International Journal of Engineering & Technology, 7 (4) (2018) 5594-5597

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.21478
Research paper

Survival analysis on heart transplantation patients of different ages using Bayesian survival techniques

Jishnu Raj [1] *, R.Roseline Mary [1]

[1] *Department of Computer Science, Christ University, Bangalore, India*
**Corresponding author E-mail: jishnuraj.mca@christuniversity.in*

Abstract

This study focuses to find the survival chances of the heart transplant patient using different survival analysis. The results are later compared with the survival analysis to check the accuracy. The age-based estimation of survival analysis is also performed. The research carried out using Bayesian Survival analysis with survregbayes, anovaDDP and Kaplan Meir methods for survival analysis. The obtained results have shown that, Bayesian survival analysis is more accurate than the normal survival analysis strategy. Based on variations in mismatched scores, the survival time fluctuates for different age groups. A plot for survregbayes model was generated for which, those below the age of 28 have higher chances of survival. The future enhancement would deploy a PH model for estimation than the currently used PO model. At the later stage the project also aims to find out the cause of hazard from an increased number of variables.

Keywords: *Heart Transplant, Survival Analysis; AnovaDDP; Survregbayes.*

1. Introduction

Heart diseases are the biggest influencers in this era with a changing life-style. For those patients who can't be cured with available medication, heart transplantation is the only solution [3]. The study on the survival rates of patients who have been previously subjected to heart transplantation [8] would surely be a generator event for those patients who are in a dilemma on whether to choose heart transplantation as a permanent solution to their problems [12]. The study would also been eye-opener for those medical-practitioners to set norms for the eligibility required to undergo the surgery.

The first successful heart-transplantation was conducted 35 years ago. Today, cardiac-transplants have become a part of the routine medical-practice. Survival Analysis is defined as the branch of statistics that studies the expected life-time duration of a human-being as well as mechanical systems [15]. The first survival analysis was performed with the emergence of Bill of Mortality and life-table. Most research fields under the medical-science and statistical-domains, today use life-tables [6]. The period of world-war II had posed a trust-seeker query for military equipment. Thus, researches began to be conducted on the life-time or durability of equipment used in the industry. The same research was found out to be applicable in the life-time analysis of cancer patients as well. The only difference being the change of name from "lifetime analysis "to "survival analysis".

Survival time is the biggest deciding factor behind survival analysis. The time duration between starting time and occurrence of an event is measured as survival time. A lot of researches have been previously done to estimate the survival function. Survival analysis is the process of identifying the occurrence of an event in a subject during a specified period of time. The linear regression model cannot be chosen as the ultimate solution to survival analysis problems as it is a function of a set of predictor variables. Linear Regression becomes an effective choice of transformation, as most times the values generated in a survival analysis incident is positive. The breakage or inconsistencies in information cannot be handled by ordinary linear regression. So there comes Bayesian statistics in the Survival analysis.

2. Literature review

Emphasis on survival analysis is done on the duration for which a given event occurs and is continued for a specific period of time [9]. The observation during the event are noted partly states as the censoring. Censoring mechanism is used to treat transplant data. This mechanism is used for every clinical trial [2]. The censoring of observations cannot be handled along a line function by linear regression. Two important aspects touched by survival analysis is the time for which an event occurs and the extend for which an event occurs [7]. Both these functionalities are based on one single factor, time. The main aim of the survival analysis function is to give survival results at a given point of time. Outcomes such as median survival can also be estimated from a survival function. The predictor variables and survival function are also linearly related. The medical domain is a frequent beneficiary of point-in-time survival analysis. When we develop a survival model we need to improve the efficiency [10] of existing model. Right censoring is the situation that occurs when the person is in the living state or has not been followed up at the termination of the study. This only helps in exceeding the time a little longer in survival analysis.

