also estimated risk ratio from Bayesian Extended Cox Model (black curve), suggested that time-

varying nature of $\beta_{stage\ IV}$ was more obvious than $\beta_{stage\ III}$. It dramatically increased to a peak of about 2.5 around 10 months and quickly diminished to about 1 around 5 months later, and then, it remained significantly stable to the end of the study period − that is, the high risk of death for patients in stage IV versus stage II was more obvious at the first 10 months of the follow up time. In summary, we can see that risk ratios have different behavior before and after 20 months, for time-varying prognosis factors. Conclusions from Figures 1-3 might be similar with other models that put covariate effects on survival rate, but with benefits of Bayesian framework, which gives more external validity to our results (19) (20). To the best of knowledge of researchers of the current study, this is the only study which used Bayesian Extended Cox regression on Non-small ling cancer data. We hope our findings give new insights to scientists, especially physicians, to prescribe better treatment for their patients. The main limitation of our study was that the prior information was specified half informatively from the perspective that, we just specified prior distribution whether variables are risk factors or not, based on previous studies. Narrower credible intervals might be achieved if full informative priors are applied in Bayesian Model (1) for future studies.

5. Conclusion

According to the results of the present study, Bayesian Extended Cox Model had a better fit than Cox PH Model which declares that results from Cox PH should be interpreted with care. Especially, from the beginning of the study to about 20 months later, very high risk ratio was estimated for variables, which showed PH was not

satisfying for them, that is to say, inferences about risk ratio for the type of treatment, pathology status, and disease stage variables could lead to misleading interpretations if we used a constant risk ratio based on Conventional Cox PH Model for statistical inference purpose.

Acknowledgments

References

1. Seyfried TN, Shelton LM. Cancer as a metabolic disease. Nutrition & metabolism. 2010;7(1):7.
2. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. In: Lung cancer and personalized medicine. Springer; 2016. p. 1–19.
3. Zhang H, Guttikonda S, Roberts L, Uziel T, Semizarov D, Elmore SW, et al. Mcl-1 is critical for survival in a subgroup of non-small-cell lung cancer cell lines. Oncogene. 2011;30(16):1963–1968.
4. Tucker ZC, Laguna BA, Moon E, Singhal S. Adjuvant immunotherapy for non-small cell lung cancer. Cancer treatment reviews. 2012;38(6):650–661.
5. Schiwitza A, Schildhaus H-U, Zwerger B, Rüschoff J, Reinhardt C, Leha A, et al. Monitoring efficacy of checkpoint inhibitor therapy in patients with non-small-cell lung cancer. Immunotherapy. 2019;11(9):769–782.
6. Cox DR. Analysis of survival data. Chapman and Hall/CRC; 2018.
7. Kleinbaum DG, Klein M. Survival analysis. Springer; 2010.
8. Kalantar-Zadeh K, Kuwae N, Regidor DL, Kovesdy CP, Kilpatrick RD, Shinaberger CS, et al. Survival predictability of time-varying indicators of bone disease in maintenance hemodialysis patients. Kidney international. 2006;70(4):771–780.
9. Klein JP, Van Houwelingen HC, Ibrahim JG, Scheike TH. Handbook of survival analysis. CRC Press; 2016.
10. Ibrahim JG, Chen M-H, Sinha D. Bayesian Survival Analysis. Wiley StatsRef: Statistics Reference Online. 2014;

11. Lee E, Zhu H, Kong D, Wang Y, Giovanello KS, Ibrahim JG. BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. The annals of applied statistics. 2015;9(4):2153.
12. Wang W, Chen M-H, Wang X, Yan J. dynsurv: Dynamic Models for Survival Data [Internet]. 2017. Available from: https://CRAN.R-project.org/package=dynsurv
13. Wang X, Chen M-H, Yan J. Bayesian dynamic regression models for interval censored survival data with application to children dental health. Lifetime data analysis. 2013;19(3):297–316.
14. Yang H-X, Woo KM, Sima CS, Bains MS, Adusumilli PS, Huang J, et al. Long-Term Survival Based on the Surgical Approach to Lobectomy for Clinical Stage I Non-Small Cell Lung Cancer: Comparison of Robotic, Video Assisted Thoracic Surgery, and Thoracotomy Lobectomy. Annals of surgery. 2017;265(2):431.
15. Fernandez FG, Kosinski AS, Furnary AP, Onaitis M, Kim S, Habib RH, et al. Differential effects of operative complications on survival after surgery for primary lung cancer. The Journal of thoracic and cardiovascular surgery. 2018;155(3):1254–1264.
16. Baine MJ, Verma V, Schonewolf CA, Lin C, Simone II CB. Histology significantly affects recurrence and survival following SBRT for early stage non-small cell lung cancer. Lung cancer. 2018;118:20–26.
17. Wang S, Chen A, Yang L, Cai L, Xie Y, Fujimoto J, et al. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. Scientific reports. 2018;8(1):1–9.
18. Pacheco JM, Gao D, Smith D, Purcell T, Hancock M, Bunn P, et al. Natural History and Factors Associated with Overall Survival in Stage IV ALK-Rearranged Non–Small Cell Lung Cancer. Journal of Thoracic Oncology. 2019;14(4):691–700.
19. Fine JP, Yan J, Kosorok MR. Temporal process regression. Biometrika. 2004; 91(3):683–703.
20. Peng L, Huang Y. Survival analysis with temporal covariate effects. Biometrika. 2007; 94(3):719–733.