Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:77
https://doi.org/10.28919/cmbn/9264
ISSN: 2052-2541

SOCIO-DEMOGRAPHIC AND CLINICAL FACTORS IN STROKE PATIENTS: A SURVIVAL ANALYSIS APPROACH

SUWARDI ANNAS[1], ASWI ASWI[1,*], IRWAN[2], MUTH HAIR[1], MARDATUNNISA ISNAINI[1], BOBBY POERWANTO[1], SERIFAT ADEDAMOLA FOLORUNSO[3]

[1]Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

[2]Department of Mathematics, Universitas Negeri Makassar, Makassar, Indonesia

[3]Department of Statistics, University of Ibadan, Nigeria

Abstract: Stroke is a serious medical condition that occurs when blood flow to a part of the brain is either disrupted or cut off. It is one of the leading causes of death worldwide and is classified into two main types: hemorrhagic (caused by bleeding) and ischemic (caused by blood vessel blockage). This study aims to identify the factors significantly influencing the rate of clinical improvement in stroke patients at RSKD DADI Hospital, Makassar, South Sulawesi Province. The analysis was conducted using the Extended Cox Regression model, with the dependent variable being the length of hospitalization until the patient was declared improved or discharged. The independent variables included gender, age, hypertension, cholesterol levels, diabetes mellitus, type of stroke, and uric acid levels. The results indicated that the type of stroke did not meet the proportional hazards assumption, suggesting that the effect of stroke type varies over time. Of the factors examined, patient age, cholesterol level, and type of stroke were found to significantly impact the rate of clinical improvement in stroke patients at RSKD DADI Hospital, South Sulawesi Province.

Keywords: stroke; survival; extended cox.

2020 AMS Subject Classification: 62P10.

---

*Corresponding author
E-mail address: aswi@unm.ac.id
Received March 27, 2025

1. INTRODUCTION

Stroke is a major cardiovascular disease involving the heart and vascular system. According to 2019 data from the Institute for Health Metrics and Evaluation (IHME), stroke ranks as the leading cause of death in Indonesia, accounting for 19.42% of all fatalities. The Riskesdas report highlights a significant increase in stroke prevalence, rising by 56% from 7 cases per 1,000 people in 2013 to 10.9 cases per 1,000 people in 2018 [1]. Several studies have examined factors influencing stroke. A study aimed to evaluate both short-term and long-term health outcomes in patients with acute ischemic stroke and to analyze factors associated with survival in Oman. All patients admitted with acute ischemic stroke from January 1, 2017, to August 31, 2018, were included. Data were analyzed using multivariate logistic regression, Cox proportional hazards, and Kaplan-Meier methods, which indicated that acute ischemic stroke was generally associated with a high burden of comorbidities [2]. A study examined survival outcomes and mortality predictors among adult stroke patients admitted to Jimma University Medical Center in Southwest Ethiopia from April 1, 2017, to March 31, 2022. The analysis utilized a Cox regression model to identify factors associated with survival in stroke patients. Findings indicated that stroke-related mortality rates were notably higher at the seven- and fourteen-day marks. Key independent predictors of mortality included the Glasgow Coma Scale score, increased intracranial pressure, dyslipidemia, and aspiration pneumonia [3]. Other research evaluated long-term trends in stroke mortality linked to high body mass index (BMI) in China from 1990 to 2019. Using data from the Global Burden of Disease (GBD) 2019, age-standardized mortality rate (ASMR) trends were analyzed through linear regression and an age-period-cohort model. The findings reveal a rise in ASMR for stroke and its subtypes due to high BMI, with variations by subtype, gender, and age. These trends highlight the need for targeted public health guidelines in China to address stroke risks associated with high BMI across specific subtypes, genders, and age groups [4].

A Weibull parametric survival model was employed to evaluate the relationship between patient survival and outcomes in brain stroke (BS) while accounting for competing risks. This study was conducted at Imam Khomeini Hospital in Ardabil, Iran [5]. Parametric survival analysis was employed to identify factors influencing the length of hospital stay among suspected stroke patients transported via emergency services to Qaem Hospital in Mashhad, the largest city in northeastern Iran. After evaluating Accelerated Failure Time (AFT) models, including Log-normal, Log-logistic, Exponential, and Weibull distributions, the Log-normal model was determined to be the most appropriate. Key factors significantly affecting the length of stay included admission priority, insurance coverage, season of admission, and residency status [6].

