Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 16(2) 2023: 194-205
http://ejournal.undip.ac.id/index.php/media_statistika

AN ADDITIVE SUBDISTRIBUTION HAZARDS MODEL FOR COMPETING RISKS DATA

Molydah S, Danardono
Department of Mathematics, Universitas Gadjah Mada Yogyakarta, Indonesia

e-mail: *molydah@mail.ugm.ac.id*

DOI: 10.14710/medstat.16.2.194-205

Article Info:
Received: 26 December 2022
Accepted: 20 February 2024
Available Online: 11 March 2024

Keywords:
Additive Hazard Model; Competing Risk; Cumulative Incidence Function; Subdistribution Hazard

Abstract: Competing risk failure time data occur frequently in medical a number of methods have been proposed for the analysis of these data. The classic approach is to model all cause-specific hazards and then estimate the cumulative incidence curve based on these cause-specific hazards. Unfortunately, the cause-specific hazard function does not have a direct interpretation in terms of survival probabilities for the particular failure type. In this paper, we consider a more flexible model for the subdistribution. It is a combination of the additive model and the Cox model and allows one to perform a more detailed study of covariate effects. One advantage of this approach is that our regression modeling allows for non-proportional hazards. This leads to a new simple goodness-of-fit procedure for the proportional subdistribution hazards assumption that is very easy to use. We applied this method to melanoma data and estimated the cumulative death rate for those who died from melanoma after surgical removal of the tumor. It was found that two covariates had a time-varying effect and two other covariates had a constant effect in predicting the cumulative incidence curve in patients who died of melanoma following tumor removal surgery.

1. INTRODUCTION

Survival analysis is a statistical method where the outcome variable that is considered is the time until an event occurs or survival time. In survival data, the problem that often arises is the presence of censored observations. Censored observations occur when the survival time of the individual being observed is not known with certainty. Censored is the basic concept that distinguishes survival analysis from conventional statistical methods, causing univariate, bivariate, and multivariate analysis to be invalid for analyzing survival data, so special statistical methods are needed, one of which is the Cox regression model, also known as the Cox Proportional Hazard model. In survival analysis the most popular regression method used is the semiparametric regression method, this is because in semiparametric regression it does not require assumptions about the survival time distribution, but the results of the parameter estimates are close to the parametric regression method. The semiparametric regression method that is often used in survival analysis is the Cox regression model. The application of the Cox model for some situations is sometimes inappropriate, one of which is when competing risks occur. In general,

competing risks arise when an individual can experience more than one type of event, and the occurrence of these events precludes the occurrence of other types of events. Often, the timing of an event in competing risk is influenced by one or several independent variables (covariate).

The Cumulative Event Curve (CIF) is the probability that a certain type of event occurs at or before a certain point in time and is an appropriate summary curve in analyzing competing risk. The Kaplan-Meier (KM) method has become a widely used tool for estimating survival functions and cumulative occurrence functions. This method is conceptually easy to understand and easy to compute, however if there is more than one type of event (or failure), and if these events are dependent, the KM method is biased. This bias arises because the KM method assumes that all events are independent, therefore the KM method is not appropriate for estimating the cumulative event curve.

In biomedical studies it is important to study the effect of covariate on the cumulative occurrence function of a particular failure. The standard approach is to model a cause specific hazard for all causes. The cox proportional hazard model is the most commonly used regression model for all causes. This approach is valid when all cause specific hazards are modeled correctly, because the cumulative incidence curve for a particular cause is a function of all cause specific hazards. However, this method has the disadvantage that it is difficult to identify which specific covariate have a time-varying effect on the cumulative incidence curve. Overcoming this problem, Fine & Gray (1999) proposed a model based on subdistribution that can estimate the effect of covariate on the cumulative incidence curve. In this study, the authors focus on a more flexible and general model for the hazard subdistribution function in estimating the effect of covariate on CIF. This model is a combination of the additive model and the Cox model, making it possible to carry out a more detailed study of covariate effects. The method will be applied to the censored competing risks data. Sun & Liu (2006) extended the Cox proportional model to a more general additive hazard model for hazard subdistributions with independent time covariates and used the IPCW technique for censored data. In addition, the advantages of the model allow for non-proportional hazard in this case leading to the goodness-of-fit test procedure.

2. LITERATURE REVIEW

2.1. Survival Analysis

Survival analysis is one of the statistical methods used to analyze data where the variable considered is the time until an event occurs. Time can be expressed in units of days, weeks, months, or even years, which are obtained from the initial observation of an individual until an event occurs for that individual (Kleinbaum and Klein, 2012). The purpose of survival analysis is to determine the relationship between the time of occurrence and the independent variables measured at the time of the study. In addition, it is also used to identify the factors that significantly influence an event. In survival data, one of the problems that arise is incomplete observations, which are generally grouped into censored data and truncated data. Censored data is data that cannot be observed in its entirety because the research subject is missing or for other reasons so that the data cannot be retrieved, or until the end of the study, the subject has not experienced a certain event. According to Collet (2003), censored data types are divided into three types, namely:

1. The left sensor is a sensor that is carried out when the initial time of the observation subject is not observed, but the full failure time event can be observed before the research study ends.

