

Atanu Bhattacharjee^{1,2}, GajendraK.Vishwakarma³, Souvik Banerjee³ and Sharvari Shukla⁴

¹Section of Biostatistics, Centre for Cancer Epidemiology, Tata Memorial Centre, Mumbai, India

²Homi Bhabha National Institute, Mumbai, India.

E-mail: atanustat@gmail.com

³Department of Mathematics & Computing, Indian Institute of Technology (ISM), Dhanbad, India

E-mail: vishwagk@rediffmail.com, souvik.stat@gmail.com

⁴Symbiosis Statistical Institute, Symbiosis International University, Pune, India

E-mail: sharvareeshukla@gmail.com

Abstract

Background: The constant news about the coronavirus is scary. It is not possible to separate treatment for Cancer due to COVID-19. An effective treatment comparison strategy is needed. We need to have a handy tool to understand cancer progression in this unprecedented scenario. Linking different events of cancer progression is the need of the hour. It is a methodological challenge.

Methods: This article is dedicated to explore the time lag effect and make statistical inference about the best experimental arm using Accelerated failure time model and regression methods. The work is presented as the occurrence of other events as a hazard rate after the first event (relapse). The time lag effect between the events is linked and analyzed.

Results: The results are further analyzed and presented to show the efficacy of our study. An accelerated failure time model is also applied with the transition states and the dependency structure between the gap times are explained using auto-regression. The effects of Arms are compared using the coefficient of auto-regression and accelerated failure time (AFT) models.

Conclusions: We provide the solutions to overcome the issue with interval between two consecutive events in motivating head and neck cancer (HNC) data. COVID-19 is not going to leave us soon. We have to carry several cancer clinical trials in presences of COVID-19. We presented a comprehensive analytical strategy to analyze cancer clinical trial data during this pandemic.

Keywords: COVID-19, Accelerated Failure Time, Proportional Hazard Model, Bayesian, Auto-Regression.

1. Background

Cancer patients were more likely to develop COVID-19 because they are immunocompromised¹. The individual risk of COVID-19 infection varies from patient to patient. It is required to assess the risk of both COVID-19 and tumor control on a case-by-case basis with the patient. The treatment effect of head and neck cancer (HNC) is obtained in the presence of multiple events like locoregional control (LRC), progression free survival (PFS), and overall death (OD). These events are analyzed separately by Kaplan-Meier and Cox PH. COVID-19 makes these events related. Time lag/interval between different types of event are to be explored. This article is dedicated to explore the time lag effect and make statistical inference about the best experimental arm using Accelerated failure time model and regression methods. The work is presented as the occurrence of other events as a hazard rate after the first event (relapse). The time lag effect is linked. It is true that local relapse biologically triggers cancer progression and death. But we have never measured. Now all events are likely to be influenced by COVID-19 infection. These cannot be isolated.

Relapse trigs disease progression. Similarly, disease progression accelerates death. Further, two-time points are generated as the duration between relapse to progression and progression to death. These are transition periods. The Cox proportional hazard model and Accelerated failure time model are useful to work on transition states. The treatment effect is comparable.

In this manuscript, the statistical model is considered to handle both the time points and explore the relations between gap durations. Further, a Cox regression model is applied to understand the different shapes of the transition hazard model. The time-varying covariates are considered separately. The results are further analyzed and presented to show the efficacy of our study. An accelerated failure time model is also applied with the transition states and the dependency structure between the gap times are explained using auto-regression. The effects of Arms are compared using the coefficient of auto-regression and AFT models. The complete analysis using Bayesian techniques is executed with R open-source software and OpenBUGS.

