BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2213-2228
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2213-2228

SURVIVAL ANALYSIS OF CORONARY HEART DISEASE PATIENTS USING THE KAPLAN-MEIER METHOD AND COX PROPORTIONAL HAZARDS REGRESSION (BRESLOW METHOD)

Sudianto Manullang[1*], Marlina Setia Sinaga[2], Pardomuan N. J. M. Sinambela[3], Fauza Nadya[4]

[1,2,3,4]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan Jln. William Iskandar, Pasar V, Kenangan Baru, Percut Sei Tuan, Deli Serdang, Sumatra Utara, 20221, Indonesia*

Corresponding author's e-mail: * ***sudianto.manullang@unimed.ac.id***

Article Info

Article History:

Received: 9th August 2025
Revised: 27th November 2025
Accepted: 9th March 2026
Available online: 8th April 2026

Keywords:

Breslow;
Coronary Heart Disease;
Cox Proportional Hazard;
Kaplan-Meier;
Survival analysis.

ABSTRACT

This study aims to analyze the survival of CHD patients during hospitalization using survival analysis methods. The Kaplan–Meier method was applied to estimate survival probabilities, and group differences were tested using the log-rank test. Furthermore, the Cox proportional hazards model with the Breslow approach was used to assess the effect of clinical factors on survival, with assumptions verified using Schoenfeld residuals. By integrating nonparametric and semiparametric survival methods, this study provides a more comprehensive assessment of CHD patient survival compared with previous studies that relied on a single analytical approach. Data were collected retrospectively from 150 inpatients at Haji General Hospital, Medan, between 2021 and 2022, with 45 cases identified as censored. The Kaplan–Meier analysis revealed a progressive decline in survival probability during hospitalization, with the survival rate decreasing from 69.3% on the first day to 5.3% by day 40. The log-rank test results indicated that only hypertension had a statistically significant effect on patient survival ($p < 0.001$), while age, gender, and cholesterol status were not significant ($p > 0.05$). The Cox regression analysis confirmed these findings, showing that CHD patients with hypertension had more than three times higher risk of death (HR = 3.13; 95% CI: 2.06–4.78) compared to those without hypertension. These findings highlight hypertension as the most dominant risk factor reducing survival among CHD patients during hospitalization. This supports prioritizing early detection and intensive monitoring for hypertensive CHD patients to improve in-hospital clinical outcomes. However, this study has limitations due to its single-center retrospective design and the use of only four variables, leaving out other clinical factors that may influence survival outcomes.

How to cite this article:

S. Manullang, M. S. Sinaga, P. N. J. M. Sinambela, F. Nadya, and W. N. Lubis, "SURVIVAL ANALYSIS OF CORONARY HEART DISEASE PATIENTS USING THE KAPLAN-MEIER METHOD AND COX PROPORTIONAL HAZARDS REGRESSION (BRESLOW METHOD)", *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 2213-2228, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Coronary heart disease (CHD) is a degenerative disease associated with lifestyle and the socioeconomic status of the population. This disease is a major health problem in developed countries. The World Health Organization (WHO) reported that more than 7 million people died from CHD worldwide in 2002. This number is estimated to increase to 11 million by 2020 [1]. Whereas data from the American Heart Association in 2022 indicate that, in 2020, cardiovascular diseases accounted for about 19.1 million deaths worldwide, while around 244.1 million individuals were living with ischemic heart disease. In Indonesia, the Ministry of Health (2020) reported that the mortality rate from CHD was relatively high, with 1.25 million deaths recorded out of a population of 250 million**.** The 2001 National Health Survey also showed that three out of every 1,000 people in Indonesia suffer from CHD [2].

The primary risk factors for coronary disease consist of smoking, high blood lipid levels, hypertension, diabetes mellitus, advanced age, obesity, physical inactivity, and a family history of early coronary artery disease, such as acute myocardial infarction. Recently, additional and sometimes unexpected risk factors have been discovered, for instance, a history of premature birth. Identifying one or more of these newer factors should encourage clinicians to place greater emphasis on reducing the more traditional major risk factors [1]. One of the commonly used methods for analyzing recovery rates is survival analysis, which is a field of statistics concerned with estimating time-to-event distributions while accounting for censoring and truncation [3]. As most research in this field aims to improve patients' survival times or reduce the likelihood of recurrence [4], survival analysis is best applied to research questions involving both the occurrence and timing of events [5].

However, survival times are not always fully observed due to censoring. A key goal of survival analysis is to estimate the survival function, the probability that a patient survives beyond a specified time based on censored data [6]. The Kaplan–Meier curve provides such an estimate when applied to a group of patients with specific characteristics [7]. One major advantage of the Kaplan–Meier method is that it does not require assumptions about the underlying distribution of survival times [8]. Nevertheless, this method also has important limitations. It can only handle a limited number of categorical covariates and is unable to simultaneously assess the effect of multiple clinical factors on survival outcomes. Moreover, when the number of subgroups increases or when continuous predictors are involved, the Kaplan–Meier approach becomes less practical and may lead to unreliable estimates due to insufficient events in each subgroup [9].

To overcome these limitations and allow for a more comprehensive analysis that incorporates several predictors simultaneously, the Cox proportional hazards regression model is often used [10]. This semi-parametric method enables researchers to evaluate the influence of multiple independent variables on the hazard (risk) of an event occurring over time [11]. This model allows researchers to analyze the influence of several independent variables on the hazard (risk) of an event occurring [12]. In this study, the Breslow method was applied to handle tied event times. This approach was chosen because it offers a good balance between computational simplicity and accuracy. Compared to the Efron and Exact methods, Breslow is computationally more efficient for moderate sample sizes and provides stable estimates when the number of tied events is not excessively large. The Efron method generally performs better only when ties are frequent, while the Exact method, though more precise, is computationally intensive and less practical for datasets of this size. Therefore, the Breslow method was considered the most appropriate for the data characteristics in this study. The Kaplan–Meier analysis and the Cox proportional hazards regression model are among the most commonly used methods in survival analysis [13].

Previous studies on survival analysis have been widely conducted. One such study by [14] employed the Kaplan–Meier method and the Cox proportional hazards regression model to analyze the survival of patients with dengue hemorrhagic fever (DHF), considering various suspected factors. The study aimed to estimate the survival function and identify factors influencing the recovery rate of DHF patients. Based on model selection, the study found that age was the only variable that had a significant effect on the survival model. In addition, [15] analyzed factors influencing the length of hospital stay for CHD patients using the Cox proportional hazards regression model with both the Breslow and Exact methods. The results indicated that age had the greatest influence on hospitalization duration in both models. The Breslow method showed a coefficient of 0.6105 for age, while the Exact method showed a higher coefficient of 0.8447. Other variables were not statistically significant in either method.

Based on the above background, this study aims to analyze the survival of coronary heart disease (CHD) patients using the Kaplan–Meier method and the Cox proportional hazards regression model with the

Breslow method. This research is expected to provide valuable information regarding patient survival patterns and their influencing factors.