2. Right censorship occurs when the subjects included in the observation can be observed in full but, until the end of the study, had not experienced an incident.
3. An interval sensor is a sensor whose survival time is within an interval of certain.

Another problem with survival data is that the truncated data is divided into left- and right-truncated data. Left-truncated data occurs when the individual's incident time is less than the left-truncated time, while right-truncated data occurs when the individual's incident time is more than the right-truncated time. Individuals who experienced a left or right slit were not included in the observation.

2.2. Survival Function and Hazard Function

In survival analysis, there are two basic functions, namely the survival function and the hazard function. The survival function is a basic function used to describe the phenomenon of its occurrence. The survival function is denoted by $\boldsymbol{S(t)}$, which is the opportunity for an individual to survive longer than time $\boldsymbol{t}$ (Kleinbaum & Klein, 2012). The survival function is a probability, so its value is always in the interval [0, 1]. The survival function is defined as follows:

$$S(t) = P(T > t) \tag{1}$$

If T is a continuous random variable, then the survival function is the complement of the cumulative distribution function, where the cumulative distribution is defined as $F(t) = P \leq t)$, so the survival function can be written as follows:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \tag{2}$$

Furthermore, the value of $S(t)$can be obtained through the probability density function $f(t)$ as follows:

$$S(t) = P(T > t) = \int_0^\infty f(x)dx \tag{3}$$

Theoretically the survival function is a non-increasing function with respect to time *t* with the following characteristics:

1. The survival function is a decreasing monotone function.
2. For $t = 0$ then $S(t) = S(0) = 1$, meaning that the chance for a research unit to survive at $t = 0$ is 1.
3. For $t \to \infty$ maka $\lim_{x\to\infty} S(t) = 0$, shows that as time goes by, the chance for a research unit to survive will get smaller, so that if t is very large, then the chance for a research unit to survive will be close to zero.

An important measure in survival analysis besides the survival function is the hazard function. The hazard function often referred to as the hazard rate, is denoted by $h(t)$. The hazard function can be defined as the rate of occurrence of an event if it is known that an object of research survives until time t. Systematically, it can be written as follows:

$$h(t) = \lim_{\Delta t\to 0} \frac{P(t \leq T < (t + \Delta t)|T \geq t)}{\Delta t} \tag{4}$$

2.3. Estimation of Survival Function and Hazard Function

The Kaplan-Meier estimator or Product Limit Estimator is a nonparametric estimator that is often used to estimate survival functions. The Kaplan-Meier estimator is given as follows (Klein & Moeschberger, 2005):

$$\begin{cases} 1 & , \quad t \le t_i \\ \prod_{t_i \le t} (1 - \frac{d_i}{Y_i}) \, ; t_i \le t & , \quad t_i \le t \end{cases} \tag{5}$$

where d_i is the number of individuals who experience the event at time t_i and Y_i is the number of individuals at risk at time t_i where t_i is the survival time observed in object, for $i = 1, 2, \ldots, k$. The Kaplan-Meier estimator is a function of the ladder that goes down when there is an event.

To estimate the survival function using the Kaplan-Meier estimator $S(t)$, first calculate the standard error or variance of the survival function. The variance of the Kaplan-Meier estimator can be found using the Greenwoods formula (Klein & Moeschberger, 2005):

$$Var\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)} \tag{6}$$

or you can use the following formula as an alternative:

$$Var\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{1 - \hat{S}(t)}{Y(t)} \tag{7}$$

2.4. Competing Risk Model

Competing risk is a situation where an individual may experience more than one type of event, and the occurrence of one type of event precludes the occurrence of another. If failure were the different causes of death, then only death from the first cause would be observed and considered an event of interest, whereas death from any other cause would be considered a competing risk. The existence of competitive risks implies that the usual methods of survival must be applied with caution, and the consequences of their use must be recognized. One method that is often used to estimate the probability of survival over a certain period of time is the Kaplan-Meier method.

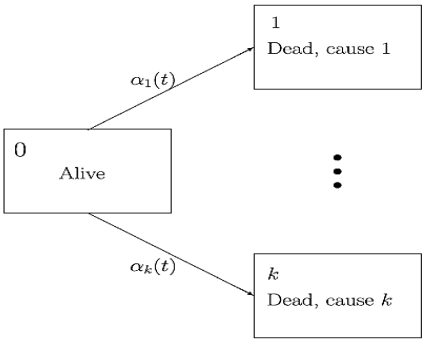

Figure 1. Figure Illustrating the Competing Risks Model

In the case of competing risks, the Kaplan-Meier method considers failures from other causes or competing risks as censored observations, but this causes bias. Kalbfleisch and Prentice (1980) suggest an approach that can be used in calculating competing risk, namely the cumulative event function (CIF), by using this technique, the probability of each event that occurs is partitioned into the probability for each type of event. In general, the standard approach that is often used in estimating and modeling the cumulative event function is by estimating and modeling the cause specific hazard $k(t) for\ k = 1, \ldots, K$ cause as shown in Figure 1.