2. Methods

2.1. Dependency modelling

It is difficult to stop the spread of the infection of COVID-19 from vulnerable cancer patients. We cannot deny treatment to several thousand cancer patients. It is really challenging to handle several thousand cancer and COVID-19 patients separately. There is a very minimal chance that cancer patients will not be infected by COVID-19 in the long run. There is no doubt we have to run several clinical trials in the presence of COVID-19 infection. Suppose we are having n number of cancer individuals. Events may occur as locoregional relapse, progression, and death. The events are marked as 1, 2 and 3 respectively. The events are ordered which implies that the Loco-regional event cannot happen post-progression or death, and death as a terminal event. Let, be the total number of events that happened to a patient? Thus, an individual I corresponds to a cluster of the size of those events. Here, our interest is to measure the event occurrence rate at each of the interval or gap time between two events. Let, d_i be the true event time for individual i . The j could be 1, 2 or 3. Also, we consider that all the individuals have experienced at least one event. The intervals between two subsequent events are defined as follows:

$$G_{i,1} = T_{i,1} \text{ and } G_{i,j} = T_{i,j} - T_{i,j-1} \text{ for } i = 1, 2, \dots, n; j = 1, 2 \quad (1)$$

In our study, the gap times are assumed to be dependent. Events are ordered. To model the dependency structure, we assume that the 1st event corresponds to $G_{i,1}$, the duration from beginning of the study to occurrence of 2nd event, 2nd event corresponds to $G_{i,2}$ and so on. So, the dependency structure is considered among $G_{i,1}$, $G_{i,2}$ etc.

We assume that a simple linear regression model between G_1 and G_2 . The regression model is

$$G_{2,0} = \beta_{0,0} + \beta_{1,0}G_{1,0}, G_{2,1} = \beta_{0,1} + \beta_{1,1}G_{1,1} \quad (2)$$

We fit two separate linear regression model for two different Arms. $\beta_{1,0}$ defines the change in $G_{2,0}$ for unit change in $G_{1,0}$ for Arm 0. The same inference can be drawn for $\beta_{1,1}$. So, ignoring the intercept term in the regression model, the difference between the coefficients $\beta_{1,1} - \beta_{1,0}$ signifies the change in dependent gap time due to change in the Arm. We fit AFT models for G_1 and G_2 and obtain the corresponding coefficients of Arm to measure the change in survival due to change in treatment.

2.2. AFT Model with Gap Time

The accelerated failure time (AFT) model is a popular alternative of proportional hazard model to analyse survival data. It is also applicable in the current COVID-19 scenario. To observe the dependency pattern between observed times, it is more efficient to model the survival time rather than hazard rate. In AFT model, it is assumed that the effect of covariate is to accelerate or decelerate the survival duration by some constants. The AFT model can be expressed as,

$$Y_i = \log(G_i) = \mu + \beta x_i + \varepsilon_i \quad (3)$$

Here, G_i denotes the survival time for i^{th} individual, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ is the unknown regression coefficient, μ is the intercept term, x_i is the covariate for i^{th} individual or subject ($i = 1, 2, \dots, n$), ε_i is the error component, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent and identically distributed with density $g(\varepsilon)$. So, for given covariates, the response times are independent. σ is the scale parameter. In our study, we consider the gap time to fit AFT models for different event occurrences.

For AFT model, the survival function is

$$S(t|x_i) = S_0[\exp\{-(\mu + \beta x_i)\} t] \quad (4)$$

We consider the Bayesian approach to estimate the parameter estimates for the AFT model are obtained from the posterior distributions based on Markov Chain Monte Carlo (MCMC) simulation method. To conduct data analysis using Bayesian techniques, we need to specify the prior distributions of the parameters. We used independent Gaussian prior distributions with mean 0 and variance 1×10^6 for the scale parameter μ and other regression coefficients $\beta_1, \beta_2, \beta_p$. The models are compared and the best fit model is chosen on the basis of Akaike Information Criterion (AIC).