Previous studies on CHD survival analysis have generally focused on a limited number of risk factors, such as age or gender, and often relied on a single statistical approach, which restricted the depth of interpretation. Moreover, few studies have examined in-hospital CHD survival using an integrated combination of non-parametric and semi-parametric methods. To address this gap, the present study applies a combined approach of Kaplan–Meier, log-rank, and Cox regression with the Breslow method to offer a more comprehensive and reliable analysis of CHD patient survival.

This study is explicitly focused on addressing the main research question of how clinical factors, specifically age, gender, hypertension status, and cholesterol levels, influence the survival probability of CHD patients during hospitalization. The analysis is conducted using an integrated approach that includes the Kaplan–Meier estimator, the log-rank test, and the Cox proportional hazards model with the Breslow method. The findings are anticipated to assist healthcare professionals, researchers, hospital administrators, and policymakers in developing more effective strategies to manage and reduce mortality from heart disease.

2. RESEARCH METHODS

2.1 Data Sources and Variables

This study employs a quantitative approach using survival analysis methods to assess the survival probability of Coronary Heart Disease (CHD) patients during hospitalization at Haji General Hospital, Medan, from 2021 to 2022. The survival function was estimated using both the Kaplan–Meier method and the Cox proportional hazards regression model with the Breslow method. Comparisons between patient groups were conducted using the log-rank test. Research data were obtained retrospectively from the medical records of inpatients diagnosed with CHD who met the inclusion criteria, namely patients with complete data on admission and discharge dates, as well as final status (alive or deceased). Data collection was carried out by reviewing medical record archives using a custom-designed observation form developed specifically for this study.

The variables in this study consist of both dependent and independent variables. The dependent variable is survival, measured as the duration of patient hospitalization (in days). The independent variables include age, gender, cholesterol status, and hypertension. Age is categorized into two groups: ≤ 45 years old and ≥ 45 years old. Gender is classified as female or male. Cholesterol status is categorized as normal or high based on clinical cut-off values (total cholesterol ≥ 200 mg/dL is considered high). Hypertension indicates whether the patient has hypertension or not, determined based on a physician's diagnosis or a blood pressure measurement of ≥ 140/90 mmHg.

2.2 Research Variables

In this study, data from 150 hospitalized patients were analyzed, of which 45 cases represented censored data patients who did not experience the event by the end of the observation period or were lost to follow-up for specific reasons. The variables examined in this study include length of hospital stay (in days), age, gender, cholesterol status, and hypertension. A detailed description of these variables is provided in Table 1.

Table 1. Research Variables

No.	Variable	Description	Data Format
1.	Censor	The censor variable indicates whether the data is censored or not, based on the occurrence of the event. In this study, the event is defined as a patient being discharged in improved condition. Therefore, censored data refers to patients who either died or were lost to follow-up.	0 : Uncensored data 1 : Censored data
2.	Survival	Duration of patient hospitalization	Numeric
3.	Age	Patient age categorized as below or above 45 years old.	0 : ≤ 45 years old 1 : ≥ 45 years old
4.	Gender	Female or Male	0 : Female 1 : Male

No.	Variable	Description	Data Format
5.	Cholesterol Status	Blood cholesterol level	0 : No high cholesterol 1 : High cholesterol
6.	Hypertension	Presence of hypertension	0 : No hypertension 1 : Hypertension

2.3 Survival Analysis

Survival analysis is a statistical method used to examine data by considering the time from the beginning of observation until the occurrence of a specific event [16]. The data utilized in this analysis is *survival time data*, which acts as the independent variable [17]. Time is defined as the period of observation until the occurrence of a particular event, and it is typically measured in days, weeks, months, or years [18]. The term *event* refers to specific outcomes, such as relapse, recovery, or death [19].

2.4 Survival Function

The survival function $S(t)$ represents the probability that a random variable denoting survival time exceeds a specified time (t) [20]. It is defined as follows:

$$S(T) = 1 - F(t) = 1 - P(T \leq 1) = \int_t^\infty f(u)\, du\,. \tag{1}$$

2.5 Probability Density Function

The probability density function is often also called the Probability Density Function, which is the probability of an individual dying or experiencing a momentary event in a time interval t until $t + \Delta t$ [21]. The probability density $f(t)$ is formulated as follows [22]:

$$f(t) = \lim_{\Delta t \to 0} \frac{P\big(t < T < (t + \Delta t)\big)}{\Delta t} = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}. \tag{2}$$

Based on Eq. (2), we get:

$$f(t) = \frac{d\big(F(t)\big)}{dt} = F'^{(t)}. \tag{3}$$

2.6 Kaplan-Meier Method

The Kaplan–Meier method is a non-parametric technique used to estimate survival functions [23], particularly useful for handling incomplete (censored) data and applicable to small sample sizes [24].

2.6.1 Estimation Using the Kaplan-Meier Method

This method, also known as the product-limit estimator, was introduced by Kaplan and Meier in 1958. It is used to calculate the survival curve based on survival time data [25]. The equation used in the Kaplan–Meier method is as follows:

$$S(t) = S_{t-1} \times \frac{(n_i - d_i)}{n_i}, \tag{4}$$

with :

$S(t)$	: Probability of a person surviving from the beginning of the study;
S_{t-1}	: Cumulative probability of surviving until just before time t;
n_i	: The number of individuals at risk of failure at time i;
d_i	: The number of individuals who experience failure at time-i;
i, t	: Range from 1 to k, where k represents the total number of distinct event times observed in the study.

2.6.2 Log-Rank Test

The log-rank test was used to compare survival distributions between two or more groups in various situations [26]. This test determines whether there is a significant difference in survival times and survival functions between the groups being compared [27]. The primary purpose of using the log-rank test is to assess whether the Kaplan–Meier curves for different categories differ significantly. The hypotheses used in the log-rank test are as follows:

$H_0 : S_1(t) = S_2(t)$, There is no significant difference between the two survival probabilities.

$H_1 : S_1(t) \neq S_2(t)$, There is a significant difference between the two survival probabilities.

The test result is determined by the $p -$ value at the significance level $= 0.05$ as follows :

If the $X^2 < X_{\alpha,G-1}$ and calculated $p -$ value > 0.05, there is no difference between the two survival functions; therefore, accept H_0 and reject H_1.

If the $X^2 > X_{\alpha,G-1}$ and calculated $p -$ value < 0.05, there is a difference between the two survival functions; therefore, accept H_1 and reject H_0.

2.7 Hazard Function

The hazard function is a basic quantity that is the basis of survival analysis [28]. It represents the risk of an event (such as failure or death) occurring within a very small time interval, assuming the individual has survived up to that point. Let $h(t)$ denotes the hazard function at time t, and $S(t)$ denotes the survival function, which gives the probability of surviving beyond time t. The hazard and survival functions are related as follows:

$$h(t) = -\frac{d}{dt} \ln S(t) \text{, or equivalently } S(t) = \exp\left(-\int_0^t h(u)du\right) \quad (5)$$

In other words, the hazard function describes how the probability of survival decreases over time, while the survival function can be derived from the cumulative hazard.

