Review Article

What is the best method for long-term survival analysis?

G Nural Bekiroglu, Esin Avci[1], Emrah G Ozgur

Department of Biostatistics, Marmara University, Medical School, İstanbul, [1]Department of Statistics, Giresun University, Faculty of Art and Sciences, Giresun

Correspondence to: Emrah G Ozgur, E-mail: emrahgokayozgur@gmail.com

Abstract

In the Cox proportional hazards regression model, which is the most commonly used model in survival analysis, the effects of independent variables on survival may not be constant over time and proportionality cannot be achieved, especially when long-term follow-up is required. When this occurs, it would be better to use alternative methods that are more powerful for the evaluation of various effective independent variables, such as milestone survival analysis, restricted mean survival time analysis (RMST), area under the survival curve (AUSC) method, parametric accelerated failure time (AFT), machine learning, nomograms, and offset variable in logistic regression. The aim was to discuss the pros and cons of these methods, especially with respect to long-term follow-up survival studies.

Keywords:
Cox regression, long-term follow-up, restricted mean survival time analysis, survival analysis, area under the survival curve method

Introduction

One of the most used methods for survival analysis is the Cox proportional hazards regression (CPHR) model, which is a semiparametric model where the survival time (the outcome) is assumed to have a known distribution. CPHR is used to investigate the effect of various variables on time when a particular event occurs.[1,2] However, when it comes to survival in the long term, the effects of independent variables on survival are not always proportional over time, in which case the application of CPHR is not correct. There are many methods in the literature that can be used in such a situation. Almost all these methods have an effect on long-term survival and contain mutually supportive information that can complement and/or support each other.

In this study, alternative survival analysis methods that require long-term follow-up are discussed, and the pros and cons of these methods are examined.

Access this article online	
Website: www.indianjcancer.com	Quick Response Code:
DOI: 10.4103/ijc.IJC_22_21	

Materials and Methods

In this section, CPHR model analysis is described particularly in terms of its limitations, especially in long-term follow-up, and therefore, the following alternative methods in survival analysis are highlighted: offset variable in logistic regression, milestone survival analysis, restricted average survival time analysis (RMST), area under the survival curve (AUSC) method, parametric accelerated failure time (AFT) models, machine learning, and nomograms. All these methods predict the probability of survival and the effect of various independent variables on survival more powerfully where the proportionality assumption of CPHR model analysis is violated.[3-6] The importance of the methods has been explained.

How to cite this article: Bekiroglu GN, Avci E, Ozgur EG. What is the best method for long-term survival analysis? Indian J Cancer 2022;59:457-61.

Submitted: 08-Jan-2021 **Revised:** 07-Feb-2021
Accepted: 24-Mar-2021 **Published:** 17-Feb-2023

Cox proportional hazards regression model analysis

The CPHR model, which is a semiparametric method, does not assume a specific "survival model," but rather assumes that the effects of independent variables on survival remain constant over time, thus proportionally adding these effects to survival.[2]

Hazard ratio (HR) is a comparison between the probability of an event in a treatment group and the probability of an event in a control group. If the survival function of a statistically significant variable has the pattern shown in Figure 1a, the effects of the independent variable on survival remain constant over time, and the condition of proportionality assumption is ensured. In this case, the CPHR model analysis can be applied to variables.

However, if the effects of independent variables on survival are not constant over time, and if the proportionality is not ensured as in Figure 1b, the model estimates made by CPHR model analysis cause misleading inferences. This is a very common situation especially in studies requiring long-term follow-up.[3]

In cases where proportionality cannot be achieved, there are many alternative methods. Some of these are described below.

Milestone survival analysis

Milestone survival analysis refers to a cross-sectional evaluation of overall survival (OS) when cases from a predetermined time point reach the end point.

