Journal of Mathematics and Scientific Computing with Applications

E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 7, Number 1, 2026
DOI: 10.53806/jmscowa.v7i1.1550
Page: 106–113

COMPARATIVE EVALUATION OF THE PRENTICE WILLIAMS PETERSON METHOD AGAINST COX PH IN RECURRENT STROKE SURVIVAL MODELLING

Atika Nur Royyanah[1], **Jerry Dwi Trijoyo Purnomo**[2], **Shofi Andari**[3]

[1,2,3]Department of Statistics, Institut Tekonologi Sepuluh Nopember, Surabaya, 60111, Indonesia

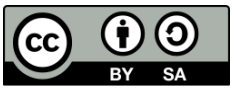
Article Info

Article history:
Received : June 6, 2026
Revised : June 26, 2026
Accepted : July 11, 2026

Keywords:
Cumulative Hazard Curve;
PWP Model;
Recurrent Events Stroke

ABSTRACT

The global burden of recurrent stroke, reflected in increased mortality, long-term disability, and healthcare costs, demands accurate risk prediction. This study compares the effectiveness of the Prentice-Williams-Peterson (PWP) models with the standard Cox Proportional Hazards model in analyzing recurrent stroke events, using 10-month data from 353 patients with a total of 429 recurrent events at a hospital in Surabaya. The objective is to identify which framework better captures the risk of sequential events based on information criteria. The PWP-Total Time (PWP-TT) model proved to be the most optimal approach, yielding an AIC of 4187.686 (compared to 5042.718 for Cox PH and 4252.759 for PWP-GT) and a BIC of 4219.691 (compared to 4915.213 for Cox PH and 4285.292 for PWP-GT). Based on the PWP-TT as the best-fitting model, the smoking variable was found to influence the response variable, namely the time until a patient experiences a stroke.



Corresponding Author:
Atika Nur Royyanah,
Department of Statistics,
Institut Tekonologi Sepuluh Nopember
Email: atikanur.ryynh@gmail.com

1. INTRODUCTION

Survival analysis models the time until a specific event occurs, such as death, relapse, or failure; its defining feature is censoring, where the event time is not fully observed because the subject survives past the study period [1]. Recurrent event data extends this framework to cases where a subject experiences repeated events, such as recurrent stroke, infection, or asthma relapse [2], [3]. The Cox Proportional Hazards model [4] is the most widely used semi-parametric survival model, valued for not requiring a specified baseline hazard distribution. However, it models only the time to the first event; subjects with multiple events are analyzed based solely on their initial occurrence [5]. For recurrent conditions such as stroke, this discards information from subsequent events and can bias risk estimates [5].

Extensions address this limitation differently. The Andersen–Gill model treats each event as a separate observation through a counting process but assumes independence between events [4]. The Wei–Lin–Weissfeld model fits each event sequence marginally but does not capture within-subject correlation. The Prentice–Williams–Peterson (PWP) model (1981) [6] instead stratifies risk by event order, using either total time since baseline (TT) or gap time between events (GT), allowing risk to vary across event sequence for example, a higher risk after a second stroke than a first [7],

Stroke is the second leading cause of global mortality, with over 12 million new cases and 6.5 million deaths in 2019 [8]. In Indonesia it is the leading cause of death, with a national prevalence of 10.9 per 1,000 residents. Cumulative recurrence risk reaches 30–40% within five years [9], [10], driven by hypertension, diabetes, hypercholesterolemia, and low adherence to secondary prevention [11].

Indonesian research on stroke risk factors has relied largely on first-event models logistic regression and standard Cox PH applied to conditions such as Type II Diabetes Mellitus [12], Dengue Hemorrhagic Fever [13], and non-clinical settings using stratified Cox models with frailty [14]. Therefore, the contribution of this study lies in its methodological application rather than methodological innovation, as it applies the PWP-TT and PWP-GT models within a clinical context constrained by data limitations, and evaluates which model performs better using predetermined metrics (AIC and BIC).

This study has three objectives: (1) estimate and test parameters within the PWP-TT framework; (2) compare PWP-TT, PWP-GT, and Cox PH in analyzing recurrent stroke events using AIC and BIC; and (3) apply these models to Indonesian recurrent stroke data to identify significant predictors of recurrence. The results are intended to contribute to survival analysis methodology and inform secondary stroke prevention strategies.