The better fit among candidate models are performed through the Akaike information criterion (AIC)⁴⁻⁵ as

$$AIC = -2 \ln\{p(x|\hat{\theta})\} + 2k. \quad (5)$$

The number of parameters is represented by k . The random variable and maximum likelihood estimate are presented by x and $\hat{\theta}$. The minimal value of AIC shows better fit of the model. The Bayesian extension of the Cox proportional hazard model is presented as

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

$$(6)$$

where the parameter of interest is defined as θ . The term Y is the observed evidence and the marginal probability of Y is defined as $P(Y)$. The prior is $P(Y)$ and the likelihood function is $P(Y|\theta)$ ⁶. Mean, standard deviation, credible interval and the highest posterior density (HPD) are computed for each parameter. An alternative of the AIC in the context of Bayesian model selection method is Deviance Information Criteria (DIC). The Deviance Information Criteria (DIC) is defined as,

$$DIC = -2\ln\{p(x|\hat{\theta})\} + p_D \quad (7)$$

where,

$$p_D = E[-2\ln\{p(x|\hat{\theta})\}] + 2\ln\{p(x|\hat{\theta})\}. \quad (8)$$

As the number of independent parameters in a Bayesian hierarchical model is not clearly defined, DIC estimates the effective number of parameters by difference of the posterior mean of the deviance and deviance of posterior means.

2.3. Bayesian Cox PH regression separately for each event

The Cox proportional hazards model is applied in time-to-event data analysis⁷⁻⁹. It is defined as

$$\lambda_i(t|Z_i) = \lambda_0(t)\exp(Z_i\beta) \quad (9)$$

or,

$$\log \lambda_i(t|Z_i) = \log \lambda_0(t) + Z_i\beta ; \quad i = 1, 2, \dots, n. \quad (10)$$

The baseline hazard and hazard at time t is defined by $\lambda_0(t)$ and $\lambda_i(t|Z_i)$ for i^{th} patient. Z_i is the covariate for an i^{th} patient with the regression coefficient β . The hazard ratio is defined as a predicted hazard function under different predictor variables. The partial likelihood function is adopted to fit the Cox model. A high p-value for the coefficient is defined as less significance of the variable in interest. The better fit among candidate models is performed through the Akaike information criterion (AIC) as discussed. Similarly, DIC is used for model comparison while using Bayesian techniques.

Different Cox PH models with several factors like age, gender, node, etc are performed and posterior means of the parameters obtained through the models are provided in Table-1.

Table 1: Posterior Estimate generated through different models through Cox PH model

Bayesian Estimate						MLE
Response	Parameter	Posterior Mean (SD)	95% HPD	DIC	pD	
LRC	Arm	-0.31 (0.14)	(-0.61, -0.03)	2187.7	0.99	-0.31 (-0.60, -0.02)
	Age	-0.20 (0.17)	(-0.53, 0.13)			
	Gender	0.41 (0.23)	(-0.06, 0.86)			
PFS	Arm	-0.31 (0.13)	(-0.58, -0.05)	2614.46	0.99	-0.31 (-0.57, -0.04)
	Age	-0.38 (0.15)	(-0.64, -0.09)			
	Gender	0.49 (0.22)	(0.05, 0.91)			
OS	Arm	-0.16 (0.13)	(-0.42, 0.08)	2610.64	0.99	-0.16 (-0.42, 0.08)
	Age	-0.38 (0.15)	(-0.68, -0.09)			
	Gender	0.28 (0.20)	(-0.13, 0.66)			

Table 2: Posterior Estimates generated for different gap times through AFT model

Response	Parameter	Posterior Mean(SD)	95% HPD
PFS	Intercept	5.32 (0.109)	(5.10, 5.53)
	Arm	0.10 (0.15)	(-.18, 0.40)
	τ (shapeparameter) for weibull distribution	2.46 (0.41)	(1.71, 3.34)
OS	Intercept	4.66 (0.121)	(4.42, 4.89)
	Arm	0.14 (0.17)	(-0.19, 0.47)
	τ (shapeparameter) for weibull distribution	0.83 (0.09)	(0.65, 1.03)

3. Results

Dataset was presented to resemble a motivating example of head and neck cancer (HNC). A total of 148 patients treated with two chemotherapeutic arms were illustrated. The clinical trial was aimed to perform the progression-free survival (PFS) between two types of therapy. The therapies were (I) 'Arm-A' (n=43 subjects) or (II) 'Arm-B' (n=31 subjects). The covariates considered were (a) Arm, (b) Age and (c) Gender. Subjects were followed continuously and the occurrence of relapse, disease progression and death were monitored under the clinical trial protocol. For the sake of the analysis, we omitted the missing observations.