2.5. Additive Subdistribution Hazard Model

In the case of competing risks, assuming two types of failures k $(k = 1, 2, \dots)$ the cumulative incidence function for cause 1 given a set of covariates x is given by

$$F_1(t; \boldsymbol{X}, \boldsymbol{Z}) = P(T^0 \leq t, \varepsilon = 1 | \boldsymbol{X}, \boldsymbol{Z})$$

With X and Z as covariates. Cumulative event function inference for other types of failures can be done in a similar way. To estimate F_1, Fine & Gray (1999) proposed a model of the hazard function F_1, as a substitute for the cause-specific hazard function. The advantage of this model is that it can estimate F_1 directly without simultaneously estimating subdistributions corresponding to other failure types, defined as follows:

$$\lambda_1(t; \boldsymbol{X}, \boldsymbol{Z}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{(t \leq T^0 \leq t + \Delta t, \varepsilon = 1 | T^0 \leq t \cap \varepsilon \neq 1), \boldsymbol{X}, \boldsymbol{Z}\}$$
$$= -\frac{d}{dt} log\{1 - F_1(t; \boldsymbol{X}, \boldsymbol{Z})\}$$

with $F_1(t; \boldsymbol{X}, \boldsymbol{Z}) = 1 - \exp\{-\Lambda_1(t; \boldsymbol{X}, \boldsymbol{Z})\}$

where $\Lambda_1(t; \boldsymbol{X}, \boldsymbol{Z}) = \int_0^t \boldsymbol{\lambda_1}(u; \boldsymbol{X}, \boldsymbol{Z}) du$. In the following, it assumes that λ_1 has the form

$$\lambda_1(t; \boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{\alpha^T}(t)\boldsymbol{X} + \lambda_{10}(t)\exp(\boldsymbol{\beta_0^T} Z) \quad (8)$$

where $\boldsymbol{\alpha}(t)$ is the q unknown dimension vector of the time-dependent component representing the covariate effect $\boldsymbol{X}$, where as β_0 is the dimension vector p of the unknown regression parameter showing the covariate effect $\boldsymbol{Z}$, and λ_{10} is the baseline hazard function that is undefined.

2.6. Regression Parameter Estimation

Let $T_i, C_i, \varepsilon_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\}$, $i = 1, \dots, n$ be n be n independent, and $\{T_i, \delta_i, \delta_i \varepsilon_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\}$ is the observed data. It is assumed that for simplification the censoring variable C is independent of the survival time of T, and the covariate $\boldsymbol{X}$ and the covariate $\boldsymbol{Z}$. Define $N_i(t) = I(T_i^0 \leq t, \varepsilon = 1)$, $Y_i(t) = 1 - N_i(t)$, and $G(t) = P\{C \leq t\}$. Define $\hat{G}(t)$ shows the Kaplan-Meier estimator of $G(t)$.

It should be noted that $N_i(t)$ and $Y_i(t)$ are usually not fully observed when censorship is present and the risk indicator is equal to 1 as long as no type 1 event occurs. Next is defined

$$r_i(t) = I(C_i \geq T \wedge t)$$

$r_i(t) = 1$ then $N_i(t)$ and $Y_i(t)$ can be counted up to time t, and if $r_i(t) = 0$ then individuals are observed until time C_i and for $N_i(t)$ and $Y_i(t)$ are not observed. Next, we define the time-dependent weight function

$$R_i(t) = r_i(t)\hat{G}(t)/\hat{G}(T_i \wedge t)$$

Let $dN_i^*(t) = R_i(t)\, dN_i(t)$ and $Y_i^* = R_i(t)\, Y_i(t)$, with $\boldsymbol{N}^* = (N_1^*, \dots, N_n^*)^T$,

$\boldsymbol{X}^* = (Y_1^* \boldsymbol{X}_1, \dots, Y_n^* \boldsymbol{X}_n)^T, \boldsymbol{Z}^* = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n)^T, \phi_i = \phi_i(\boldsymbol{\beta}) = Y_i^* \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$

$\phi = \phi(\boldsymbol{\beta}) = (\phi_1, \dots, \phi_n)^T, \boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{\beta}) = diag(\phi_i), \boldsymbol{W} = diag(w_i)$

$\boldsymbol{V} = diag(v_i)$, with $\boldsymbol{w} = (w_1, \dots, w_n)^T$ and $\boldsymbol{v} = (v_1, \dots, v_n)^T$ as a weight function.