The advantage behind survival analysis is that they permit both broken and unbroken observations to generate quality models. The two parts of dependent event in survival analysis include the event time and the event status. The survival and hazard functions are timebased [1]. The hazard function generates the survival up to a specific time. The survival function provides an insight on the possibility of surviving up to a given time period. Relationship of a factor of interest is generated to the time to the event in the pres-

ence of factors such as age, gender, race. etc. A set of models are present that generate the relation between set of predictor variables and survival time. Parametric, non-parametric and semi-parametric approaches are used in these scenarios. Modelled survival times are compared with various parametric distributions like Weibull, lognormal ...etc.

Bayesian statistics are usually used instead of frequentist approach to design perform analysis and report clinical trials. The additional difficulty of generating Bayesian survival analysis is the need to create a model for survival distribution. After the entry of Bayesian statistics in survival model, a great deal of attention is received to the spatial survival analysis. The package spatsurv [1] and SpBayesSurv [4] in R used for the Bayesian spatial survival analysis. There are many more packages in R to support Bayesian survival analysis. It is more beneficial for the modelling of survival analysis using the parametric, semi-parametric and nonparametric model [11]. The Bayesian nonparametric model is different from other models by fitting single model that can adapt its complexity to the data [14].

With the extensive accessibility and advancement of statistical computing and the of personal computers, usage of sophisticated parametric and non-parametric methods began to be used in analyzing survival data. The computing process takes lesser time. Even bio-medical researchers have moved to the usage of non-parametric procedures in their experiments. Lesser assumptions are identified in case of non-parametric methods. Distributed assumption is not imposed on survival times. Considering all the factors that have led to the failure or death of an individual would be a Herculean task. Non-parametric methods are based on only a few assumptions than parametric methods. Distributional assumptions are not imposed on survival times. It would definitely be a very difficult task to convert all the disease-causing factors into mathematical function formats.

Though non-parametric and semi-parametric methods have advantages and are highly popular, parametric modelling offers useful alternatives. When the propositional hazard is in question, parametric models offers alternatives to Cox model.

3. Materials and methods

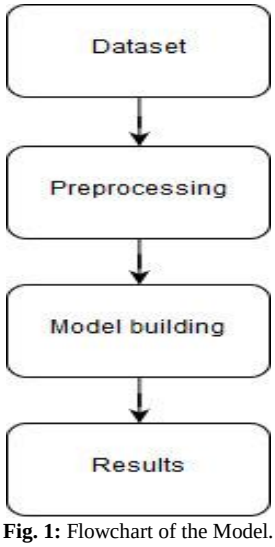

Fig. 1: Flowchart of the Model.

Fig 1 shows the methodology followed for predicting the survival analysis in this study. A systematic approach is deployed to ensure accurate and appropriate results. The heart transplant dataset is chosen and then preprocessed for removing discrepancies which is then analyzed using anovaDDP and survregbayes model and the acquired results are compared for better survival chances.

3.1. Data collection

The data for this research is collected from Stanford University heart transplantation dataset. It is an open source database which also comes along with the survival package in R. In this research the data has been gathered from the Stanford university heart transplantation dataset. The attribute time is the censoring time which is the survival time of heart transplant patient during the observation. The attribute mismatch is the antibody response of the patient. The attribute age refers to the patient age in years and the attribute status is the class label which is intended to predict the survival of the patient after heart transplantation. The class label is very important attribute in the context of survival analysis techniques to predict the survival rate.

Table 1: Dataset Attributes

Attributes	Description
Time	censoring time
Status	status of censoring
Age	Patients age in years
mismatch	Heart transplant mismatch score

3.2. Data pre-processing

In order to build the survival analysis based on the dataset collected it must be pre-processed in such a way that it will be used for further analysis. The data collected from the Stanford university as pre-processed to ensure that the data is clean and fit to use. The dataset had few NA values and those values were removed and the age attribute is transformed to three different levels such as age_group1, age_group2 and age_group3. Censoring is mostly used in survival analysis which eliminates the missing data which is not observed for various reasons. Type-I censoring, right censoring, is considered here. So now the data is free from inconsistent information and missing values.