2.7.1 Cox Proportional Hazards Model

The Cox proportional hazards model is termed the Cox model since it assumes that the hazard functions across individuals are proportional, implying that the ratio between the hazard functions of any two individuals remains constant [29]. The Cox model can be written as follows:

$$h(t, \boldsymbol{X}) = h_0(t)\, exp\left(X_1\beta_1 + X_2\beta_2 + \cdots + X_p\beta_p\right) = h_0(t)\, exp\left(\sum_{i}^{p} X_i\beta_i\right). \quad (6)$$

2.7.2 Parameter Estimation using the Breslow Approach

The Breslow method is the simplest among the available methods, as it is not computationally intensive. In addition, this method provides good estimation results when the number of tied events is small. In general, the Breslow method can be expressed as follows:

$$L(\beta_{breslow}) = \prod_{j \in D} \frac{exp(\beta S_k)}{\left[\sum_{j \in R_k} exp(\beta X_j)\right]^{d_i}} \quad (7)$$

With:

X	: Covariate vector (explanatory variables);
β	: The regression parameters to be estimated;
D	: The set of indices j of all event times;
R_k	: The risk set consists of all individuals who have not yet experienced the event at a given time;
S_K	: The sum of covariate x in the case of tied events;
d_i	: The number of tied events at time t_j.

2.7.3 Proportional Hazards Assumption

According to Kleinbaum and Klein, the proportional hazards assumption can be assessed using a goodness-of-fit estimation method [29]. This method employs test statistics to evaluate the assumption, making it more objective than graphical methods. The test statistics used in this method are the Schoenfeld residuals, with the testing steps outlined as follows:

1. Construct the Cox proportional hazards model and compute the Schoenfeld residuals for each individual and for each covariate.
2. Create a variable representing the rank of survival times.

Test the correlation between the Schoenfeld residuals and the survival time ranks from step 2.

Hypothesis:

$H_0 : \rho = 0,$

$H_1 : \rho \neq 0.$

Test statistic:

$$r = \frac{\sum_j^n (R_{ij} - \overline{R}_{ij})(RT_j - \overline{RT}_j)}{\sqrt{\sum_j^n (R_{ij} - \overline{R}_{ij})^2}\sqrt{\sum_j^n (RT_j - \overline{RT}_j)^2}}, \quad (8)$$

with:

r : correlation coefficient;
n : the number of individuals who experienced the event;
R_{ij} : the residual of the i covariate for the j individual;
RT_j : the rank of survival time for the j individual;
$\overline{RT_j}$: the average rank of survival times for the j individual.

Rejection criteria:

H_0 is rejected if $|r_{count}| > r_{n-2,\frac{a}{2}}$ or $p - value < a$ with $a = 0.05$.

2.7.4 Testing of Parameters

Parameter testing is conducted using the likelihood ratio test and the Wald test [30]. These tests are performed to determine whether the independent variables have a significant effect on the model.

1. Likelihood ratio test

   Hypothesis:

   $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0;$

   H_1: there is at least one $\beta_j \neq 0$, by $j = 1,2,3, \dots, p$.

   Test statistic G for the likelihood test ratio is as follows:

$$G^2 = -2\left[\ln L_R - \ln L_f\right]. \quad (9)$$

   Rejection criteria: H_0 is rejected if $G \geq \chi^2_{(\alpha,p)}$ or $p - value \leq a = 0.05$.

2. Wald test

   Hypothesis:

   $H_0 : \beta_j = 0;$

   $H_1 : \beta_j \neq 0; j = 1,2,3, \dots, p.$

   Test statistic:

$$|W| = \left|\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right| \quad (10)$$

   Rejection criteria: H_0 is rejected if $W > Z_{\frac{\alpha}{2}}$ or $p - value < a = 0.05$.

3. RESULTS AND DISCUSSION

3.1 Kaplan-Meier Method

This method provides insights into the probability of patient recovery at each time interval, thereby facilitating the analysis of treatment effectiveness and recovery duration.

3.1.1 Kaplan-Meier Estimation Method

Table 2 presents the Kaplan-Meier estimation results illustrating the patients' survival probability to reach recovery during the hospitalization period.

Table 2. Estimated Survival Probability of Coronary Heart Disease Patients

Time	n.risk	n.event	survival	Std.err	Lower 95%	CI upper 95% CI
1	150	46	0.6933	0.0376	0.6233	0.771
2	90	5	0.5448	0.0393	0.5821	0.737
3	79	5	0.6134	0.0410	0.5381	0.699
4	69	6	0.5600	0.0428	0.4821	0.651
5	60	4	0.5227	0.0438	0.4435	0.616
6	56	3	0.4947	0.0444	0.4150	0.590
7	49	6	0.4341	0.0453	0.3538	0.533
8	42	5	0.3824	0.0454	0.3030	0.483
9	37	3	0.3514	0.0451	0.2732	0.452
10	33	3	0.3195	0.0446	0.2430	0.420
11	30	41	0.2769	0.0435	0.2035	0.377
13	35	3	0.2437	0.0423	0.1734	0.342
14	22	1	0.2326	0.0418	0.1636	0.331
15	21	2	0.2104	0.0406	0.1441	0.307
17	16	2	0.1841	0.0396	0.1208	0.281
21	13	2	0.1558	0.0382	0.0964	0.252
22	11	2	0.1275	0.0361	0.0731	0.222
25	6	1	0.1062	0.0358	0.0549	0.206
34	4	1	0.0797	0.0354	0.0334	0.190
40	3	1	0.0531	0.0320	0.0163	0.173

The Kaplan–Meier analysis was conducted to evaluate the survival probability of patients with coronary heart disease (CHD) during hospitalization. Based on the estimation results, on the first day, the patients' survival probability was 69.3% (95% CI: 62.3%–77.1%) on the first day of hospitalization. This probability gradually declined over time, reaching approximately 5.3% (95% CI: 1.6%–17.3%) by day 40, indicating that most recovery events occurred earlier during the hospital stay. These results suggest that the likelihood of experiencing the event (recovery or discharge) increases as the length of hospitalization progresses.

The Kaplan–Meier curve also reveals a steep decline in the early days of hospitalization, suggesting that most events occurred within the initial period. After presenting the survival probability estimates in tabular form, the trend can be more clearly visualized through the Kaplan–Meier curve shown in Fig. 1. This curve illustrates the changing probability of recovery for CHD patients throughout their hospital stay.

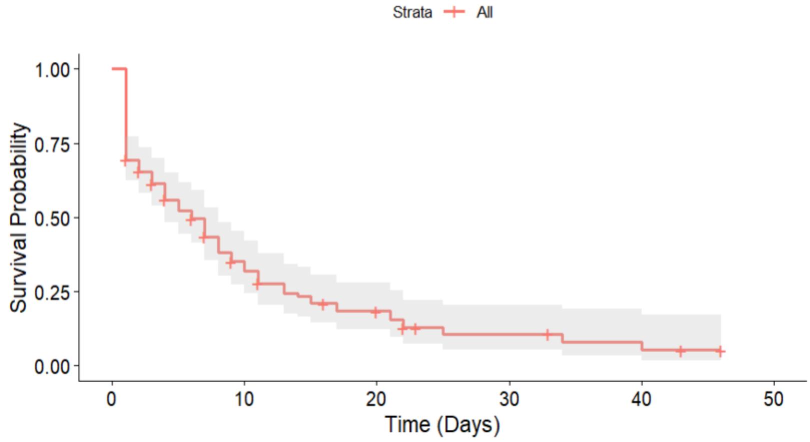

Figure 1. Survival Curve of Coronary Heart Disease Patients

The shaded gray area around the curve represents the 95% confidence interval, which widens over time as the number of remaining patients decreases, increasing the uncertainty of the estimates. Overall, these results reflect a relatively low survival rate among hospitalized CHD patients and underscore the importance of early intervention and intensive monitoring during the initial stages of hospitalization.