Milestone Survival Analysis is a descriptive analysis, as it calculates OS over a predetermined time segment using the Kaplan–Meier method, rather than evaluating all survival data. Therefore, this method can generally be regarded as an interim analysis because it allows for various adjustments to be made at certain points during the study. The correct time to be determined as the milestone is the time point at which it is believed that the treatment benefit is likely to remain constant, especially from a clinical point of view. In this context, the selection of the milestone must be based on clinically significant criteria. Another important point is that the milestone does not necessarily require long-term survival time.[7,8]

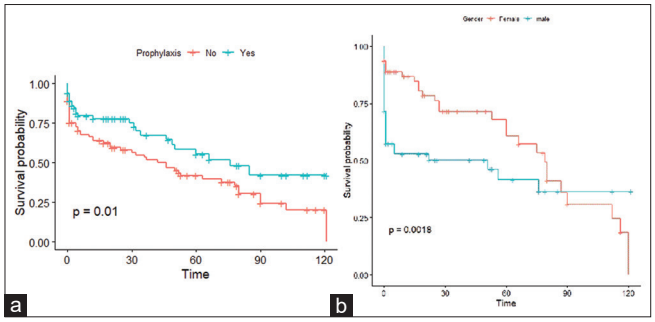


Figure 1: (a) The survival function with proportionality. (b) Survival function where proportionality cannot be achieved

Restricted mean survival time analysis

RMST is interpreted as the average disease-free survival time up to a predetermined, clinically significant time point. In simple terms, it is equivalent to the area under the Kaplan–Meier curve in the disease-free period from the start of the study to a certain time point. RMST can be applied to both progression-free survival and OS.

The most important difference of RMST compared with other survival analyses is that the disease-free survival time offers insight into the extent to which the treatment provides an absolute benefit or disadvantage in comparison with the control. The treatment effect calculated by RMST with the 95% confidence interval can be easily understood and interpreted by clinicians. RMST difference is the difference in areas under the Kaplan–Meier plots for each group and represents a survival gain for treated patients versus controls. The results of RMST are equally as powerful as those of tests where the proportional hazard assumption is valid, such as the log-rank test or HR. It even provides powerful results when the proportional hazard assumption is violated. This advantage is methodologically important.[9-13]

Effective communication of treatment in clinical trials is a prerequisite for making joint treatment decisions. RMST is a robust measure that represents the average overall and/or disease-free survival time over a predetermined period and provides useful information about treatment effect by complementing conventional measures such as relative and absolute risk reductions.[14]

Area under the survival curve (AUC) method

In patients receiving long-term treatment, especially some anticancer treatments (e.g., immunotherapies), the survival probability decreases, and graphically the survival curve appears as a plateau in the right tail. To display the presence of a plateau in survival curves, Damuzzo *et al*. (2019), proposed the ratio between the AUSC and median survival as a new parameter.[7]

Although the AUSC method is similar to the restricted AUC (RMST), the difference between them lies in the fact that the AUC may include the extrapolation of survival from a predetermined point to infinity, whereas the restricted AUC is calculated from Time 0 until the predetermined point in the follow-up.[15,16] Both AUC and RMST[17] are relatively simple but can be combined with models dependent on Gompertz and Weibull, which require complex mathematical analysis.

Key Message

In survival analysis, especially for long term survival analysis there are a lot of alternatives regression methods when Cox proportional hazards regression model assumptions are violated.

The AUSC method, similar to HR, can provide information on the ratio between the AUSC values of the treatment group and the control group. However, if for one arm of the trial, the ratio of the AUSC over the median is greater than 1, it means that the long-term survivors have a greater impact on the OS pattern.[16,17]