3.3. Model creation

Fig. 2 shows the flow of survival analysis of heart transplant patient. Crucial factor for the survival analysis is survival function which is the also the main aspect for Bayesian survival analysis. Here survival function is used with AnovaDDP and survregbayes method. AnovaDDP is a Bayesian non-parametric model which is used in non-spatial data in which ANOVA type dependence is defined using the dependent Dirichlet Process, along the various random measures. Used the Kaplan-Meier method to compare the median survival probability with AnovaDDP method. Inclusion of exchangeable Gaussian exchangeable frailties, for fitting right censored survival data is done using the survregbayes method. Semiparametric proportional odds (PO) model fall into the dimensions set by this function.

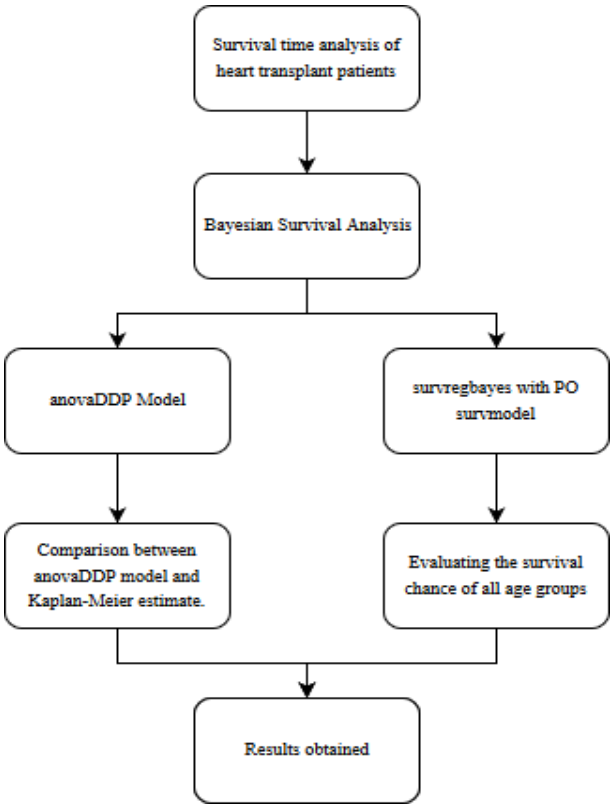

Fig. 2: Implementation of Algorithms in Survival Analysis.

3.3.1. Survival function

$$S(t) = P(T > t) = 1 - F(t) \quad (1)$$

Where the time of event occurred is denoted by random variable T and t is some time. Using the Kaplan-Meier estimate we can plot the graph from the life-table obtained by using the Survival function.

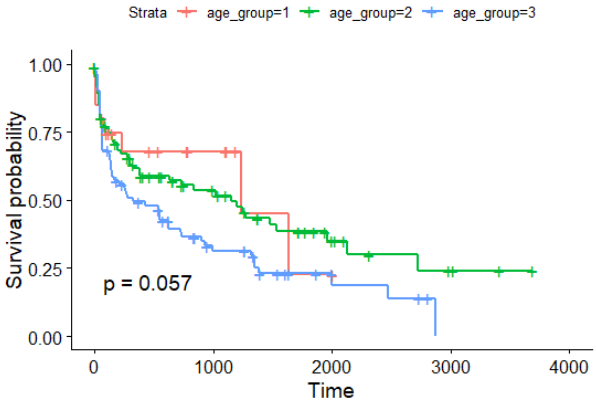

Fig. 3: Kaplan-Meier Graph for Different Age Group in Heart Transplant Data.

The above figure states that the median survival time of the age_group-1 is higher than other age groups.