3.1.2 Log Rank Test

To determine whether there are significant differences in patient recovery probabilities based on age category, gender, hypertension, and cholesterol status, the log-rank test was applied. This test is used to compare survival functions between groups in order to assess the influence of these variables on the survival time of patients with coronary heart disease.

3.1.2.1 Log Rank Test Based on Age

The following are the results of the comparative analysis of survival functions between two age groups: patients under 45 years old and those aged 45 and above. The results of the log-rank test based on age are presented in Table 3.

Table 3. Log-Rank Test Based on Age

Variable	Category	n (Subject)	Chi-square	df	$p-value$	Information
Age	$0: \leq 45$	69	2.3	1	0.100	not significant
	$1: > 45$	81				($p > 0.05$)

Based on the log-rank test results presented in Table 3, a Kaplan–Meier curve was subsequently generated to visualize the survival patterns by age group.

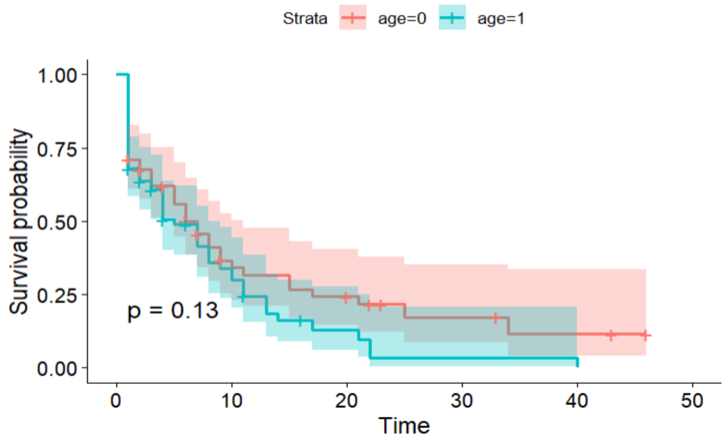

Figure 2. Kaplan-Meier Curve by Age

In Fig. 2, the Kaplan–Meier curve shows that patients aged ≤ 45 years (age = 0) exhibit a higher survival probability compared to those aged > 45 years (age = 1). However, based on the log-rank test, the obtained $p-$ value is 0.10, indicating that the difference between the two age groups is not statistically significant. This suggests that there is insufficient evidence to conclude that age has a significant effect on patient survival during hospitalization.

3.1.2.2 Log-Rank Test Based on Gender

In this analysis, patients were classified into two groups based on their gender, namely female and male. This classification enables a comparison of survival experiences between the two groups, providing insight into whether gender has a significant impact on patient survival during hospitalization.

Table 4. Log-Rank Test Based on Gender

Variable	Category	n (Subject)	Chi-square	df	$p-value$	Information
Gender	0: female	64	3.5	1	0.060	not significant
	1: male	86				($p > 0.05$)

Based on the log-rank test results presented in Table 4, a Kaplan–Meier curve was subsequently generated to visualize the survival patterns based on gender.

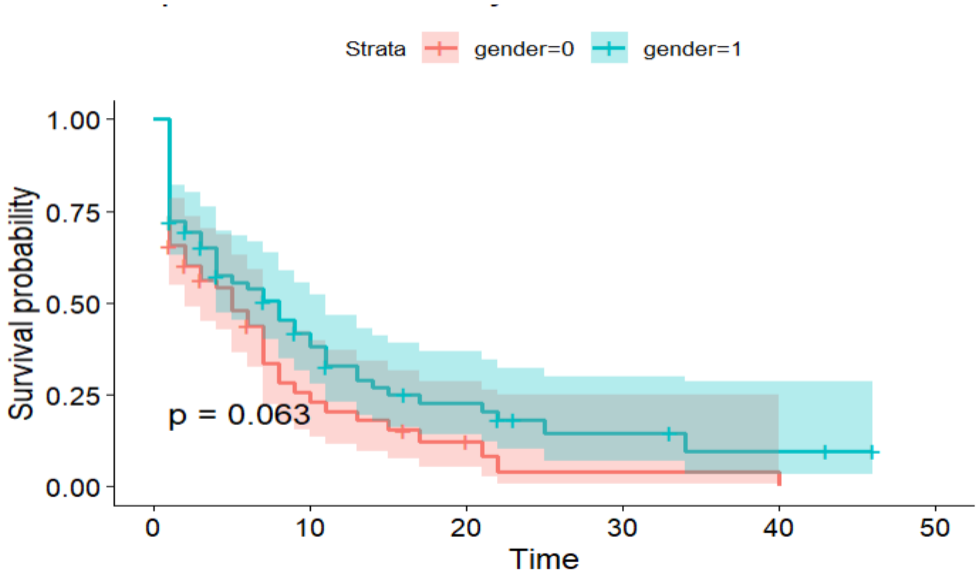

Figure 3. Kaplan-Meier Curve by Gender

Based on the Kaplan–Meier curve shown in Fig. 3, male patients (gender = 1) appear to have a slightly higher survival probability compared to female patients (gender = 0) during the hospitalization period. However, the log-rank test yielded a $p-$ value of 0.06, indicating that the difference in survival time between genders is not statistically significant ($p > 0.05$). Therefore, there is insufficient evidence to conclude that gender has a significant effect on the survival of CHD patients during hospitalization.

3.1.2.3 Log Rank Test Based on Cholesterol

To assess the effect of cholesterol levels on patient survival, an analysis was conducted using the Log-Rank test. This test compares the survival curves between two groups of patients: those with normal cholesterol levels and those with high cholesterol levels.

Table 5. Log-Rank Test Based on Cholesterol

Variable	Category	n (Subject)	Chi-square	df	$p-value$	Information
Cholesterol	0: No high cholesterol	86	0.8	1	0.400	not significant
	1: high cholesterol	64				($p < 0.05$)

Based on the log-rank test results presented in Table 5, a Kaplan–Meier curve was subsequently generated to visualize and compare the survival patterns of patients according to their cholesterol status, facilitating the identification of differences in survival probabilities between patients with normal and high cholesterol levels.