The duration between treatment initiation to the time of progression or the last follow-up visit for patients who had not progressed was considered. The progression is defined as RECIST criteria version 1.1. Disease-free survival was considered as the duration while the person experienced a complete remission. We considered the duration between LRC and progression as T_1 and between progression and death as T_2 .

One of the aims of the trial was to investigate the best effective arm to prolong the PFS. The trial was followed by the loco-regional recurrence and overall survival. In this example, we measured the LRC as the duration between dates of registration to the date of first loco-regional relapse. Similarly, the date of registration to date of progression is defined as PFS. The OS was defined as the last date of follow-up or date of death from the date of registration. The Cox proportional hazard model and accelerated failure time model were considered for different states in the context of Bayesian frameworks. The states were defined as a dead state (state 3), living with the progressed disease (state 2) and living with locoregional recurrence, not with distant metastasis/progression (state1). The direct transition from state 1 to state 3 is possible. However, as mentioned earlier, we considered only those patients for which all 3 states were apparent.

The Cox proportional hazard model applied in this dataset is defined as,

$$\lambda(t|x) = \lambda_0(t) \exp(\beta_1 * Arm + \beta_2 * Age + \beta_3 * Gender) \quad (11)$$

The 3 covariates considered for the modeling are Arm, Age, and Gender. The results are illustrated in Table 1. The survival curves corresponding to LRC and PFS are shown in Figure 1, Figure 2. The Kolmogorov-type supremum test is performed to obtain the p-value.