3.3.2. AnovaDDP

This function is used to fit Bayesian Nonparametric model. So, the data that are used for this model is non-spatial right censored time-to-event data. We assume that a non-spatial LDDPM model [16] with cdf is followed by x_i and $y_i = \log t_i$,

$$F_{x_i}(t) = \int \Phi\left(\frac{\log t - x_i^T \beta}{\sigma}\right) dG\{\beta, \sigma^2\}, \quad (2)$$

Where $\Phi(\cdot)$ is the cdf of the standard normal, and G follows the Dirichlet Process (DP) prior. This Bayesian nonparametric model treats the conditional distribution F_x as a function-valued parameter and allows its variance, skewness, modality and other features to flexibly vary with the x covariates.

3.3.3. Survregbayes

Survregbayes PO model survival and density functions:

$$S_{X_{ij}}(t) = \frac{e^{-x_{ij}^T \beta + v_i} S_0(t)}{1 + (e^{-x_{ij}^T \beta + v_i} - 1) S_0(t)}, \; f_{X_{ij}}(t) = \frac{e^{-x_{ij}^T \beta + v_i} f_0(t)}{[1 + (e^{-x_{ij}^T \beta + v_i} - 1) S_0(t)]^2}, \quad (3)$$

v_i denotes an unobserved frailty associated with s_i, β denotes the regression co-efficient vector and the baseline survival with density $f_0(t)$ denoted by $S_0(t)$. The survregbayes function implements the iid prior distributions:

$$(v_1 \ldots . v_m)^T | \tau \sim IID(\tau^2), \tau^{-2} \sim \Gamma(a_\tau, b_\tau) \quad (4)$$

The independent Gaussian prior $IID(\tau^2)$ is used for non-spatial data and defined as,

$$v_1, v_2, \ldots, v_m \sim N(0, \tau^2) \quad (5)$$

3.3.4. MCMC likelihood function

The MCMC is used to fit the data to the model. The self-tuning adaptive Markov Chain Monte Carlo methods carry out the model parameters and statistical inference, manual tuning is not required here.

$$\mathcal{L}(w_L, \theta, \beta, v) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} [S_{x_{ij}}(a_{ij}) - S_{x_{ij}}(b_{ij})]^{I\{a_{ij} < b_{ij}\}} f_{x_{ij}}(a_{ij})^{I\{a_{ij} = b_{ij}\}}. \quad (6)$$

Markov chain on the hidden variable that has the posterior as its equilibrium distribution defines a Markov-chain on hidden variables [13]. Gibbs sampling is an efficient methodology of MCMC Sampling where conditional distribution of each hidden variable given the others and the observations are identified and presented.

4. Results and discussion

The proposed method shows that the result obtained for the traditional survival analysis and Bayesian survival analysis of the heart transplant patients are almost same. The figure-4 shows that the median survival time of age_group-1 in anovaDDP and Kaplan-Meier are same. According to the result obtained by the Kaplan-Meier their occurred an enormous descent in the age_group-1 but in the case of anovaDDP there is only a slight descent in the curve which infers the age_group-1 has more chances of survival than the other. On the other hand, in Kaplan-Meier it shows that the age_group-2 people have more survival chance than others.

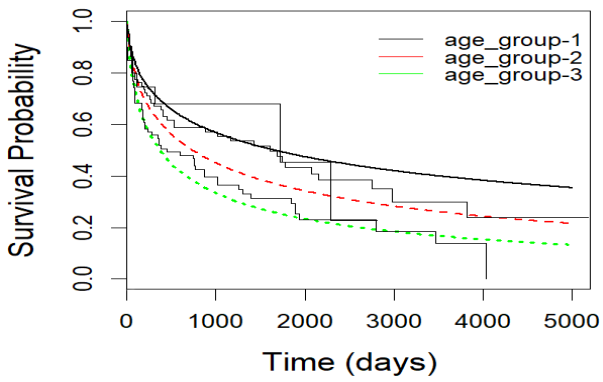

Fig. 4: Graph Shows the Comparison between the Survival Curves of Anovaddp and Kaplan-Meier Curve At Different Age Groups in Heart Transplant Data.