2. RESEARCH METHOD

This section outlines the methodological framework employed to achieve the research objectives and ensure the reproducibility of the study. A robust quantitative design was adopted to analyze the complex underlying patterns within the data, utilizing established statistical frameworks alongside advanced modeling techniques. The following subsections detail the research design, data sources, collection procedures, analytical formulations, and software tools utilized to process and validate the findings systematically.

2.1. Research Design

This study employs a quantitative research design based on real-world medical record data, without data simulation. This design was selected because it enables an objective comparison of the effectiveness of two variants of the PWP model in handling recurrent events characterized by event-order dependence. Eligibility criteria were applied to preserve data integrity and minimize selection bias. Subjects were included if they were inpatients with a primary diagnosis of stroke recorded in the medical records of a hospital in Surabaya between January 1 and October 30, 2024; experienced at least one stroke event and a maximum of three recurrences during the observation period, with trackable recurrent events; and had complete data on stroke onset dates and follow-up periods sufficient to determine the exact time between events. Patients were excluded if they had incomplete data on admission date, stroke onset, or follow-up period that precluded calculation of survival time; exhibited inconsistencies, duplications, or unverifiable errors in their medical records; or had stroke events recorded on the exact same date that could not be clinically distinguished as separate events.

2.2. Data Source and Variables

Table 1: Number of Patients in Each Event Stratum

1 times event	2 times event	3 times event	4 times event	5 times event	Total
288 patients	54 patients	11 patients	5 patients	3 patients	361 patients

Data collection was conducted at a hospital in Surabaya, East Java, Indonesia. A rigorous de-identification protocol was applied during data extraction: all personal identifiers, including patient names, full addresses, and national identification numbers, were entirely removed from the dataset. Each subject was represented by a randomized alphanumeric code (Patient ID) already present in the medical records, ensuring that individual identities could not be traced. From an initial cohort of 361 patients identified during the observation period, the distribution of stroke events per patient was as follows: 288 patients (79.8%) experienced a single event, 54 patients (15.0%) experienced two events, 11 patients (3.0%) experienced three events, 5 patients (1.4%) experienced four events, and 3 patients (0.8%) experienced five events (Table 1). A marked decline in the number of patients was observed beyond the third event, with the risk set for the fourth and fifth event strata comprising only 5 and 3 patients, respectively. Such small subgroups are insufficient to support stable parameter estimation in stratified recurrent-event models such as PWP-TT and PWP-GT, where hazard ratio estimates for higher-order strata can become unreliable when the underlying risk set is very small [15]. Accordingly, patients experiencing more than

three events ($n = 8$, 2.2%) were excluded, and the analytic scope was restricted to a maximum of three events per subject (initial event through second recurrence). This threshold retained the substantial majority of the cohort (353 patients, 97.8%) while excluding the sparsest and least stable event strata. The final analytic sample therefore comprised 353 patients.

The data used are secondary medical records from a hospital, covering a 10-month observation period from January through October 30, 2024. During the study period, 782 records were collected, including both stroke events and censored observations, originating from 353 distinct stroke patients, ranging from a single occurrence to a second recurrence (third event). Among these subjects, 18% were hospitalized more than once, indicating recurrent stroke events, with the number of recurrences per patient ranging from one to two (up to three total events). The unit of observation was defined as the time until a patient was diagnosed with an initial or recurrent stroke. The independent variables included in the model were sex, patient age, comorbidity history (hypertension, coronary heart disease, and diabetes mellitus), adherence to routine medical check-ups, and smoking history status. Time-to-event was recorded in days, calculated from baseline observation until the occurrence of the event or right censoring.

We acknowledge that the 10-month observation window (January–October 2024) is a limitation of this study. Stroke recurrence patterns typically manifest over a period of several years; a shorter observation period may therefore limit the model's ability to capture long-term recurrence trajectories or late-onset events. The findings of this study consequently reflect short- to medium-term recurrence dynamics within the sampled population, and future research with a multi-year follow-up period is warranted to validate longer-term patterns. With 7 predictor variables and 429 recorded events, the events-per-variable (EPV) ratio in this study is 61.3. Based on the rule-of-thumb minimum of 10 EPV for Cox-based models and their extensions [16], this ratio substantially exceeds the adequacy threshold, indicating that coefficient estimate reliability is not compromised, including for variables with relatively low prevalence such as coronary heart disease or diabetes mellitus, even if their proportion in the sample is small.