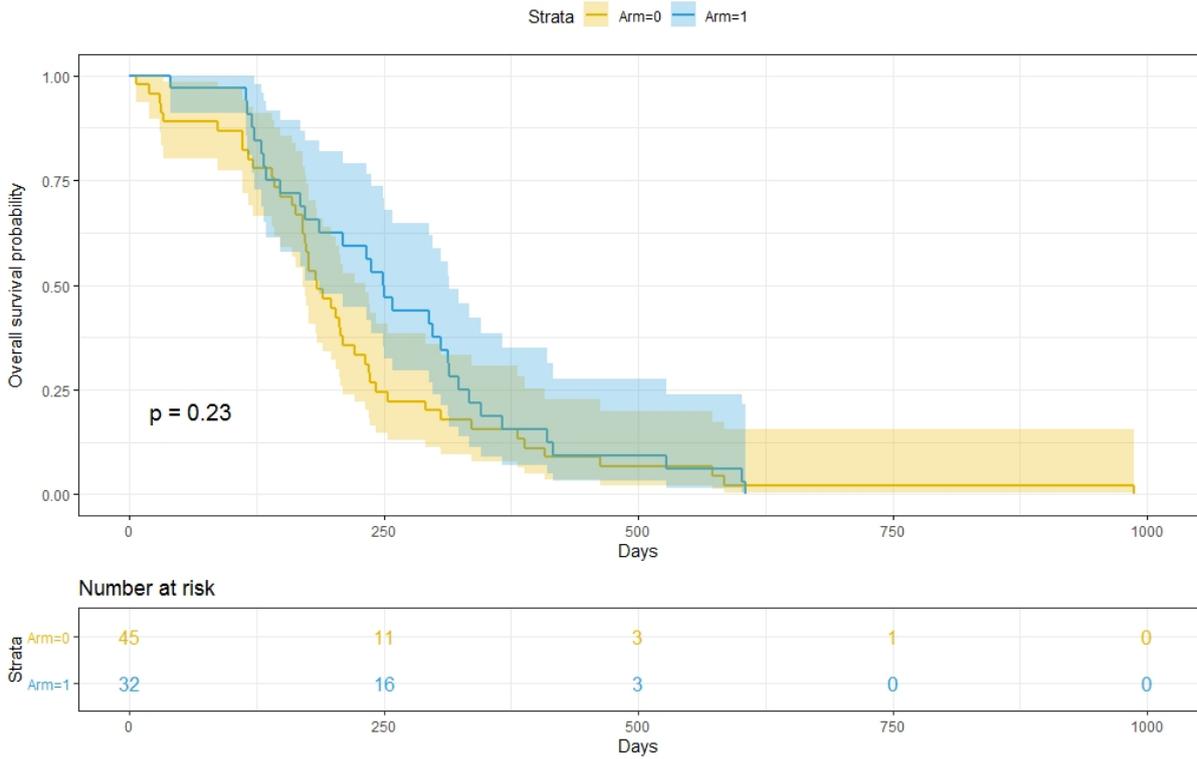


Fig. 1. Locoregional relapse progression

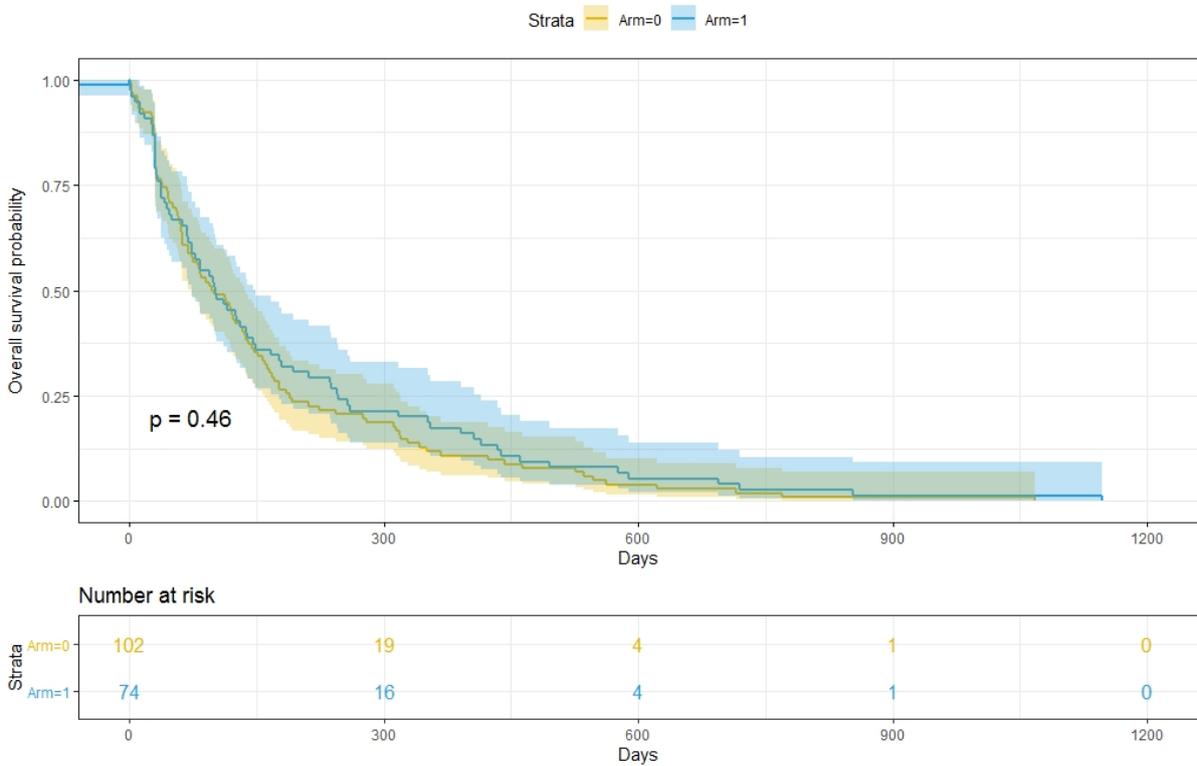


Fig. 2. Progression Free Survival

The AFT models are computed considering Arm as the only covariate. The model is,

$$Y = \log(G) = \mu + \beta * Arm + \epsilon \quad (12)$$

The posterior mean and standard deviation of the Arm effect is obtained from both the AFT and regression model and from the density plots of the difference of Arm effect from both the model shown in figure3.