To estimate the parameter c$\beta_0, A(t) = \int_0^t \alpha(u) du$, and $\Lambda_{10}(t) = \int_0^t \lambda_{10}(u) du$ with Inverse Probability of Censoring Weighting (IPWC) technique by using the following score function:

$$\int_0^t \boldsymbol{Z}^{*T}\boldsymbol{\Phi}\mathbf{V}\{\mathrm{d}\mathbf{N}^* - \boldsymbol{X}^* dA - \phi d\Lambda_{10}\} = 0$$
$$\boldsymbol{X}^*\boldsymbol{W}\{d\boldsymbol{N}^* - \boldsymbol{X}^* dA - \phi d\Lambda_{10}\} = 0$$
$$\phi^T\boldsymbol{W}\{d\boldsymbol{N}^* - \boldsymbol{X}^* dA - \phi d\Lambda_{10}\} = 0$$

where τ is a prespecified constant such that $P\{T \geq \tau\} > 0$. For a given $\boldsymbol{\beta}$, solving the second and third score equations gives a weighted Aalen estimator

$$\hat{A}(t;\widehat{\boldsymbol{\beta}}) = \int_0^t (\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}d\boldsymbol{N}^*(u) \tag{9}$$

for A, and a Breslow estimator

$$\widehat{\Lambda}_{10}(t;\boldsymbol{\beta}) = \int_0^t (\phi^T\boldsymbol{W}\phi)^{-1}\phi^T\boldsymbol{W}\boldsymbol{H}d\boldsymbol{N}^*(u) \tag{10}$$

for Λ_{10}, where $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Q}$ and $\boldsymbol{Q} = 1 - \phi(\phi^t\boldsymbol{\beta}\boldsymbol{W}\phi)^{-1}\phi^T\boldsymbol{\beta}\boldsymbol{W}$**.** By plugging Equation (9) and (10) into the first score equation for $\boldsymbol{\beta}_0$ is obtained,

$$U(\boldsymbol{\beta};\tau) = \int_0^\tau \boldsymbol{Z}^{*T}\boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{H}d\boldsymbol{N}^*(t) \tag{11}$$

By letting $w_i = hv_i$ for function h, $U(\beta;\tau)$ is reduced to Cox-like scoring equation.

$$U(\boldsymbol{\beta};\tau) = \sum_{i=1}^n \int_0^\tau \left\{\boldsymbol{Z}_i - \frac{\sum_{j=1}^n w_j\phi_j\boldsymbol{Z}_j Y_j \exp(\beta^T\boldsymbol{Z}_j}{\sum_{j=1}^n w_j\phi_j Y_j \exp(\beta^T\boldsymbol{Z}_j}\right\} v_i\phi_i d\widehat{N}_i(t) = 0$$

where $d\widetilde{\boldsymbol{N}} = \boldsymbol{H}d\boldsymbol{N}^*$. Let $\widehat{\boldsymbol{\beta}}$ show the solution of Equation (11), after $\widehat{\boldsymbol{\beta}}$ is obtained, the researcher can estimate the cumulative function baseline hazard of $\widehat{\Lambda}_{10}$ and the additive component $A(t)$ using the Brelow estimator $\widehat{\Lambda}_{10}(t) = \widehat{\Lambda}_{10}(t;\widehat{\boldsymbol{\beta}})$ and the Aalen estimator $\hat{A}(t) = \hat{A}(t,\beta)$. If there is only one cause of failure, then Equations (9) - (11) are reduced to a score equation (Martinussen & Scheike, 2002).

To investigate asymptotic nature of $\widehat{\boldsymbol{\beta}}$, $\widehat{\Lambda}_{10}(t)$ and $\hat{A}(t)$, Suppose $\hat{\phi}, \widehat{\Phi}, \widehat{\boldsymbol{Q}}$, and $\widehat{\boldsymbol{H}}$ be defined as ϕ,$\boldsymbol{\Phi}$,$\mathbf{Q}$ and $\boldsymbol{H}$ with $\boldsymbol{\beta}$ replaced by $\widehat{\boldsymbol{\beta}}$. $n^{1/2}U(\boldsymbol{\beta}_0;\tau)$ has an asymptotic normal distribution with mean zero and a covariance matrix that can be consistently estimated by

$$\hat{\Sigma}_U = n^{-1}\sum_{i=1}^n \widehat{\Psi}_{1i}(\tau)\widehat{\Psi}_{1i}^T(\tau) \tag{12}$$

Therefore, $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})$ has an asymptotic normal distribution with mean zero and covariance matrix that can be consistently estimated by

$$\hat{\Sigma}_{\boldsymbol{\beta}} = n^{-1}I^{-1}(\widehat{\boldsymbol{\beta}})\sum_{i=1}^n \widehat{\Psi}_{1i}(\tau)\widehat{\Psi}_{1i}^T(\tau)I^{-1}(\widehat{\boldsymbol{\beta}})^T \tag{13}$$