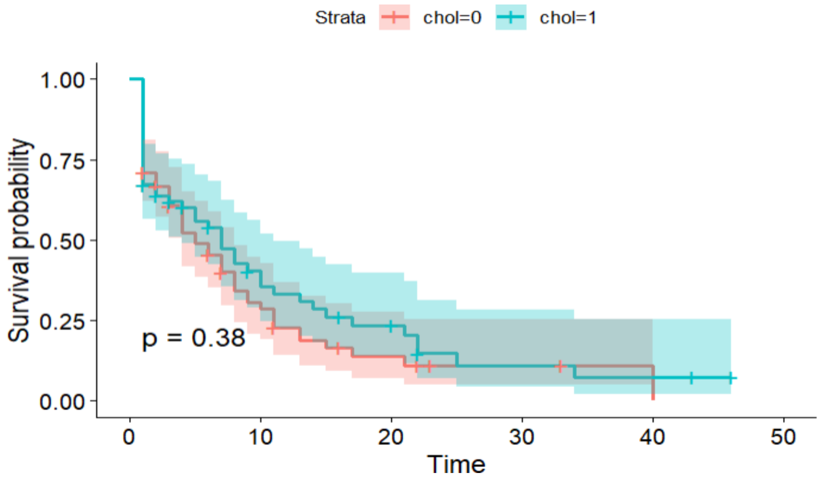

Figure 4. Kaplan-Meier Curve by Cholesterol

The Kaplan–Meier curve in Fig. 4 shows no substantial difference in survival probabilities between patients with high cholesterol levels (chol = 1) and those with normal levels (chol = 0). This observation is supported by the log-rank test, which yielded a $p-$ value of 0.4, indicating that cholesterol status does not have a statistically significant effect on patient survival ($p > 0.05$). Thus, cholesterol status does not significantly influence the risk of death among CHD patients during hospitalization.

3.1.2.4 Log-Rank Test Based on Hypertension Status

To determine the effect of hypertension on patient survival, a Log-Rank test was conducted based on hypertension status.

Table 6. Log-Rank Test Based on Hypertension Status

Variable	Category	n (Subject)	Chi-square	df	$p-value$	Information
Hypertension	0: No hypertension	70	30 .9	1	3×10^{-08}	significant
	1: hypertension	80				($p < 0.001$)

Based on the log-rank test results presented in Table 6, a Kaplan–Meier curve was subsequently generated to visualize the survival patterns based on hypertension status.

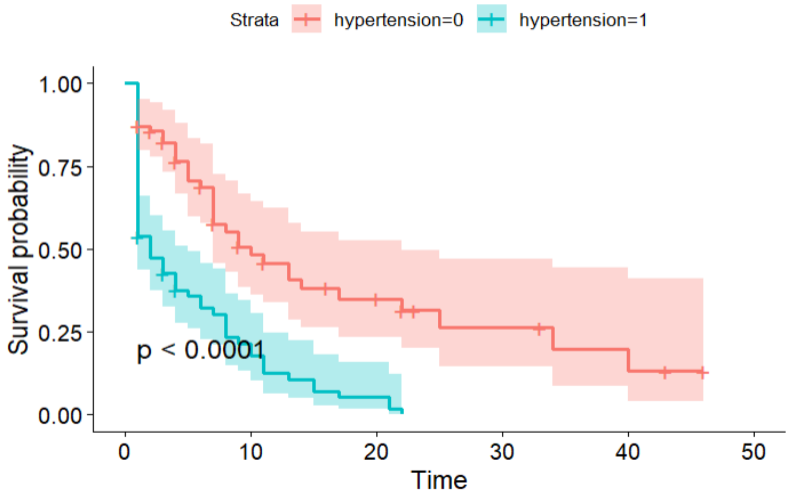

Figure 5. Kaplan-Meier Curve by Hypertension

The results presented in the table and figure above indicate a highly significant effect of hypertension on patient recovery. As shown in Fig. 5, patients without hypertension (hypertension = 0) recover more rapidly compared to those with hypertension (hypertension = 1). The log-rank test yielded a $p-$ value of 3×10^{-8}, which is far below the 0.05 significance threshold, indicating that this difference is statistically highly significant. Therefore, hypertension is a major factor that prolongs hospitalization or delays recovery in CHD patients, rather than reducing survival. The curve visually reflects this, showing faster recovery in patients without hypertension, indicated by the steeper decline in the survival probability.

3.2 Cox Proportional Hazards Regression Modelling

3.2.1 Initial Cox Proportional Hazards Regression Model Using the Breslow Method

To evaluate the effect of the independent variables (age, gender, cholesterol, and hypertension) on the risk of death in patients with coronary heart disease, an initial Cox proportional hazards regression model was developed using the Breslow method. The parameter estimation results of this model are presented in Table 7.

Table 7. Model Parameter Estimation Using the Breslow Method

Variable	coef	z	Lower limit	Upper limit	Pr(>|z|)
Age	0.2881	1.415	0.8949	1.988	0.1572
Gender	−0.4004	−1.95	0.4475	1.003	0.0519
Cholesterol	−0.1661	−0.79	0.5609	1.279	0.4291
Hypertension	1.1420	5.308	2.0553	4.777	1.11e-07

The initial Cox proportional hazards regression model using the Breslow method is as follows:

$$h(t, X) = h_0(t) \cdot exp(0.2881 \cdot X1 - 0.4001 \cdot X2 - 0.1661 \cdot X3 + 1.1420 \cdot X4).$$

3.2.2 Proportional Hazard Assumption Testing

The following presents the interpretation of the proportional hazards (PH) assumption test results for the Cox regression model in Table 8, based on both numerical output and graphical analysis using Schoenfeld residuals. This evaluation is crucial to determine whether the effect of each covariate on the hazard remains constant over time, an essential assumption of the Cox model.

Table 8. Testing the Proportional Hazards Assumption in The Cox Model

	chi-square	$p -$ value	df
X_1	0.916	0.34	1
X_2	0.305	0.58	1
X_3	0.702	0.40	1
X_4	0.262	0.61	1
Global	1.950	0.74	4

From Table 8 above, it can be observed that the $p -$ value for each variable, as well as the overall (global) $p -$ value, are all above 0.05. This indicates that there is no significant evidence of a violation of the proportional hazards assumption. Additionally, the low chi-square values further support the absence of any relationship between time and the covariate effects on the hazard. Therefore, the Cox model is considered valid and suitable for further interpretation.

Subsequently, the Schoenfeld residual plot is presented to illustrate the behavior of the regression coefficients (β) over time. The solid line represents the estimated trend of the covariate effect, while the dashed lines denote the 95% confidence interval. If the solid line remains within the dashed bands and does not exhibit a consistent upward or downward pattern, it can be concluded that the proportional hazards assumption has not been violated.

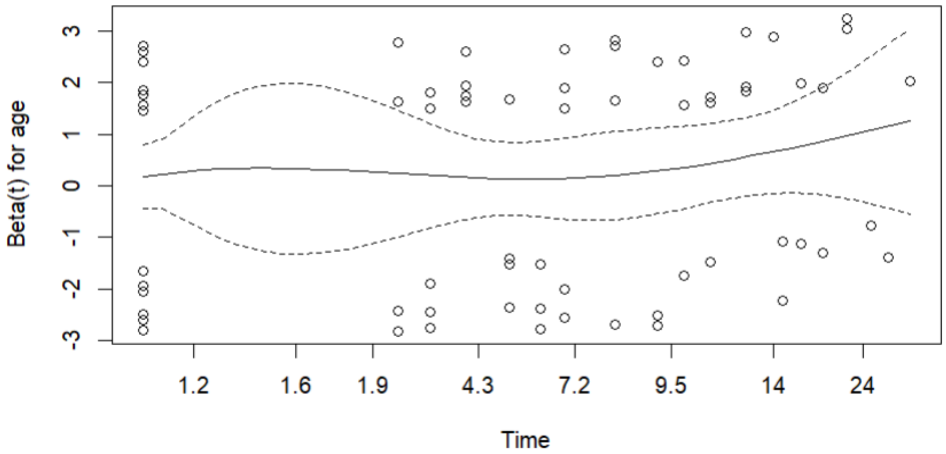

Figure 6. Schoenfeld Residual Plot for Age

Fig. 6 illustrates the relationship between the effect of the age variable and time. The trend line appears relatively flat, and the entire solid line remains within the 95% confidence interval band (dashed lines). The residual points are scattered randomly, showing no systematic pattern. These findings indicate that the effect of age on the hazard of the event (discharge or death) does not vary significantly over time. Therefore, the proportional hazards assumption for the age variable is considered to be met.