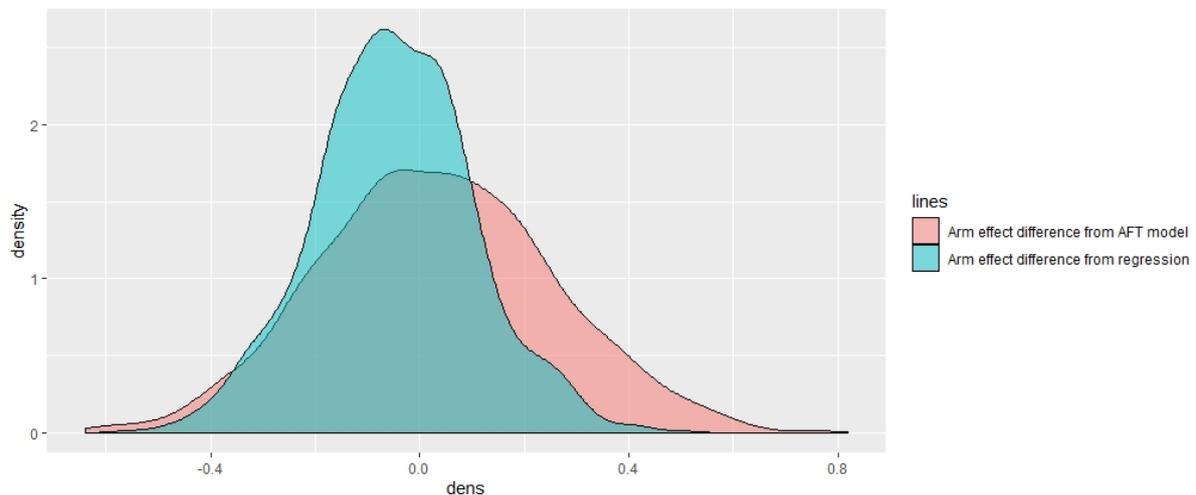


Fig. 3. Plot of Arm effect difference from AFT model and auto-regression model

We can draw this inference that the dependency of gap times is translated through the regression structure. So, adding the Arm effect from the AFT survival model for the first gap time and the Arm effect obtained from the regression model, we will obtain the same output for the AFT survival model for the second gap time. Thus, given the information of time between LRC and PFS and dependency structure between gap times, the survival duration between PFS and OS can be predicted. The details of the posterior means obtained using the Bayesian AFT model are given in table 2. The analysis was done using R opensource software¹⁰ and Bayesian computing was performed using OpenBUGS.

3.1. Discussion

The novel corona virus that causes COVID-19 appeared more than twice as high among individuals with cancer than the general population¹¹. In survival analysis of disease-related to oncology, the patients commonly experience multiple events like locoregional relapse, progression, death across the follow-up period. The interest lies in the prediction of survival duration for a particular event and evaluating effective treatments. The analysis is carried

out by assuming the independence of the events. However, due to a missed visit of the patients, information regarding the complete follow-ups of the patient is often unknown. So, their survival duration cannot be predicted based on the analysis carried out on the previously occurred events. The dependency modeling of the durations between consecutive events will assist to predict the occurrence of the next event.

The generalized version of the multi-state model is well-documented¹²⁻¹³. The simplest form is defined as mortality model having two states, 'alive without disease' and 'dead' and linked transition between these two states. The competing risk model can be defined as a provision where individual may die due to another causes¹⁴⁻¹⁶. The widely accepted form of the multi-state model is the illness-death model or disability model.

The associated package to work in these directions are 'mstate' is useful for multistate regression and to get prediction probability. Another package 'survdim' is useful to perform type-specific Cox models. The parametric multistate model can be performed through 'msm' and 'flexsurv'. This work is performed with open source software OpenBugs to serve the Bayesian.