Next, it will be shown that $n^{1/2}\{\hat{A}_{10}(t) - A_{10}(t)\}$ converges to a zero-mean Gaussian process for which the covariance function at (t_1, t_2) can be estimated consistently with

$$\hat{\sigma}_A(t_1,t_2) = n^{-1}\sum_{i=1}^n \widehat{\Psi}_{2i}(t_1)\widehat{\Psi}_{2i}^T(t_2) \tag{14}$$

Similarly, $n^{1/2}\{\widehat{\Lambda}_{10}(t) - \Lambda_{10}(t)\}$ converges weakly to a zero-mean Gaussian process whose covariance function at (t_1, t_2) can be estimated consistently with

$$\hat{\sigma}_{\Lambda_{10}}(t_1, t_2) = n^{-1} \sum_{i=1}^{n} \hat{\Psi}_{3i}(t_1) \hat{\Psi}_{3i}^{T}(t_2) \tag{15}$$

2.7. Prediction of Cumulative Incidence Functions

One of the main goals in survival analysis is to predict certain survival probabilities for future subjects. To predict $\mathcal{F}_1$ under (8) for a patient with a set of covariates $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Z} = \boldsymbol{z}$, one can first estimate the cumulative distribution hazard $\Lambda_1(t; \boldsymbol{x}, \boldsymbol{z})$ by

$$\hat{\Lambda}_1(t; x, z) = \int_0^t x^T(u) d\hat{A}(u) + \int_0^t \exp\left(\hat{\beta}^T z(u)\right) d\hat{\Lambda}_{10}(u) \tag{16}$$

The predicted cumulative incidence is then given by $\hat{\mathcal{F}}_1(t; \boldsymbol{x}, \boldsymbol{z}) = 1 - \exp\{-\hat{\Lambda}_1(t; x, z)\}$. Furthermore, an estimate $\hat{t}_p$ for the 100pth percentile t_p of $\mathcal{F}_1(t|\boldsymbol{x}, \boldsymbol{z})$ can be obtained by solving the equation $\hat{\mathcal{F}}_1(t; \boldsymbol{x}, \boldsymbol{z}) = 1 - p$, where $0 < p < 1$ is such that $t_p < \tau$. Using the functional δ-method, one can show that for a known, monotone, absolutely continuous transformation $g, n^{1/2}\left\{g(\hat{F}_1(t; x, z)\right) - g(F_1(t; x, z))\}$ converges weakly to a zero-mean Gaussian process whose covariance function at t_1, t_2 can be consistently estimated by

$$\begin{aligned} \hat{\sigma}_F(t_1, t_2) = n^{-1}\, & g((\hat{F}_1(t_1; x, z)) g(\hat{F}_1(t_2; x, z))(1 - \hat{F}_1(t_1; x, z)) \\ & \times (1 - \hat{F}_1(t_1; x, z)) \sum_{i=1}^{n} \hat{\Psi}_{4i}(t_1) \hat{\Psi}_{4i}^{T}(t_2) \end{aligned} \tag{17}$$

The above transformation g is usually chosen to stabilize the variance and to ensure that pointwise and simultaneous confidence intervals for the probability $F_1(t; x, z)$ are bounded between 0 and 1. One commonly used choice is $g = log(-log)$. Sometimes one is interested in constructing confidence bands for (t), $\Lambda_{10}(t)$, $F_1(t; x, z)$ or t_p. This may be analytically difficult since the limiting Gaussian processes for $n^{1/2}\{\hat{A}(t) - A(t)\}$, $n^{1/2}\{\hat{\Lambda}_{10}(t) - \Lambda_{10}(t)\}$, and $n^{1/2}\{g(\hat{F}_1(t; x, z) - F_1(t; x, z)\}$ do not have independent increments. To this end, we propose to use the following simulation approach to approximate these limiting distributions as in Lin, et al. (1994) and Scheike & Zhang (2003). Let be $\{G_i; i = 1, \dots, n\}$ be a simple random sample of size n from the standard normal distribution and independent of the observed data. Then one can construct the simultaneous confidence bands for (t), $\Lambda_{10}(t)$, $F_1(t; x, z)$ or t_p or tp by replacing $\hat{M}_i(t)$ and $\hat{M}_i^c(t)$ with $G_i \hat{M}_i(t)$ and $G_i \hat{M}_i^c(t)$, respectively, and repeatedly generating normal random samples $\{G_i; i = 1, \dots, n\}$ given the observed data. Note that since $\Lambda_{10}(t)$ is nonnegative, one may want to use the log transformation for the construction of its confidence bands.