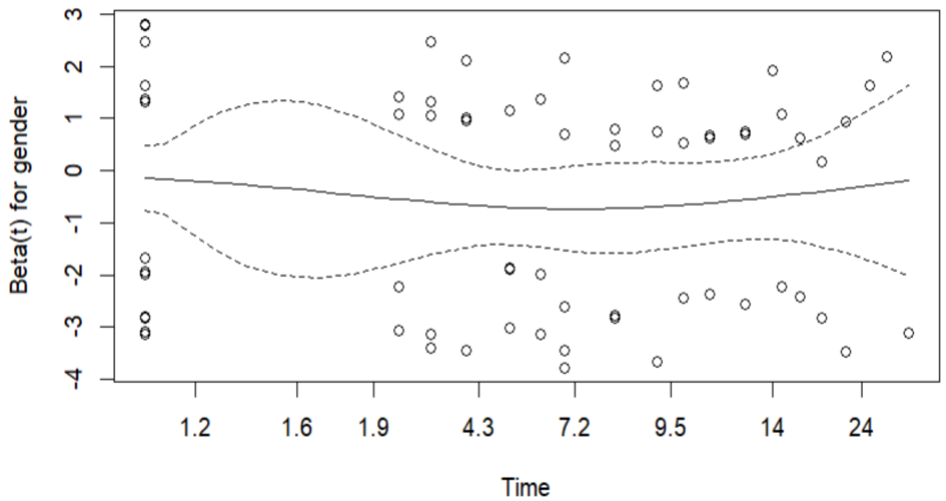

Figure 7. Schoenfeld Residual Plot for Gender

In Fig. 7, the trend of the β coefficient for the gender variable also appears flat and remains within the 95% confidence interval bounds. The residual points are randomly scattered, showing no systematic pattern. This suggests that the effect of gender on the hazard remains constant over time. Therefore, the gender variable satisfies the proportional hazards assumption in the model.

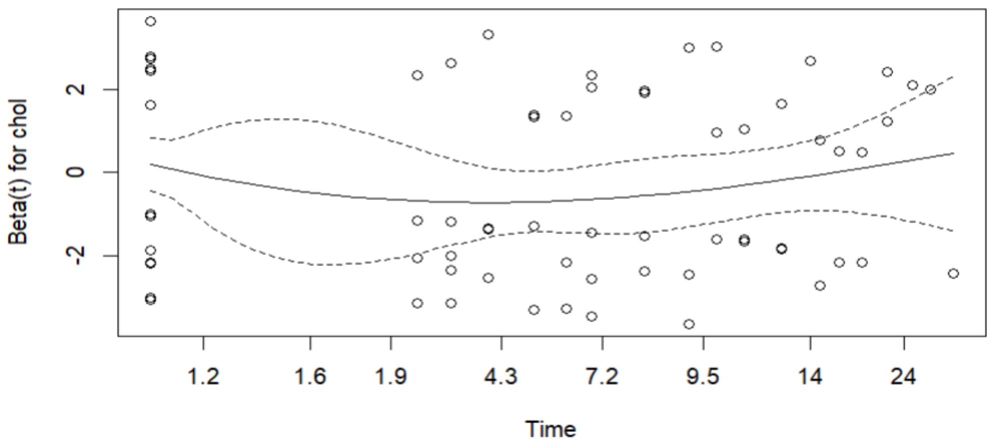

Figure 8. Schoenfeld Residual Plot for Cholesterol Status

The plot for the cholesterol (chol) variable, Fig. 8, shows a slight upward trend toward the end; however, it remains within the 95% confidence interval bounds. The residual points are randomly scattered and do not exhibit a systematic increasing or decreasing pattern. This suggests that the cholesterol variable does not violate the proportional hazards assumption, and its effect on survival can be considered constant over time.

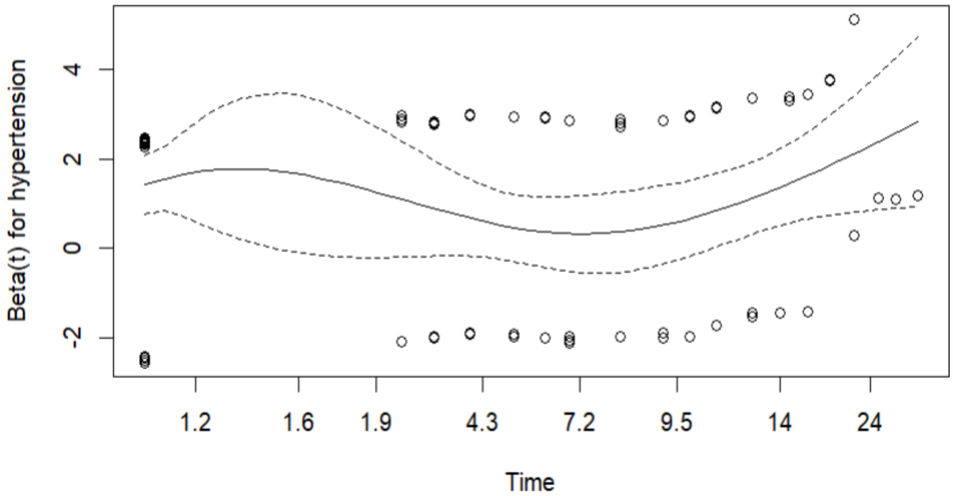

Figure 9. **Schoenfeld Residual Plot for Hypertension**

In the hypertension plot, Fig. 9, the trend line shows a slight upward movement toward the end of the time period but remains within the 95% confidence interval band. Although there is a concentration of residual points at the beginning and end of the timeline, no systematic pattern suggests a violation of the assumption. Therefore, it can be concluded that the effect of hypertension on the risk of events does not vary over time, and the proportional hazards assumption is satisfied for this variable.

Based on both the statistical test results and the Schoenfeld residual plots, it can be concluded that all variables in the model, age, gender, cholesterol status, and hypertension, satisfy the proportional hazards (PH**)** assumption. The consistently high $p-$ value and flat trend lines are strong indicators that the effects of these covariates on the hazard function remain constant throughout the observation period. Consequently, the Cox proportional hazards model used in this study is considered valid and appropriate for analyzing the factors influencing the survival of patients with coronary heart disease (CHD).

3.2.3 Parameter Testing

1. Likelihood ratio test

   The Likelihood Ratio Test yielded a test statistic of 36.74 with 4 degrees of freedom (df) and a $p-$ value of 2×10^{-07}, indicating that the model is significantly better than the null model (without predictors).