4. Conclusions

The constant news about the corona virus is scary. It is not possible to separate treatment for Cancer due to COVID-19. An effective treatment comparison strategy is needed. We presented a handy tool to understand cancer progression in this unprecedented scenario. Linking different events of cancer progression is the need of the hour. It is a methodological challenge. We provide the solutions to overcome the issue with interval between two consecutive events in motivating head and neck cancer (HNC) data.

Abbreviations: AFT: accelerated failure time, AIC: akaike information criterion, DIC: deviance information criteria, HNC: head and neck cancer, HPD: highest posterior density, LRC: locoregional control, MCMC: markov chain monte carlo, OD: overall death, PFS: progression free survival, SD: standard deviation.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: Not applicable.

Competing interests: The authors declare that there are no financial and competing interests.

Funding: Authors are thankful to Science and Engineering Research Board, Department of Science & Technology, Government of India, for providing necessary support to carry out the present research work through project No. MSC/2020/000063 but not for APC.

Authors' contributions: AB planned the study, AB and SB performed the study, GKV prepare the manuscript. SS writes the methodological details to finalized the manuscript.

Acknowledgments: Authors are thankful to Dr. (Md.) Mojibur Rahman, Professor of English Language, Department of Humanities and Social Sciences, Indian Institute of Technology (ISM) Dhanbad, India for editing the English language and improving the grammar of this manuscript.

Author details: The details of all authors are given on the title page.

Reference

- [1]. H. Wang and L. Zhang, Risk of covid-19 for patients with cancer. *The Lancet Oncology*, 21(4), e181, 2020.
- [2]. H. Putter, J. van der Hage, G.H. de Bock, R. Elgalta, and C.J. Van de Velde, Estimation and prediction in a multistate model for breast cancer. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 48(3), 366-380, 2006.
- [3]. C.A. Perez, M. Bauer, S. Edelstein, B.W. Gillespie, and R. Birch, Impact of tumor control on survival in carcinoma of the lung treated with irradiation. *International Journal of Radiation Oncology Biology and Physics*, 12(4), 539-547, 1986.
- [4]. H. Akaike, A bayesian extension of the minimum aic procedure of autoregressive model fitting, *Biometrika*, 66(2), 237-242, 1979.
- [5]. H. Akaike et al., Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1), 3-14, 1981.
- [6]. M. Stokes, F. Chen, and F. Gunes, *An introduction to Bayesian analysis with sas*. Stat® Software, 2014.

- [7]. J. Kasza, D. Wraith, K. Lamb, and R. Wolfe, Survival analysis of time-to-event data in respiratory health research studies. *Respirology*, 19(4), 483-492, 2014.
- [8]. D.Y. Lin, L.-J. Wei, and Z. Ying, Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3), 557-572, 1993.
- [9]. B. George, S. Seals, and I. Aban, Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4), 686-694, 2014.
- [10]. R.C. Team et al., *R: A language and environment for statistical computing*. 2013.
- [11]. C. Wu, X. Chen, Y. Cai, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang et al., Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china. *JAMA internal medicine*, 2020.
- [12]. P.K. Andersen, O. Borgan, R.D. Gill, and N. Keiding, Statistical models based on counting processes, 2012.
- [13]. P. Hougaard, Analysis of multivariate survival data, 2012.
- [14]. A. Bhattacharjee, Bayesian competing risks model: an application to breast cancer clinical trial with incomplete observations. *Journal of Statistics and Management Systems*, 18(4), 381–404, 2015.
- [15]. P.K. Andersen, S.Z. Abildstrom, and S. Rosthøj, Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2), 203–215, 2002.
- [16]. H. Putter and C. Spitoni, Non-parametric estimation of transition probabilities in non-markov multi-state models: The landmark aalen–johansen estimator, *Statistical Methods in Medical Research*, 27(7), 2081–2092, 2018.