While the study was conducted outside Indonesia, several related studies on stroke have also been carried out within Indonesia. A study investigated predictors of in-hospital mortality among ischemic stroke (IS) patients using medical records from the National Brain Centre Hospital in Jakarta, Indonesia. Employing Cox regression and Fine-Gray models while adjusting for age and sex, the researchers calculated hazard ratios to assess the impact of each risk factor. The study identified male sex, NIH Stroke Scale (NIHSS), uric acid levels, cardiovascular disease, pneumonia, sepsis, BMI, and Glasgow Coma Scale (GCS) on admission as key determinants of in-hospital mortality among IS patients [7]. A study examined the link between socioeconomic status (SES) and stroke severity in Indonesia, focusing on patients diagnosed at the National Brain Centre Hospital in Jakarta in 2020. SES was determined by factors such as marital status, occupation, education, payment source, and hospital class, with smoking status and sex as confounders. The findings indicated that patients with higher SES had a lower risk of severe stroke compared to those with lower SES, emphasizing the importance of SES improvement in stroke management.

A parametric Weibull parametric survival analysis has been applied to identify factors influencing stroke patient recovery. The analysis utilized data from stroke patients treated at Sultan Daeng Raja Hospital in Bulukumba, South Sulawesi, Indonesia, in 2018 [8]. The optimal model developed in this study identified two significant predictor variables: age and BMI. Thus, age and BMI were determined to be critical factors affecting stroke recovery. Several related studies on stroke have also been carried out in Makassar City, Indonesia. A study conducted at Dadi Hospital, Makassar City, aimed to identify significant factors in classifying stroke types namely, non-hemorrhagic and hemorrhagic strokes with stroke type as the response variable. The predictor variables included cholesterol levels, blood sugar levels, body temperature, length of hospitalization, pulse rate, and gender. Using logistic regression analysis, the study found that cholesterol levels and length of hospitalization significantly affected stroke classification [9]. By Using Kernel Logistic regression, it was found that four predictors affected the type of stroke significantly, namely cholesterol level, temperature, length of stay, and disease history [10]. Another study identified that a binary logistic regression model incorporating medical history and blood sugar levels was the more appropriate model, as it had the lowest Akaike Information Criterion (AIC) value and statistically significant covariates. Medical history and blood sugar levels were shown to significantly influence stroke type [11]. A study aimed to classify stroke types based on six predictors: age, gender, cholesterol levels, length of stay (LOS), medical history, and blood sugar levels. Using a resilient backpropagation neural network (RBNN), selected for its suitability to the data, the classification model achieved optimal performance with 5 nodes in the hidden layer, yielding a low error rate of 9.26 and a strong overall prediction accuracy [12].

The literature indicates that semiparametric Extended Cox survival analysis has not yet been widely applied to investigate key factors affecting stroke recovery rates in the Makassar region. Unlike logistic regression, which solely models the occurrence of an event [13], the Extended Cox survival model incorporates the timing of the event, providing more detailed insights. Additionally, it effectively handles censored data, where the exact time of the event is unknown for some subjects. The "Extended" feature enhances the model's flexibility by allowing the inclusion of time-varying covariates or effects, addressing scenarios where the proportional hazards assumption is violated, thereby increasing its applicability compared to the standard Cox model [14]. This study aims to (1) describe the general characteristics of stroke patients hospitalized at one hospital in Makassar city, namely, RSKD Dadi Makassar, (2) obtain a survival model for stroke patient outcomes at RSKD Dadi Makassar, and (3) identify significant factors influencing the rate of clinical improvement in stroke patients hospitalized at RSKD Dadi Makassar.

2. METHODS

2.1. DATA

This study used data from medical records of inpatient stroke patients at Dadi Hospital, a stroke center, in 2022. The dataset originally included 440 stroke patients, but 37 cases with incomplete clinical information were excluded, leaving 403 cases for analysis. Table 1 provides the operational definitions of the study variables.

Table 1. Research Variables

Variable Name	Scale	Description
Survival Time	Ratio	The duration of a stroke patient is hospitalized until discharge, with a status (S): 1 if improved, 0 if deceased.
Gender (X_1)	Nominal	0 = Female 1 = Male
Patient Age (X_2)	Ratio	Age of the patient at the time of stroke.