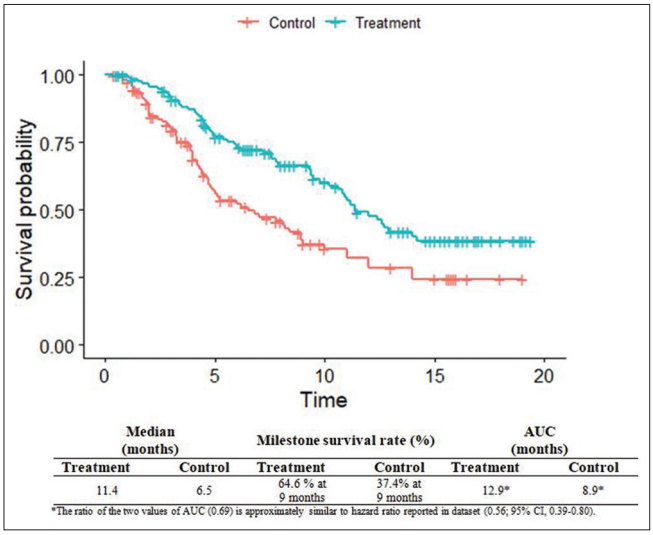
To describe these two methods, which have different and incompatible perspectives of analysis, survival curves for data on non–small-cell lung cancer are given in Figure 2. In this figure, if the milestone is placed between the values of the two medians (i.e., at 9 months) where the magnitude of survival probability difference is maximized, it can be concluded from the findings that "the median is the message." Therefore, there would be no advantage in placing the milestone at a longer follow-up point.

In the AUC analysis, the AUC/median ratio was determined to be 1.13 in the treatment group and 1.36 in the control group, demonstrating greater AUC than the median value in both groups, as expected.

The ratio of both AUC was approximately similar to the HR.

Parametric accelerated time of failure models

The most important features of these models are that the hazards, which are a constraining assumption as in the CPHR model, do not have proportionality, and the parameter estimates are as powerful as in the CPHR model. Another advantage of these models is that estimates are calculated and compared in terms of survival times rather than HRs.[1,18-20]



Median (months)		Milestone survival rate (%)		AUC (months)	
Treatment	Control	Treatment	Control	Treatment	Control
11.4	6.5	64.6 % at 9 months	37.4% at 9 months	12.9*	8.9*

*The ratio of the two values of AUC (0.69) is approximately similar to hazard ratio reported in dataset (0.56; 95% CI, 0.39-0.80).

Figure 2: The comparison of performance between the milestone method and the area under the curve (AUC) method

Although interpretations are made with HR in the CPHR model, in the AFT model interpretations are made with time ratio (TR). TR is a comparison of the ratios of the distance "travelled" by participants on the survival curve. A survival function that starts with 100% survival at Time 0 descends in the direction of 0% survival as time increases. The "baseline" for each participant is the same in the survival curve of the AFT model, although some participants may have progressed faster than others. Some variables may have an effect that further accelerates the progression in the survival curve. For the continuous variables, TR >1 implies that the effect of the variable slowed down or extended the duration of the event, whereas TR <1 indicates that the previous event is more likely to occur. A TR of 0.75 means that each unit increase in the variable has a 0.25-fold less effect on progression in the survival curve.[18]

Hazard distributions are important in parametric AFT models. Some commonly used hazard distributions include exponential, Weibull, log-normal, log-logistic, and generalized gamma models.

Whichever of these distributions fits the hazards' distributions, AFT is applied to that model and the prediction model is implemented.

Machine learning

The increase in the size and complexity of data in the medical field in recent years has exceeded human capabilities. Due to this increase, new technologies for data processing and new algorithms and methods for analysis and modeling are being developed. Machine learning, which is an example of this, is an application of artificial intelligence (AI).[21] It places importance on the development of computer programs that can access data and use them to learn for themselves. The application of machine learning algorithms offers different solutions through the use of different mathematical algorithms. Some examples of machine learning algorithms are linear regression, logistic regression, decision tree, support vector machine (SVM), naive Bayes, *K*-nearest neighbors (KNN), *K*-means, random forest, gradient boosting, AdaBoost, and so on. Deep learning is also a type of machine learning that works based on the

structure and function of the human brain. Machine learning techniques have been used in recent years to construct classification models from survival data and obtain survival probabilities.