2.3. Data Collection Procedure

Data were collected through a medical record extraction method that captured the chronic history of each subject. This procedure ensured that the exact sequence of events, along with corresponding covariates, was accurately recorded. Where specific data points were initially missing, the required information was recovered by cross-referencing other fields or supplementary files within the patient's medical record under the same registration number. The dataset was subsequently structured into a longitudinal/counting-process format to accommodate the recurrent event data structure.

2.4. Analytical Methods or Algorithms

The analytical framework involves a comparative analysis between two primary statistical modeling approaches:

a. PWP-GT (Gap Time) Model: Analyzes the risk of event occurrence based on the time elapsed since the previous event (waiting time).
b. PWP-TT (Total Time) Model: Analyzes the risk based on the time elapsed since the start of the study (calendar time).

Parameter estimation for both models is executed using the Partial Likelihood method derived from their respective hazard functions. Model validation is performed by comparing the Akaike Information Criterion (AIC) values and Bayesian Information Criterion (BIC) with evaluating the consistency of the regression coefficients across each event stratum.

The baseline formulation for the hazard function of the PWP model for the k-th event is mathematically expressed as follows:

$$h_k\left(\mathrm{t}, \mathrm{X}, \beta\right) = \mathrm{h_{0k}}\left(\mathrm{t}\right) \exp\left(\beta_k^{T'} \mathrm{X}\right) \tag{1}$$

where $\mathrm{h}_{0k}(t)$ represents the stratum-specific baseline hazard function for the k-th event, X denotes the vector of explanatory covariates, and $\boldsymbol{\beta}_k$ represents the vector of regression coefficients for the k-th stratum.

Parameter estimation in the PWP model is performed using the partial likelihood method developed by Cox (1972) [17]. This procedure allows for estimation without explicitly defining the baseline hazard distribution. Once parameters are obtained, hypothesis testing is conducted using three primary approaches: the Wald test, Likelihood Ratio Test (LRT), and Score test, each offering different advantages depending on sample size and model complexity [5], [7]. Nevertheless, like the standard Cox model, PWP relies on the proportional hazards (PH) assumption. Violations of this assumption can lead to biased coefficient interpretations, necessitating diagnostic tests such as Schoenfeld residuals or stratified Kaplan–Meier curves [18]. Furthermore, the choice between PWP-TT and PWP-GT must consider the research context: PWP-TT is more suitable if event risk is influenced by

the total time from baseline, while PWP-GT is more appropriate when the focus is on the duration between recurrences.

All hazard ratios in the best model (HRs) reported in the results section are accompanied by 95% confidence intervals (CIs), calculated based on the standard errors of the partial likelihood coefficient estimates.

2.5. Software and Tools

All stages of data preprocessing, statistical modeling, and visualization were performed using the R programming language (R version 4.1.1 (2021-08-10)). The primary packages utilized include `survival` for the PWP model estimation, `dplyr` for data manipulation, and both `survminer` and `ggplot2` for hazard curve visualization and model assumption diagnostics.

2.6. Ethical Considerations

This study obtained formal ethical approval from the institutional review board (IRB) / ethics committee of the participating hospital. To ensure strict compliance with health research ethics standards and protect patient confidentiality, all subject identities were completely anonymized prior to analysis.

3. RESULT AND ANALYSIS

The primary objective of this analysis was to compare the performance of the Cox Proportional Hazards (Cox PH), PWP-Total Time (PWP-TT), and PWP-Gap Time (PWP-GT) models to identify the most accurate approach for modeling the research data. The selection of the superior model is based on the evaluation of information criteria.

Prior to conducting the model comparison analysis, this section presents descriptive statistics to provide an overview of the characteristics and distribution patterns of the data. Understanding this data distribution is essential as a fundamental step before evaluating the accuracy of the Cox PH, PWP-TT, and PWP-GT models in representing the research data. The summary of these data characteristics was as follows.

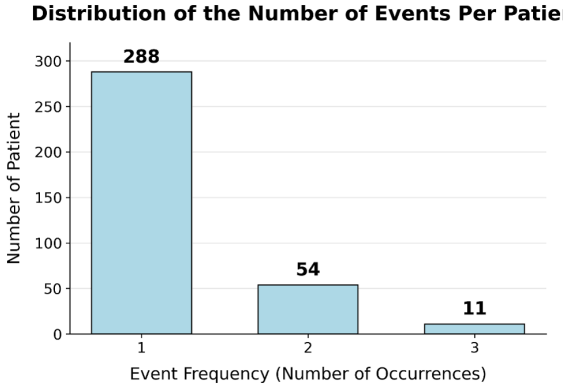


Figure 1. Event Frequency (Number of Occurences)

The distribution showed that the frequency of events per patient was highly skewed toward a single occurrence. However, since a total of 65 patients $(54 + 11)$ experienced more than one event, the use of recurrent event models such as the PWP-TT or PWP-GT models mentioned previously is statistically justified to capture the dependency and risk factors associated with these subsequent occurrences.