2.8. Test for Model Identification

This section considers the goodness of fit test of the model and the of time-varying covariates. For these, we develop some asymptotically procedures. To evaluate the goodness of fit of the covariates included in the multiplicative part of the model, following Lin, et al. (1993) and Wei (1984), consider the cumulative score processes. The observed score process is given by $n^{-1/2} U\left(\hat{\beta}; t\right)$, t), and its asymptotic distribution is equivalent distribution of

$$n^{-1/2} \sum_{i=1}^{n} \left\{\hat{\Psi}_{1i}(t) - I(\hat{\beta}, t) I^{-1}(\hat{\beta}, \tau) \hat{\Psi}_{1i}(\tau)\right\}$$

where $I(\beta, t)$ is the minus of the derivative of $n^{-1/2}U(\beta; t)$. Note that if multiplicative part of the model is appropriate, the components of the score process should behave as under the null. This suggests that we can use the following test statistics

$$\mathcal{F}_1 = \sup_{0 \le t \le \tau} \left| n^{-1/2} U_j(\hat{\beta}; t) \right|, \quad (j = 1, \dots, p) \tag{18}$$

where $U_j(\beta; t)$ denotes the jth component of $U(\beta; t)$. The percentiles of this test statistic can be estimated empirically using a number of simulated processes as discussed in the previous section, or in Lin, et al. (1993). Now consider testing if covariate j, included in the additive part of the model, is significant. For this, we suggest the test statistic

$$\mathcal{F}_2 = \sup_{0 \le t \le \tau} \left| \frac{\hat{A}_j(t)}{\hat{\sigma}^2_{A,j}(t)} \right| \tag{19}$$

where $\hat{A}_j$ is the jth component of $\hat{A}$ and $\hat{\sigma}^2_{A,j}(t)$ is the estimate of the variance of $\hat{A}_j(t)$. Sometimes one may also be interested in testing if an additive component has indeed a time-varying effect. To this end, we propose the test statistic

$$\mathcal{F}_3 = \sup_{0 \le t \le \tau} \left| \hat{A}_j(t) - \frac{\hat{A}_j(\tau)}{\tau} t \right| \tag{20}$$

Note that $\mathcal{F}_2$evaluates the departure of $\hat{A}_j(t)$ from the null, while $\mathcal{F}_3$ measures the departure between $\hat{A}_j(t)/t$ and the estimate of the constant effect under the null, $\hat{A}_j(\tau)/\tau$. Also note that the asymptotic distribution of $n^{1/2}\{\hat{A}(t) - A(t)\}$ is equivalent to the asymptotic distribution of $n^{-1/2}\sum_{i=1}^{n} \widehat{\Psi}_{2i}(t)$, where $\widehat{\Psi}_{2i}(t)$is defined in (7). Then the percentiles of the above two test statistics can be simulated as before. The proposed tests are simple to implement and are omnibus. Additionally, one can plot the estimated cumulative regression function and use the plots to visually examine whether a covariate has a time-varying effect on the cumulative incidence function.

3. MATERIAL AND METHOD

3.1. Data

The data used in this case study is secondary data entitled Melanoma obtained from the MASS package in program R. Melanoma data consists of measurements performed on patients with malignant melanoma, a type of skin cancer. Each patient underwent surgical removal of the tumor at the Department of Plastic Surgery, Odense University Hospital, Denmark during the period 1962 to 1977. The surgery consisted of the complete removal of the tumor together with approximately 2.5 cm of the surrounding skin. Among the measurements taken were the thickness of the tumor and see whether there was ulceration or not. This measure is considered an important prognostic variable in patients with thick tumors and/or an increased likelihood of death from melanoma.

3.2. Variables and Methods

The data used in this case study is secondary data entitled Melanoma Malignant obtained from the MASS package in R 4.0.5 software program. The Melanoma Malignant data consists of measurements taken on patients with Melanoma Malignant, a type of skin cancer. Each patient underwent surgical removal of the tumor at the Department of Plastic Surgery, Odense University Hospital, Denmark during the period 1962 to 1977. The surgery consisted of the complete removal of the tumor together with approximately 2.5 cm of the surrounding skin. Among the measurements taken were the thickness of the tumor and seeing whether there was ulcer or not. This measure is considered an important

prognostic variable in patients with thick tumors and/or an increased likelihood of death from Melanoma.

The purpose of the analysis of this case is to assess the influence of risk factors on survival. The risk factors are gender and age of the patient and histological variables, namely thickness and ulcer.

In the case of Melanoma patients, there are three possible events, namely patients who died from Melanoma, patients who died not because of Melanoma, and patients who were still alive until the end of the study. Each patient will only experience one event out of three possibilities. The incident of concern is the patient who died of Melanoma. If the patient dies not due to Melanoma then the patient is considered a competing risk individual. Each patient is represented by one row in the data set. The following variables are used in the data, including:

a. Dependent variable is a time variable (in days), which is the duration of time the patient survives during surgery until he dies or is censored, and an event variable which is the type of event experienced by the patient.
b. Independent variable (covariate) is a variable that is thought to influence the response variable, including the variables sex, age, thickness, and ulcer. The sex variable is the sex variable of the patient. Age variable (in years), is the patient's age variable at the time of surgery. Thick variable, is the patient's tumor thickness variable (in mm) at the time of surgery. The ulcer variable is an ulcer indicator variable.