Hypertension (X_3)	Nominal	0 = No hypertension
		1 = hypertension
Cholesterol (X_4)	Nominal	0 = No high cholesterol
		1 = High cholesterol
Diabetes Mellitus (X_5)	Nominal	0 = No diabetes mellitus
		1 = Diabetes mellitus
Type of Stroke (X_6)	Nominal	1 = Ischemic stroke
		0 = Hemorrhagic stroke
Urid Acid (X_7)	Nominal	0= No elevated uric acid
		1= Elevated uric acid

2.2. SURVIVAL ANALISIS

Survival analysis is a statistical method that can be employed to identify factors influencing the rate of stroke recovery. This technique focuses on analyzing variables that affect the occurrence and timing of events, such as recovery or death, over a measured period days, weeks, months, or years. In this context, an initial event might involve the onset of disease, while the final event could include either patient recovery or death [14]. This approach aids in understanding recovery dynamics in stroke patients and in identifying key factors that impact outcomes.

2.3. RESEARCH PROCEDURE

The research procedure involves several key steps to analyze stroke patient survival. First, it includes describing patient characteristics based on factors likely to influence survival outcomes. Next, the distribution of the dependent variable, survival time, is examined. The proportional hazard assumption is then tested, either through a global test or Goodness of Fit (GOF) analysis, to identify any variables that violate this assumption. If any do, an Extended Cox regression model is constructed to account for these variables. The model parameters are estimated and tested for significance, after which the best Extended Cox model is selected. Finally, the hazard ratio is interpreted to assess the impact of each factor on survival rates.

3. MAIN RESULTS

3.1. DESCRIPTIVE ANALYSIS

The hospitalization period for stroke patients ranged from 1 to 23 days, with a mean length of stay of 6 days and a standard deviation of 2.95 days. Patient age varied from 23 to 92 years, with a mean of 59 years and a standard deviation of 59.4 years. Among 403 stroke cases recorded at

RSKD DADI, South Sulawesi Province, in 2022, 349 patients showed improvement, while 54 either died or were referred to other facilities. Male patients constituted 52% of the total cases. A history of hypertension was present in 35% of patients, while 28% had hypercholesterolemia, 20% had diabetes, and 3% had elevated uric acid levels. Ischemic stroke accounted for 87% of cases, whereas hemorrhagic stroke comprised the remaining 13%.

3.2. Checking the Distribution of Survival Time Data

Checking the distribution of survival time data is crucial in survival analysis, particularly for selecting the most suitable model. By verifying that the data conforms to a specific distribution, researchers can choose an optimal model that enhances the accuracy of the analysis. This distributional pattern also guides the selection of an appropriate modeling approach whether parametric, semi-parametric (such as the Cox proportional hazards model), or non-parametric. Based on distributional fit tests conducted on the survival time data, the findings indicate that the data does not conform to the Weibull, Gamma, Lognormal, or Exponential distributions. Consequently, the analysis can proceed using the semi-parametric Cox model, which does not require a specified survival time distribution.

3.3. Testing the Proportional Hazards Assumption

The fundamental assumption in the Cox Proportional Hazards (PH) regression model is that the hazard ratio between groups remains constant over time, meaning that the effect of each covariate on the hazard rate does not vary with time. To assess this PH assumption, the Schoenfeld residuals test is commonly employed, both for individual covariates and overall model fit. The results of the proportional hazards assumption test for each covariate are presented in Table 2.

Table 2. The results of the proportional hazards assumption test for each covariate

Variables	χ^2	*p-value*
Gender (X_1)	0.130	0.718
Patient Age (X_2)	0.149	0.699
Hypertension (X_3)	0.011	0.914
Cholesterol (X_4)	0.016	0.898
Diabetes Mellitus (X_5)	1.336	0.247
Type of Stroke (X_6)	9.423	0.002
Urid Acid (X_7)	0.544	0.460
GLOBAL	13.21	0.067

Table 2 indicates that all variables meet the proportional hazards assumption, except for the stroke type variable (X_6), which has a p-value of 0.002. In semi-parametric Cox survival analysis, if any variables fail to satisfy the proportional hazards assumption, several adjustments can be implemented to ensure model validity and enhance result interpretation accuracy. One approach involves introducing interactions between the variables that violate the assumption and a time function. The Extended Cox model accommodates time-dependent effects for specific variables by incorporating interactions between these variables and the time function (covariate × time).