2. Wald Test

   The Wald Test produced a test statistic of 33.58 (df = 4) with a $p-$ value of 9×10^{-07}, which also indicates that at least one independent variable significantly affects the hazard (risk of event).

Meanwhile, the Score (Log-Rank) Test also produced results identical to the Likelihood Ratio Test, with a test statistic of 36.74, df = 4, and a $p-$ value of 2×10^{-07}. The consistency across these three tests further reinforces the conclusion that the constructed Cox regression model is statistically significant overall.

Additionally, the concordance index (C-index) of 0.696, with a standard error (SE) of 0.032, indicates that the model possesses a reasonably good discriminatory ability in predicting the sequence of event times (e.g., deaths) among patients. A concordance value approaching 0.7 suggests that the model is fairly effective in distinguishing patients who experience earlier events from those who experience them later. In conclusion, these findings demonstrate that the Cox regression model is appropriate for use and exhibits satisfactory performance in explaining the relationship between clinical factors such as age, gender, cholesterol, and hypertension and the risk of death in patients with coronary heart disease (CHD) during hospitalization.

The findings of this study indicate that hypertension is the most significant factor influencing the survival of CHD (coronary heart disease) patients during hospitalization. This result is consistent with previous studies that also reported hypertension as a significant contributor to increased risk of prolonged hospital stay and mortality in CHD patients. Similarly, [15] found that age was a determining factor in patient outcomes, although in our study, age was not statistically significant. The difference may be due to variations in patient characteristics and sample sizes across studies. Furthermore, the robustness of our model, confirmed by the proportional hazard assumption testing, strengthens the conclusion that hypertension plays a dominant role in reducing survival probability.

4. CONCLUSION

Based on the results and discussion from the Kaplan-Meier method and Cox regression using the Breslow method, it can be concluded that:

1. The Kaplan-Meier analysis showed a decline in survival probability over the course of hospitalization. On the first day, the survival probability of CHD patients was 69.3% (95% CI: 62.3%–77.1%), decreasing to approximately 5.3% (95% CI: 1.6%–17.3%) by day 40. The log-rank test indicated that only the hypertension variable exhibited a statistically significant difference in survival between groups ($p < 0.001$). In contrast, the variables age, gender, and cholesterol status did not show statistically significant differences ($p > 0.05$). This demonstrates that the risk of death or discharge increases over time, highlighting the importance of early monitoring and intervention.
2. The results of the Cox regression analysis reinforce the findings of the Kaplan-Meier analysis. Among the four variables tested, only hypertension had a significant effect on the risk of events, with a hazard ratio more than three times higher compared to patients without hypertension. Age, gender, and cholesterol variables did not show significant effects in the multivariate model. This model indicates that hypertension is a major risk factor in reducing the survival time of coronary heart disease patients during hospitalization.

Author Contributions

Sudianto Manullang: Conceptualization, Methodology, Supervision, Project Administration, and Writing–Original Draft. Marlina Setia Sinaga: Formal Analysis, Data Curation, and Writing–Review and Editing. Pardomuan N. J. M. Sinambela: Investigation, Resources, Validation. Fauza Nadya: Visualization, Data Curation, and Writing–Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research has been funded by BOPTN Universitas Negeri Medan with contract number: 0183/UN33.8/PPKM/PPT/2023.

Acknowledgment

The authors would like to express their gratitude and appreciation to Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Negeri Medan. Special thanks are also extended to the students who assisted in the data collection and processing, whose support was instrumental to the success of this research.

Declarations

The authors declare no conflict of interest to report study.

Declaration of Generative AI and AI-assisted Technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement, including grammar, spelling, and clarity. The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. All final text was reviewed and approved by the authors.

REFERENCES

[1] M. Joseph and S. Alpert, "NEW CORONARY HEART DISEASE RISK FACTORS," *Am. J. Med.*, vol. 136, no. 4, page. p331-332, 2022. doi: https://doi.org/10.1016/j.amjmed.2022.08.002

[2] M. R. Muliantino, N. Z. Qadri, E. Afriyanti, and R. Sarfika, "SELF-EFFICACY IN INCREASING PHYSICAL ACTIVITY OF CORONARY HEART DISEASE PATIENTS: A CROSS-SECTIONAL STUDY," *J. Ners*, vol. 19, no. 3, page. 371–377, 2024. doi: https://doi.org/10.20473/jn.v19i3.51941

[3] R. Sonabend, F. J. Kiraly, A. Bender, B. Bischl, and M. Lang, "AN R PACKAGE FOR MACHINE LEARNING IN SURVIVAL ANALYSIS," *Bioinformatics*, vol. 37, no. 17, page. 2789–2791, 2021. doi: https://doi.org/10.1093/bioinformatics/btab039

[4] W. Tizi and A. Berrado, "MACHINE LEARNING FOR SURVIVAL ANALYSIS IN CANCER RESEARCH: A COMPARATIVE STUDY," *Sci. African*, vol. 21, no. e01880, 2023. doi: https://doi.org/10.1016/j.sciaf.2023.e01880

[5] Q. E. Denfeld, D. Burger, and C. S. Lee, "AN EASY START GUIDE TO ANALYZING TIME TO-EVENT DATA," *Eur J Cardiovasc Nurs. Author Manuscr.*, vol. 22, no. 3, page. 332–337, 2024. doi: https://doi.org/10.1093/eurjcn/zvad023

[6] I. K. Hasan, W. A. Pakaya, N. Achmad, and D. R. Isa, "ANALISIS SURVIVAL MENGGUNAKAN REGRESI WEIBULL PADA LAJU KESEMBUHAN PASIEN TUBERKULOSIS PARU DI RSUD ALOEI SABOE KOTA GORONTALO," *Euler J. Ilm. Mat. Sains and Teknol.*, vol. 9, no. 1, page. 40–51, 2021. doi: https://doi.org/10.34312/euler.v9i1.10758

[7] S. Annas, M. N. Bustam, and Mahrani, "METODE KAPLAN-MEIER UNTUK ANALISIS KETAHANAN HIDUP PENDERITA KANKER PAYUDARA DI RSUD KOTA MAKASSAR," *J. Stat. Its Appl. Teach. Res.*, vol. Vol.3 No 2, no. 2, page. 58–66, 2021. doi: https://doi.org/10.35580/variansiunm23853.

[8] G. D'Arrigo, D. Leonardis, S. A. ElHafeez, G. Tripepi, and S. Roumeliotis, "METHODS TO ANALYSE TIME-TO-EVENT DATA: THE KAPLAN-MEIER SURVIVAL CURVE," *Oxid. Med. Cell. Longev.*, page. 7, 2021. doi: https://doi.org/10.1155/2021/2290120

[9] E. Candès, L. Lei, and Z. Ren, "CONFORMALIZED SURVIVAL ANALYSIS," *J. R. Stat. Soc. Ser. B*, vol. 85, page. 24–25, 2023.