Regardless of the area of application, the importance of machine learning methods in analyzing big data and making predictions and decisions is increasing on a daily basis. The clinical validity of the significant independent variables determined by machine learning models, which are the output of the algorithms, has great importance. In addition, values such as the C-index and standard deviation of the models play a very important role when comparing models in decision making for model selection.[22,23]

Nomograms

Nomograms, which are simple and personalized graphs, are used in the diagnosis and prognosis of diseases in survival analysis. A nomogram is a kind of graphical calculation instrument that allows numerous combinations of the inputs of many continuously and/or categorically coded variables in nomogram format. This format distinguishes nomograms from tables, crosstabs, or decision trees, where continuously coded variables cannot be processed. Nomograms are designed to extract the maximum amount of information from data with the aim of providing the most accurate predictions.[24]

In the construction of nomograms, the CPHR model is generally used in variable selection. A good nomogram is expected to have high validity, discrimination, calibration, and clinical utility. Nomograms are usually evaluated by internal validity. The performance of a nomogram is checked by performing cross validation and bootstrap. Discrimination is the process of accurately categorizing patients and healthy people, for which concordance index (CI) and AUC are generally used. These two values range from 0 to 1, and the closer the values are to 1, the better the discrimination feature.[25,26]

Offset variable in logistic regression

The offset variable, which is a covariate, specifies the constant monitoring time, that is, a predetermined and known time. As it is a constant number, it is not necessary to estimate any beta coefficient for this "covariate" in regression. However, when a follow-up variable does not exist in the logistic regression, there will not be any censored data as in the survival analysis. There are two basic variables that are necessary to calculate the probability of survival, one is the variable that determines the patient's condition (dead/alive) during that period of follow-up and the other is the follow-up time. The constant number (e.g., 5 years), which expresses the follow-up time is included in the logistic regression with offset variable and then analyzed. Thus, in a certain follow-up time, variables that are statistically significant according to the patient's situation (dead/alive) and the risks attributed to these variables are determined.[5,6,27]

Although not a widely used method, the offset variable in the logistic regression method provides an estimation of the survival probability when there is no follow-up variable.

Discussion and Conclusion

In survival studies based on long-term follow-up, the survival probability distribution may be uneven, and it is difficult to obtain robust estimates with such a data set. In this case, researchers need to apply a variety of survival methods to reach a better diagnosis.

In this article, survival models for long-term follow-up have been discussed, including the CPHR model, offset variable in logistic regression, milestone survival analysis, RMST, the AUSC method, parametric AFT models, machine learning, and nomogram.

Each model has advantages and disadvantages. CPHR is the first method that comes to mind for survival analysis, but it is not always correct to use this because the conditions are not fully controlled. In these circumstances, the roles of the biostatistician and the biostatistical consulting process become important. Biostatistical consulting is a necessary and beneficial process for medical researchers, because various statistical analysis methods for complex data are currently applicable due to new informatics technologies. The ultimate goal of medical research is to provide the most accurate predictions possible, and the role of the biostatistician is to decide on the most appropriate and accurate model, both statistically and clinically. In this respect, it is also important that appropriate model selection in survival analyses is made according to the objectives of the study.

The models discussed in this study can be considered to be of great benefit to medical researchers not only in increasing understanding but also in practical applications.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

ORCID iDs

G Nural Bekiroglu: https://orcid.org/0000-0001-6471-6612
Esin Avci: https://orcid.org/0000-0002-9173-0142
Emrah G Ozgur: https://orcid.org/0000-0002-3966-4184