Table 2: Descriptive Characteristics of Patients by Event Order

Variables	Category	Number of 1st Event	Number of 2nd Event	Number of 3rd Event	Total
Gender	0 : Male	181	31	8	220
	1 : Female	107	23	3	133
Age	0 : Patient age < 50 years	126	16	6	148
	1 : Patient age ≥ 50 years	162	38	5	205
Hypertension	0 : No history of hypertension	115	20	4	139
	1 : Has history of hypertension	173	34	7	214
Coronary Heart Disease	0 : No history of coronary heart disease	272	52	9	333
	1 : Has history of coronary heart disease	16	2	2	20
Diabetes Mellitus	0 : No history of diabetes	64	13	4	81
	1 : Has history of diabetes	224	41	7	272
Control Compliance	0 : Never attended follow-up/control	90	36	9	135
	1 : Attended follow-up/control	198	18	2	218
Smoking Track	0 : Never smoked	228	44	10	282
	1 : Has smoked / Ever smoked	60	10	1	71

The data suggested that Control Compliance and Age were major factors associated with the frequency of recurrences. Non-compliant patients and those over 50 years old showed a higher tendency to progress from a first event to subsequent recurrences. This descriptive evidence supports the necessity of using the PWP models to further analyze how these risk factors specifically influence the ”gap time” or ”total time” between stroke recurrences. and recommendations for future research.

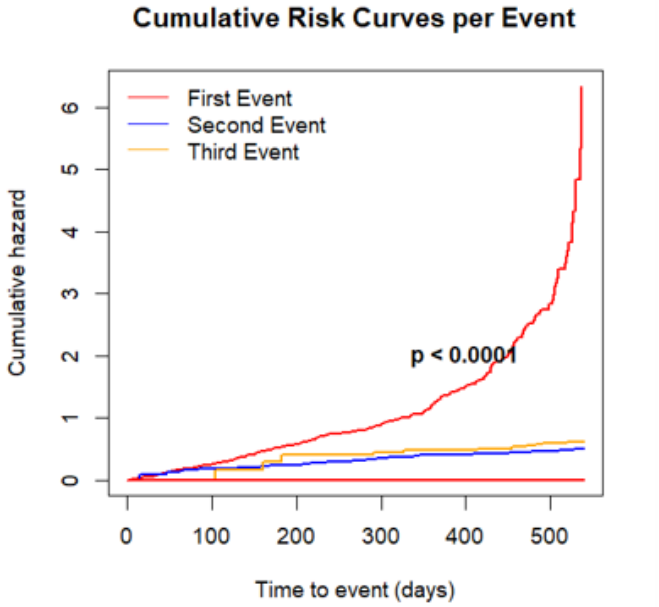


Figure 2. Cumulative Risk Curves per Event

The graph illustrated the Cumulative Hazard Curves for three sequential events over a period of approximately 550 days, revealing a stark contrast in risk patterns. The first event (red line) showed a moderate, steady increase in risk during the first 400 days, followed by a dramatic and rapid acceleration that reached a cumulative hazard of nearly 10 by the end of the study. In sharp contrast, the second (blue) and third (orange) events maintained a very low and stable trajectory, staying below a value of 1.0 throughout the entire duration. With a highly significant p-value of 0.0001, the data confirms that the occurrence of the first event is statistically much more frequent and carries a significantly higher risk over time compared to subsequent recurring events. It should be noted that the cumulative hazard estimate is not bounded by 1 and does not represent a probability. Unlike the survival function, which ranges between 0 and 1, the cumulative hazard can exceed 1 and in this study reaches values approaching 10 for the first-event stratum reflecting the accumulated instantaneous risk over the observation period rather than a direct probability of event occurrence.