[10] I. Dormuth, T. Liu, J. Xu, M. Pauly, and M. Ditzhaus, "A COMPARATIVE STUDY TO ALTERNATIVES TO THE LOG-RANK TEST," *arXiv:2210.13258v1 [stat.ME]*, 2022. doi: https://doi.org/10.1016/j.cct.2023.107165

[11] Z. Zhang, "SEMI-PARAMETRIC REGRESSION MODEL FOR SURVIVAL DATA: GRAPHICAL VISUALIZATION WITH R," *Ann Transl Med*, vol. 4, no. 23, 2016. doi: https://doi.org/10.21037/atm.2016.08.61

[12] A. R. Faisal, M. N. Bustan, and S. Annas, "ANALISIS SURVIVAL DENGAN PEMODELAN REGRESI COX PROPORTIONAL HAZARD MENGGUNAKAN PENDEKATAN BAYESIAN (STUDI KASUS: PASIEN RAWAT INAP PENDERITA DEMAM TIFOID DI RSUD HAJI MAKASSAR)," *J. Stat. Its Appl. Teach. Res.*, vol. 2, no. 2, page. 62–70, 2021. doi: https://doi.org/10.35580/variansiunm14629

[13] S. W. Lee, "KAPLAN-MEIER AND COX PROPORTIONAL HAZARDS REGRESSION IN SURVIVAL ANALYSIS: STATISTICAL STANDARD AND GUIDELINE OF LIFE CYCLE COMMITTEE," *Life Cycle*, vol. 3, no. e8, 2023. doi: https://doi.org/10.54724/lc.2023.e8

[14] F. Khairunnisa, F. Saumi, and Amelia, "SURVIVAL ANALYSIS OF DENGUE HEMORRHAGIC FEVER PATIENTS (DHF)," *Barekeng*, vol. 16, no. 3, page. 897–908, 2022. doi: https://doi.org/10.30598/barekengvol16iss3pp897-908

[15] M. Nainggolan and S. Manullang, "ANALISIS SURVIVAL PASIEN PENYAKIT JANTUNG KORONER DENGAN METODE BRESLOW DAN EXACT PADA MODEL REGRESI COX PROPORTIONAL HAZARD," *Innov. J. Soc. Sci. Res.*, vol. 4, page. 2350–2360, 2024.

[16] S. Lee and H. Lim, "REVIEW OF STATISTICAL METHODS FOR SURVIVAL ANALYSIS USING GENOMIC DATA," *Genomic Inform.*, vol. 17, no. 4, 219M. doi: https://doi.org/10.5808/GI.2019.17.4.e41

[17] M. Muhajir and Y. D. Palupi, "SURVIVAL ANALYSIS OF CHILD PATIENT DIARRHEA USING KAPLAN-MEIER METHOD AND RANK LOG TEST," *EKSAKTA J. Sci. Data Anal.*, vol. 18, page. 74–84, 2018. doi: https://doi.org/10.20885/eksakta.vol18.iss1.art8

[18] P. Wang, Yanli, and C. K.Reddy, "MACHINE LEARNING FOR SURVIVAL ANALYSIS: A SURVEY," *ACMComputing Surv.*, vol. 51, no. 6, page. 36, 2019. doi: https://doi.org/10.1145/3214306

[19] C. Kartsonaki, "SURVIVAL ANALYSIS," *Elvesier*, vol. 22, no. 7, page. 263–270, 2016. doi: https://doi.org/10.1016/j.mpdhp.2016.06.005

[20] E. Giunchiglia, A. Nemchenko, and M. van der Schaar, "A DEEP RECURRENT MODEL FOR SURVIVAL ANALYSIS," *Lect. Notes Comput. Sci.*, page. 23–32, 2018. doi: https://doi.org/10.1007/978-3-030-01424-7_3

[21] M. J. Simpson, J. A. Sharp, and R. E. Baker, "SURVIVAL PROBABILITY FOR A DIFFUSIVE PROCESS ON A GROWING DOMAIN," *Am. Phys. Soc.*, 2015. doi: https://doi.org/10.1103/PhysRevE.91.042701

[22] M. Ren, Q. Zhang, and J. Zhang, "AN INTRODUCTORY SURVEY OF PROBABILITY DENSITY FUNCTION CONTROL," *Syst. Sci. Control Eng.*, vol. 7, no. 1, page. 158–170, 2019. doi: https://doi.org/10.1080/21642583.2019.1588804

[23] D. Koletsi and N. Pandis, "SURVIVAL ANALYSIS PART 2:KAPLAN-MEIER METHOD AND THELOG-RANK TEST," *Stat. Res. Des.*, vol. 152, no. 4, 2017. doi: https://doi.org/10.1016/j.ajodo.2017.07.008

[24] I. Firsawan, N. N. Debataraja, and S. W. Rizki, "ANALISIS SURVIVAL PADA DATA TERSENSOR TIPE I DENGAN METODE KAPLAN-MEIER," *Bimaster Bul. Ilm. Mat. Stat. and Ter.*, vol. 11, no. 1, page. 19–26, 2022.

[25] J. Zeea and S. X. Xieb, "THE KAPLAN-MEIER METHOD FOR ESTIMATING AND COMPARING PROPORTIONS IN A RANDOMIZED CONTROLLED TRIAL WITH DROPOUTS," *HHS Pub lic Access*, vol. 2, no. 1, page. 23–33, 2018. doi: https://doi.org/10.1080/24709360.2017.1407866

[26] P. Schober, MMedStat, and Thomas, "KAPLAN-MEIER CURVES, LOG-RANK TESTS, AND COX REGRESSION FOR TIME-TO-EVENT DATA," *anesrhesia analg.*, vol. 132, no. 4, 2021. https://doi.org/10.1213/ANE.0000000000005358

[27] W. W. Cahyani and F. Rakhmawati, "ANALISIS SURVIVAL MENGGUNAKAN REGRESI WEIBULL PADA LAJU KESEMBUHAN PASIEN JANTUNG KORONER," *Sains and Mat.*, vol. 9, no. 2, page. 39–45, 2024.

[28] R. Demicheli, W. Hrushesky, M. Retsky, and E. Biganzoli, "INTERPRETING BREAST CANCER SURVIVAL DATA BY THE HAZARD FUNCTION: REMARKABLE FINDINGS FROM EVENT DYNAMICS," *Medicina (B. Aires).*, vol. 56,

no. 9, 2020. doi: https://doi.org/10.3390/medicina56090468

[29] I. Kuitunen, V. T. Ponkilainen, M. M. Uimonen, A. Eskelinen, and A. Reito, "TESTING THE PROPORTIONAL HAZARDS ASSUMPTION IN COX REGRESSION AND DEALING WITH POSSIBLE NON-PROPORTIONALITY IN TOTAL JOINT ARTHROPLASTY RESEARCH: METHODOLOGICAL PERSPECTIVES AND REVIEW," *Kuitunen al. BMC Musculoskelet. Disord.*, vol. 22, page. 1–7, 2021. doi: https://doi.org/10.1186/s12891-021-04379-2

[30] D. W. Vermunt, V. D. Schmittmann1, and Jeroen K. Vermunt, "STATISTICAL POWER OF LIKELIHOOD RATIO AND WALD TESTS IN LATENT CLASS MODELS WITH COVARIATES," *Behav Res*, vol. 49, page. 1824–1837, 2017. doi: https://doi.org/10.3758/s13428-016-0825-y.