Table 1 is a description of the data for categorical variables.

Table 1. Table of Frequencies and Coding of Categorical Variables

Variables	Category	Code	Total
Event	Censored	0	134
	Died of Melanoma	1	57
	Died not because of Melanoma	2	14
Sex	Male	0	79
	Female	1	126
Ulcer	No Ulcer	0	90
	Ulcer	1	115

Altogether 205 patients with four risk factors will be entered into the regression model to see the effect of covariates and predict the likelihood of the cumulative incidence of Melanoma Malignant patients. The data will be processed using the available packages in the `timereg` and `cmprsk` packages in R 4.0.5 software. In this case, the researchers only focused on cause 1, namely patients who died after undergoing surgery to remove a tumor due to melanoma.

4. RESULTS AND DISCUSSION

Descriptive analysis is used to describe the characteristics based on factors that are thought to influence the cumulative event function of Melanoma patients at Odense University Hospital, Denmark in 1962-1977. Table 2 is the result of a descriptive analysis of continuous data. Table 2 provides information that the average length of time Melanoma patients are treated at Odense University Denmark after undergoing surgery to remove the tumor is 2153 days. The average thickness of Melanoma patients is 2.91 cm, and the average age of Melanoma patients is 52 years.

Table 2. Descriptive Analysis Results

Variable	N	Minimum	Maximum	Mean
Time	205	10	5565	2153
Thick	205	0.10	17.42	2.91
Age	205	4	95	52

After being analyzed using the additive model, the following results are obtained:

Table 3. Results of Analysis with the Additive Model

Variable	Supremum test	P-value
Ulcer	3.73	0.004
Age	1.30	0.743
Sex	2.01	0.338
Thick	3.61	0.003

Based on the results of the analysis in table 3 above, by testing the hypothesis for each independent variable, it can be concluded that the sex and age variables have a constant effect, while the thick and ulcer variables have a time-varying effect on the cumulative incidence function in patients who died from Melanoma after surgical removal of the tumor.

Next, additive subdistribution hazard modeling will be carried out based on the previous analysis, along with the results of the analysis:

Table 4. Additive Subdistribution Hazard Model

Model	Variable	Coefficient	P-value
Parametric	Sex	-0.010	0.077
	Age	0.001	0.817
Nonparametric	Thick	3.610	0.003
	Ulcer	3.660	0.004

Table 4. above shows that the age and sex variables have a constant effect on the cumulative event function, so they are included in the multiplicative model section. As for the thick and ulcer variables, they have a time-varying effect on the cumulative incidence function in patients who died of Melanoma Malignant after surgical removal of the tumor. After being analyzed using additive and multiplicative models, the following results are presented in Table 5.

Table 5. Analysis Results with Additive and Multiplicative Models

Model	Variable	Coefficient	P-value
Multiplicative	Sex	-0.672	0.017
	Age	0.530	0.260
Addtive	Thick	3.61	0.003
	Ulcer	3.66	0.004

Based on the results of the analysis in Table 4.2 above, by testing the hypothesis for each independent variable, based on the multiplicative model, it can be concluded that the sex variable has a significant effect on the cumulative incidence function of patients who died of melanoma, while the age variable has no significant effect on cumulative incidence function. As for the other two variables, namely thick and ulcer variables, based on the additive model it was concluded that both had a significant effect on the cumulative incidence function in patients who died of Melanoma after surgery after tumor removal.

To further check the suitability of the model, an analysis was performed by comparing male and female patients who had ulcers and those who did not have ulcers.

The data used is the average age of Melanoma patients and the average thickness of Melanoma patients. The prediction results based on the additive subdistribution hazard model can be seen in the Figure 1(a).

Based on the Figure 1(a), it can be concluded that the probability or risk of Melanoma Malignant patients dying after surgery to remove the tumor for male patients who have ulcers is 0.59, while for male patients who do not have ulcers it is 0.29.