3.4. Parameter Estimation of the Extended-Cox Regression Model

Results of the Cox regression model parameter estimation are given in Table 3.

Table 3. Results of the Cox Regression Model Parameter Estimation

Variables	($\hat{\beta}$)	exp($\hat{\beta}$)	*P-value*	Interpretation
Gender (X_1)	-0.1362	0.8727	0.214	Not Significant
Patient Age (X_2)	-0.0159	0.9843	0.002	Significant
Hypertension (X_3)	-0.0143	0.9858	0.9013	Not Significant
Cholesterol (X_4)	0.3255	1.3850	0.0087	Significant
Diabetes Mellitus (X_5)	-0.1500	0.8607	0.2755	Not Significant
Type of Stroke (X_6)	18.980	1.756e+08	<2e-16	Significant
Urid Acid (X_7)	-0.1708	0.8430	0.5543	Not Significant
Type of Stroke (X_6)*g(t)	-8.5750	0.00018	<2e-16	Significant

Table 3 indicates that age, cholesterol level, stroke type, and the interaction between stroke type and a time-dependent function $g(t)$ are significantly associated with survival time. Age significantly impacts survival, with each additional year of age corresponds to a roughly 1.57% decrease in the hazard of discharge (calculated as 1 - 0.9843). This finding implies that older patients tend to have a slightly lower likelihood of discharge. High cholesterol levels are also a significant predictor, raising the hazard by approximately 38.5%, suggesting that patients with high cholesterol levels are more likely to be discharged sooner. Stroke type proves to be a highly significant factor. Patients with ischemic stroke have vastly higher odds of survival than those with hemorrhagic stroke. Lastly, the significant interaction between stroke type and the time-dependent function g(t) suggests that the influence of stroke type on hazard changes over time, with ischemic stroke patients showing a decreasing hazard as time progresses.

3.5. Selection of the Best-Fit Model

The optimal model was selected using the backward elimination method, beginning with the

Complete Model (referred to as Model 1), which includes all available independent variables and has an AIC value of 3760.81. In cases where a variable has a p-value exceeding the predetermined significance level (α=0.05), it is considered insignificant and removed from the model. Typically, the variable with the highest p-value (i.e., the least significant) is removed first. Based on Table 3, variable X_3 has the highest p-value (0.9013). Model 2, which excludes variable X_3, has an AIC value of 3758.828. Model 3 (excluding X_3 and X_7) has an AIC of 3757.209, while Model 4 (excluding X_1, X_3, and X_7) has an AIC of 3756.346. The final model, Model 5 (excluding X_1, X_3, X_5, and X_7), achieves the lowest AIC value of 3755.785.

To select the best model, the model with the lowest AIC value is preferred. Among the five models generated, Model 5 is identified as the optimal model. Results of the Cox Extended-Survival Model Parameter Estimation for the best model which the incorporation of patient age (X_2), cholesterol (X_4), type of stroke (X_6), and type of stroke (X_6)*g(t) are given in Table 4.

Table 4. Results of the Extended Cox Regression Model Parameter Estimation for selected model

Variables	$(\hat{\beta})$	$\exp(\hat{\beta})$	*P-value*
Patient Age (X_2)	-0.0153	0.9849	0.00361
Cholesterol (X_4)	0.3231	1.3810	0.00674
Type of Stroke (X_6)	18.92	1.647e+08	< 2e-16
Type of Stroke (X_6)*g(t)	-8.556	0.00019	< 2e-16

Extended Cox Regression Model can be written as follows:

$$\hat{h}(t, x(t)) = \widehat{h_0}(t)\exp(-0.0153\text{Age} + 0.3231\text{Cholesterol} + 18.92\text{type of Stroke} - 8.556\text{type of Stroke} * \text{g(t)})$$