References

1. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis Part II: Multivariate data analysis – An introduction to concepts and methods. Br J Cancer 2003;89:431-6.
2. Cox DR. Regression models and life-tables. Breakthroughs in Statistics. New York, NY: Springer Series in Statitstics; 1992. p. 527-41.
3. Gregson J, Sharples L, Stone GW, Butman CF, Öhrn F, Pocock S. Nonproportional hazards for time –to-event outcomes in clinical trials. J Am Coll Cardiol 2019;74:2102-112.
4. Holm RM, Makikallio T, Lindsay MM, Spence MS, Erglis A, Menown IB, *et al*. Percutaneous coronary angioplasty versus coronary artery bypass grafting in the treatment of unprotected left main stenosis: Updated 5-year outcomes from the randomised, non-inferiority NOBLE trial. Lancet 2020;395:191-99.
5. Stone GW, Kappetein AP, Sabik JF, Pocock SJ, Morice MC, Puskas J, *et al*. Five-year outcomes after PCI or CABG for left main coronary disease. N Engl J Med 2019;381:1820-30.
6. Tong BC, Huber JC, Ascheim DD, Puskas JD, Ferguson TB, Blackstone EH, *et al*. Weighting composite endpoints in clinical trials: Essential evidence for the heart team. Ann Throrac Surg 2012;94:1908-13.
7. Damuzzo V, Agnoletto L, Leonardi L, Chiumente M, Mengato D, Messori A. Analysis of survival curve: Statistical methods accounting for the presence of long-term survivors. Front Oncol 2019;9:453.
8. Chen TT. Milestone survival: A potential intermediate endpoint for immune checkpoint inhibitors. J Natl Cancer Inst 2015;107:djv156.
9. Huang B, Kuan PF. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event endpoint. Pharm Stat 2018;17:202-13.
10. Keene ON. Alternatives to the hazard ratio in summarizing efficacy in time-to-event studies: An example from influenza trials. Stat Med 2002;21:3687-700.
11. Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards ssumption is in doubt. Stat Med 2011;30:2409-21.
12. Tian L, Zhao L, Wei J. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. Biostatistics 2014;15:222-33.
13. Wang ZX, Wu HX, Xie L, Wang YN, Yang LP, He MM, *et al*. Correlation of milestone restricted mean survival time ratio with overall survival hazard ratio in randomized clinical trials of immune checkpoint inhibators. JAMA Netw Open 2019;2:e193433.
14. Trinquart L, Bill-Axelson A, Rider JR. Restricted mean survival times to improve communication of evidence from cancer randomized trials and observational studies. Eur Urol 2019;76:137-9.
15. Messori A, Trippoli S, Becagli P, Tendi E. Pharmacoeconomic profile of paclitaxel as a first-line treatment for patients with advanced ovarian carcinoma. A lifetime cost-effectiveness analysis. Cancer 1996;78:2366-73.
16. Messori A, Becagli P, Trippoli S. Median versus mean lifetime survival in the analysis of survival data. Haematologica 1997;82:730.
17. Matucci M, Messori A, Donati-Cori G, Longo G, Vannini S, Morfini M, *et al*. Kinetic evaluation of four factor VIII concentrates by model-independent methods. Scand J Haematol 1985;34:22-8.
18. Karimi A, Delpisheh A, Sayehmiri K. Application of accelareted failure time models for breast cancer patients' survival in Kurdistan province of Iran. J Cancer Res Ther 2016;12:1184-48.
19. Lee ET, Go OT. Survival analysis in public health research. Annu Rev Public Health 1997;18:105-34.
20. Teshnizi SH, Ayatollahi SM. Comparison of cox regression and parametric models: Application for assessment of survival of pediatric cases of acute leukemia in Southern Iran. Asian Pac J Cancer Prev 2017;18:981-5.
21. Krittanawong C, Zhang HJ, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol 2017;69:2657-64.
22. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, *et al*. Cardiovascular event prediction by machine learning: The multi-ethnic study of atherosclerosis. Circ Res 2017;121:1092-101.
23. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, *et al*. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. Eur Heart J 2017;38:500-7.
24. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology – More than meets the eye. Lancet Oncol 2015;16:e173-80.
25. Guan T, Li Y, Qiu Z, Zhang Y, Lin W, Lai Y, *et al*. Nomograms and risk classification systems predicting overall and cancer-specific survival in primary malignant cardiac tumor. J Card Surg 2019;34:1540-9.
26. Hartaigh OB, Gransar H, Callister T, Shaw L, Schulman-Marcus J, Valenti V, *et al*. A simple-to-use nomogram for predicting 5-, 10- and 15-year survival for asymptomatic individuals undergoing coronary artery calcium scoring. J Am Coll Cardiol 2015;65 (10_ Supplement):A1159.
27. Lyles HR, Mitchell EM, Weinberg CR, Umbach DM, Schisterman EF. An efficient design strategy for logistic regression using outcome- and covariate-dependent pooling of biospecimens prior to assay. Biometrics 2016;72:965-75.