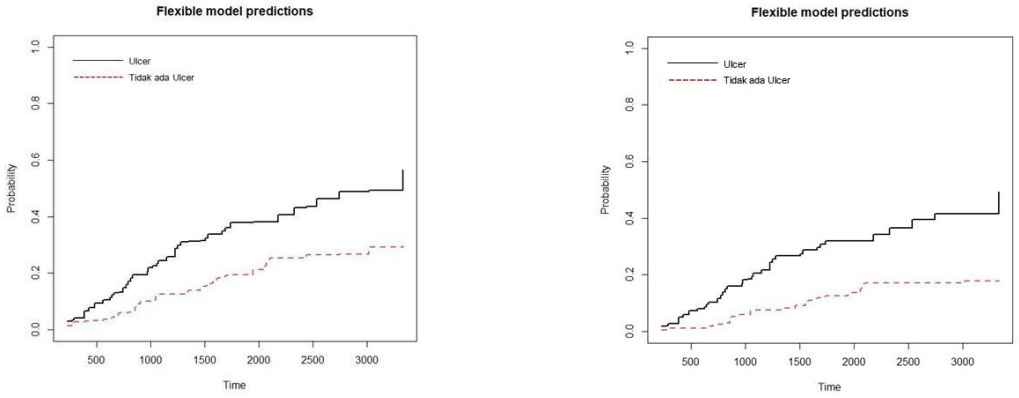

Figure 1. (a) Estimation of Cumulative Function in Male Patients
(b) Estimation of Cumulative Function in Female Patients

Based on the Figure 1(b), it can be concluded that the probability or risk of Melanoma Malignant patients dying after surgery to remove the tumor for female patients who have ulcers is 0.49, while for female patients who do not have ulcers it is 0.18. Based on the two pictures, it can be concluded that Melanoma Malignant patients either have ulcers or do not have ulcers, indicating that male patients have a greater risk of dying after surgical removal of the tumor compared to female Melanoma Malignant patients.

5. CONCLUSION

The additive subdistribution hazard regression model is a flexible model for regression analysis of competing risk failure time data. In applying additive subdistribution hazard regression analysis to melanoma patient data, it was found that the thick and ulcer variables had a time-varying effect on the cumulative incidence function of patients who died of melanoma, while the sex and age variables were not constant on the cumulative incidence function of patients who died because of melanoma after perform tumor removal surgery.

REFERENCES

Allison, P. D. (2010). *Survival Analysis*. United Kingdom: Emerald Group Publishing Ltd.

Cox, D. R. (1972). Regression Models and Life-Table. *Journal of the Royal Statistical Society: Series B (Methodologi)*, *2*(34), 187–220.

Collet, D. (2003). *Modelling Survival Data in Medical Research* (Second Edition). London: Chapman dan Hall/CRS.

Danardono, D. (2012), *Analisis Data Survival*. Yogyakarta: Program Studi Statistika, Universitas Gadjah Mada,

de Wreede, L. C., Fiocco, M., & Putter, H. (2010). The mstate Package for Estimation and Prediction in Non and Semi-parametric Multi-state and Competing Risks Models. *Journal of Computer Methods and Programs in Biomedicine*, *99*, 261–274.

Fauzani N. A. A. (2020). Pemodelan Risiko Bersaing dengan Kovariat Bergantung Waktu Menggunakan Hazard Subdistribusi. *Theses*. Yogyakarta: Magister Matematika, Universitas Gadjah Mada.

Fine, J. P. & Gray. R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association, 446*(94), 496–509.

Gray, R. J. (1988). A Class of K-sample Test for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*, *16*(3), 1141–1154.

Haesook, K.T. (2007). Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis. *Journal of Clinical Cancer Research*, *13*(2), 559–565.

Lin, D. Y., Wei, L. J., & Ying, Z. (1993). Checking the Cox Model with Cumulative Sums Martingale-Based Residuals. *Biometrika*, *80*(3), 557–572.

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data* (Second Edition). New York: Wiley.

Kleinbaum, D. G. & Klein, M. (2015). *Survival Analysis A Self Learning Text* (Second Edition), New York: Springer.

Klein, P. & Moeschberger, L. (2005). *Survival Analysis: Techniques for Censored and Truncated Data (Second Edition)*. New York: Springer.

Kleinbaum, D. G. & Klein, M. (2015). *Survival Analysis A Self Learning Text* (Third Edition). New York: Springer.

Lee. E. T. & Wang. J. W. (2003). *Statistical Methods for Survival Data Analysis* (Third Edition). New York: Wiley.

Martinussen, T. & Scheike, T. H. (2002). A Flexible Additive Multiplicative Hazard Model. *Biometrika*, *89*(2), 283–298.

Nugrahaeni, D. K. (2011). *Konsep Dasar Epidemiologi*. Jakarta: Penerbit Buku Kedokteran EGC.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Jr., Flournoy, V. T., & Breslow, N. E. (1978). The Analysis of Failure Times in the Presence of Competing Risks. *Journal of Biometrics*, *4*(34), 541–554.

Sun, L., Liu, J., Sun, J., & Zhang, M. J. (2006). Modeling The Subdistribution of a Competing Risks. *Journal of the Statistica Sinica*, *4*(16), 1367–1385.

Wei, L. J. (1984), *Testing Goodness of Fit for Proportional Hazards Model with Censored Observation*s, *Journal of The American Statistical Association*, 79(387), 649–652.