In another experiment we consider the age as 50 with different mismatch score. The figure states that the survival chance of a patient is not only related with covariant age but also mismatch score. When the survival probability decreasing there occurs an increase in survival time for the patient having the mismatch score 2.5.

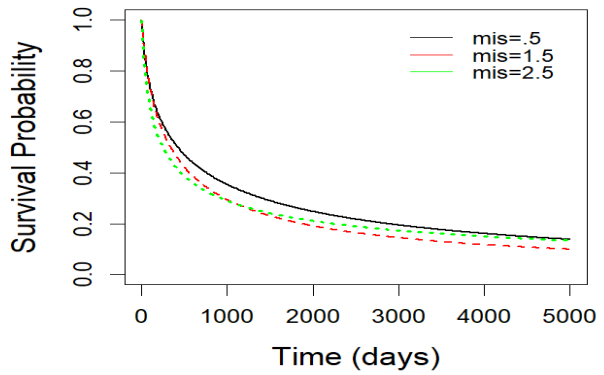

Fig. 5: Anovaddp Graph Shows the Comparison between Different Mismatch Score where Patient Is 50 Years Old.

Prediction of survival chance becomes easier with survregbayes function in spBayes Package in R [4]. Using the PO model in semiparametric model we found the chance of survival of the heart transplant patient is corelated with the variables. When we consider the age_group-1, where the age is less than the mean age, the obtained results are almost same for the patients. But for the age_group-2 the survival chance of the patients is less, but when the mismatch score is less the survival chance is higher which reach up to the age_group-1.

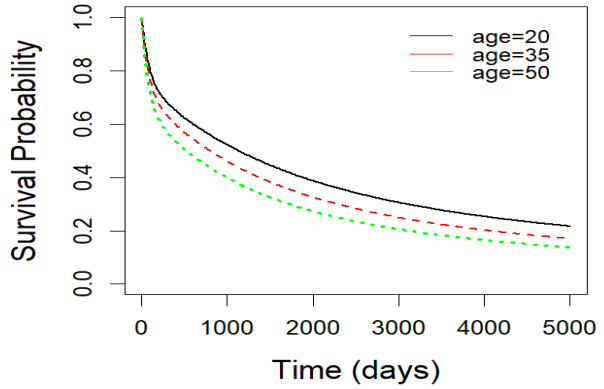

Fig. 6: Graph Shows the Survival Curves for Heart Transplant Patients at Different Ages.

The technical issue that is faced in this experiment is censoring because we are generally using partially observed the time to event data. Another problem that is concerned about is the heart transplant data, we only have few numbers of rows in the data and during pre-processing we are removing all the NA value rows which makes the data less. While applying the model to this data we could only able to obtain the appropriate result while comparing with survival analysis. But if we have a greater number of data, we could have a more accurate survival analysis result which can be used for the future. We could use the parametric models with models with loglogistic, lognormal or Weibull baseline functions for efficient fitting in the survregbayes function [5]. As for the future implementation we would find the cause of hazard from an increased number of variables using the PH survmodel from expanded number of variables.

5. Conclusion

In this paper we have used the anovaDDP and survregbayes methods to represent Bayesian survival analysis of heart transplant patients and reviewed that survival times can be estimated from a censored data. The comparison of proposed method with the Kaplan-Meier method shows that the survival chance of the age_group-2(age between 28 – 45) is far better than other age groups. This study suggests the Bayesian survival technique gives more accurate survival time than the Kaplan Meier technique. Figure-4 shows that possible survival time for each individual depends upon the covariant.

For the future implementation this method can be improved by using the PH SurvModel, which is similar to cumulative hazard model, with left-censored data and time-dependent covariates to find out the cause of hazards for the patients.