4. CONCLUSIONS

The Survival Semiparametric Extended-Cox Regression Model was chosen as the preferred approach to evaluate factors significantly affecting the rate of clinical improvement in stroke patients at RSKD DADI Hospital, Makassar, South Sulawesi Province. Key covariates included in the model are patient age, cholesterol level, type of stroke, and an interaction term, type of stroke * g(t). Results indicate that older age is associated with a slightly reduced likelihood of discharge. Elevated cholesterol is also a notable predictor, increasing the hazard by approximately 38.5%, which implies that patients with higher cholesterol levels tend to be discharged earlier. Stroke type is particularly impactful, with ischemic stroke patients showing considerably higher odds of survival compared to those with hemorrhagic stroke. The significant interaction between stroke type and the time-dependent function g(t) further suggests that the effect of stroke type on hazard is time-dependent, with ischemic stroke patients experiencing a reduction in hazard over time. Future research could extend this model by incorporating additional data from other hospitals in Makassar and by including spatial random effects to enhance the model's robustness.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Universitas Negeri Makassar. This work was funded by DIPA Universitas Negeri Makassar Number: SP DIPA-023.17.2.677523/2024, dated November 24, 2023, according to the Decree of the Rector of Universitas Negeri Makassar Number: 261/UN36/HK/2024 dated February 7, 2024.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

REFERENCES

[1] Kemenkes, Kenali Stroke dan Penyebabnya, (2023). https://ayosehat.kemkes.go.id/kenali-stroke-dan-penyebabnya.

[2] A.M. Al Alawi, I. Al Busaidi, E. Al Shibli, et al. Health Outcomes After Acute Ischemic Stroke:retrospective and Survival Analysis From Oman, Ann. Saudi Med. 42 (2022), 269-275. https://doi.org/10.5144/0256-4947.2022.269.

[3] W. Mosisa, Y. Gezehagn, G. Kune, et al. Survival Status and Predictors of Mortality Among Adult Stroke Patients Admitted to Jimma University Medical Center, South West Ethiopia: a Retrospective Cohort Study, Vasc. Health Risk Manag. 19 (2023), 527-541. https://doi.org/10.2147/vhrm.s399815.

[4] Y. Gao, K. Liu, S. Fang, Trend Analysis of Stroke Subtypes Mortality Attributable to High Body-mass Index in China From 1990 to 2019, BMC Public Health 24 (2024), 2155. https://doi.org/10.1186/s12889-024-19615-2.

[5] S. Norouzi, R. Fallah, A. Pourdarvish, et al. Survival Analysis of Patients with Brain Stroke in the Presence of Competing Risks: a Weibull Parametric Model, J. Biostat. Epidemiol. 7(2021), 235-243. https://doi.org/10.18502/jbe.v7i3.7295.

[6] Y. Razieh, S. Payam, N. Eisa, et al. Survival Analysis of the Length of Hospital Stay of Suspected Stroke Patients Transferred by EMS to Ghaem Hospital in Mashhad, Med. J. 81 (2024), 886-898.

[7] N. Yamanie, Y. Felistia, N.H. Susanto, et al. Prognostic Model of In-Hospital Ischemic Stroke Mortality Based on an Electronic Health Record Cohort in Indonesia, PLOS ONE 19 (2024), e0305100. https://doi.org/10.1371/journal.pone.0305100.

[8] R. Fajar, Parametric Survival Analysis in Stroke Patients with Weibull Distribution Approach, J. Neurol. Sci. 405 (2019), 130-131. https://doi.org/10.1016/j.jns.2019.10.684.

[9] S. Annas, A. Aswi, M. Abdy, B. Poerwanto, Stroke Classification Model Using Logistic Regression, J. Phys.: Conf. Ser. 2123 (2021), 012016. https://doi.org/10.1088/1742-6596/2123/1/012016.

[10] S. Annas, B. Poerwanto, A. Aswi, et al. Classification Model For Type of Stroke Using Kernel Logistic Regression, Commun Math Biol Neurosci. 2022 (2022). https://doi.org/10.28919/cmbn/7752.

[11] S. Annas, A. Aswi, M. Abdy, B. Poerwanto, Binary Logistic Regression Model of Stroke Patients: a Case Study of Stroke Centre Hospital in Makassar, Indones. J. Stat. Appl. 6 (2022), 161-169. https://doi.org/10.29244/ijsa.v6i1p161-169.

[12] S. Annas, A. Aswi, M. Abdy, B. Poerwanto, R.Y. Fa'rifah, Stroke Type Classification Model Based on Risk Factors Using Resilient Backpropagation Neural Networks, AIP Conf. Proc. 2977 (2023), 060006. https://doi.org/10.1063/5.0181745.

[13] D.G. Kleinbaum, M. Klein, Logistic Regression A Self-Learning Text, Springer, New York, 2010.

[14] D.G. Kleinbaum, M. Klein, Survival Analysis: A Self-Learning Text, Springer, New York, 2012.