References

[1] Taylor, B. M., & Rowlingson, B. S. (2017). spatsurv : An *R* Package for Bayesian Inference with Spatial Survival Models. *Journal of Statistical Software*, *77*(4). https://doi.org/10.18637/jss.v077.i04.

[2] Aitkin, M., Laird, N., Francis, B., & Laird, N. A. N. (2014). A Reanalysis of the Stanford Heart Transplant Data, *78*(382), 264–274.

[3] Borkon, A. M., Muehlebach, G. F., Jones, P. G., Bresnahan, D. R., Genton, R. E., Gorton, M. E., … Rowe, S. K. (1999). An analysis of the effect of age on survival after heart transplant. *The Journal of Heart and Lung Transplantation: The Official Publication of the International Society for Heart Transplantation*, *18*(7), 668–674. Retrieved from https://doi.org/10.1016/S1053-2498(99)00024-8.

[4] Zhou, H., Hanson, T., & Zhang, J. (2016). spBayesSurv: Fitting Bayesian Spatial Survival Models Using R. *Journal of Statistical*, (Taylor 2017), 18637.

[5] Tsujitani, M., & Tanaka, Y. (2013). Analysis of heart transplant survival data using generalized additive models. *Computational & Mathematical Methods in Medicine*, *2013*, 609857. https://doi.org/10.1155/2013/609857.

[6] Politi, P., Piccinelli, M., Poli, P. F., Klersy, C., Campana, C., Goggi, C., … Barale, F. (2004). Ten years of "extended" life: quality of life among heart transplantation survivors. *Transplantation*, *78*(2), 257–263. https://doi.org/10.1097/01.TP.0000133537.87951.F2.

[7] Lee, E. T., & Go, O. T. (1997). Survival analysis in public health research. *Annual Review of Public Health*, *18*, 105–134. https://doi.org/10.1146/annurev.publhealth.18.1.105.

[8] Tjang, Y. S., van der Heijden, G. J. M. G., Tenderich, G., Grobbee, D. E., & Körfer, R. (2008). Survival analysis in heart transplantation: results from an analysis of 1290 cases in a single center. *European Journal of Cardio-Thoracic Surgery*, *33*(5), 856–861. https://doi.org/10.1016/j.ejcts.2008.02.014.

[9] Kartsonaki, C. (2016). Survival analysis. *Diagnostic Histopathology*, *22*(7), 263–270. https://doi.org/10.1016/j.mpdhp.2016.06.005.

[10] Oztekin, A. (2010). Data Mining Based Survival Analysis and Simulation Modeling for Lung Transplant.

[11] Susarla, V., & Ryzin, J. V. (1976). Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations. *Journal of the American Statistical Association*, *71*(356), 897–902. https://doi.org/10.1080/01621459.1976.10480966.

[12] Medved, D., Ohlsson, M., Höglund, P., Andersson, B., Nugues, P., & Nilsson, J. (2018). Improving prediction of heart transplantation outcome using deep learning techniques. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-21417-7.

[13] Chang, I. S., Hsiung, C. A., Yuh-Jenn, W. U., & Yang, C. C. (2005). Bayesian survival analysis using Bernstein polynomials. *Scandinavian Journal of Statistics*. https://doi.org/10.1111/j.1467-9469.2005.00451.x.

[14] Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12. https://doi.org/10.1016/j.jmp.2011.08.004.

[15] Li, Q., Yang, W., Zhu, G. X., & Zhou, J. (2011). Study on survival analysis method for equipment life and its application. *2011 IEEE 18th International Conference on Industrial Engineering and Engineering Management, IE and EM 2011*, (PART 2), 848–850.

[16] De Iorio M, Johnson WO, M¨uller P, Rosner GL (2009). "Bayesian Nonparametric Nonproportional Hazards Survival Modeling." Biometrics, 65(3), 762–771. https://doi.org/10.1111/j.1541-0420.2008.01166.x.