Table 3: Comparison Across Cox PH, PWP-TT, and PWP-GT Models

Variables	**Cox PH**		**PWP-TT**		**PWP-GT**	
	HR	**p**	**HR**	**p**	**HR**	**p**
Age	0.98	0.55	0.94	0.580	0.95	0.58
Gender	1.03	0.63	0.87	0.199	0.87	0.20
Hypertension	1.02	0.68	0.95	0.670	0.96	0.70
Coronary Heart Disease	1.07	0.55	0.95	0.860	0.96	0.86
Diabetes Mellitus	0.93	0.49	0.87	0.264	0.88	0.26
Control Compliance	0.85	0.00*	0.77	0.038*	0.76	0.02*
Smoking	0.97	0.53	0.84	0.069	0.83	0.06

Based on the comparison among the Cox Proportional Hazards (PH), PWP-Total Time (TT), and PWP-Gap Time (GT) models, Control Compliance was the only variable that reached statistical significance across all three models (Cox PH: HR $= 0.85$, $p = 0.00^*$; PWP-TT: HR $= 0.77$, $p = 0.038^*$; PWP-GT: HR $= 0.76$, $p = 0.02^*$), with a consistently protective direction of effect throughout. Compliant patients showed approximately 15% to 24% lower risk of recurrence depending on the model, consistent with the protective role of routine medical follow-up reported in the secondary stroke prevention literature [8]. The remaining variables—Age, Gender, Hypertension, Coronary Heart Disease, and Diabetes Mellitus—did not show statistically significant effects in any of the three models.

Table 4: Comparison of Model Goodness of Fit Based on AIC dan BIC Values

Model	**AIC**	**BIC**
Cox PH	5042.718	4915.213
PWP-TT	4187.686	4219.691
PWP-GT	4252.759	4285.292

Based on the analysis of the model goodness-of-fit criteria, Table 4 illustrates the performance comparison between the Cox PH, PWP-TT, and PWP-GT models. In statistics, the Akaike Information Criterion (AIC) serves as a parameter to determine the best-fitting model, where a lower value indicates a more efficient model with minimal information loss.

The results demonstrated that the PWP-TT model was the most optimal for explaining this dataset, as it yielded the lowest AIC value of 4187.686 and BIC values of 4219.691. The PWP-GT (Gap Time) model ranked second with an AIC of 4252.759 and BIC of 4285.292, showing strong performance but still falling short of the PWP-TT model. In contrast, the Cox PH model recorded the highest AIC at 5042.718 and BIC at 4915.213, which suggests that this standard model is less capable of capturing the complexities or dependency structures between events as effectively as the PWP approaches.

Consequently, the PWP-TT model is recommended as the superior model and the most accurate for further analysis. Based on this justification, the subsequent analysis focuses exclusively on the PWP-TT framework to evaluate the specific risk factors influencing recurrent events. Through this optimized model, the estimation parameters are presented to identify which clinical and demographic variables hold statistically significant effects on the hazard rate. By interpreting these significant predictors, the study can precisely delineate how factors such as patient behavior and medical history accelerate or delay the onset of consecutive events over the total follow-up duration.

Table 5: Proportional Hazards Assumption Diagnostics for the Best Model

Variables	$\boldsymbol{\chi^2}$	**P-Value**
Age	0.0288	0.865
Gender	0.0215	0.883
Hypertension	0.1569	0.642
Coronary Heart Disease	0.5246	0.469
Diabetes Mellitus	0.4580	0.499
Control Compliance	0.3536	0.552
Smoking	0.3094	0.578

Results of the proportional hazards (PH) assumption diagnostic test for the best-fitting model, which is the PWP-TT model, selected based on the lowest AIC and BIC values, are presented in Table 5. The Schoenfeld residual test showed that all variables in the model, namely Age ($\chi^2 = 0.0288$, $p = 0.865$), Gender ($\chi^2 = 0.0215$, $p = 0.883$), Hypertension ($\chi^2 = 2.1569$, $p = 0.142$), Coronary Heart Disease ($\chi^2 = 0.5246$, $p = 0.469$), Diabetes Mellitus ($\chi^2 = 0.4580$, $p = 0.499$), Control Compliance ($\chi^2 = 0.3536$, $p = 0.552$), and Smoking ($\chi^2 = 0.3094$, $p = 0.578$), had p-values well above the 0.05 significance threshold. Therefore, there is no evidence that the effects of the variables in the model change significantly over the follow-up period, supporting the methodological validity of the hazard ratio interpretations reported for the PWP-TT model.

Table 6: Parameter Estimation for the PWP-TT Model

Variables	$\boldsymbol{\beta}$	**HR**	**95% CI**	**P-Value**
Age	-0.054	0.94	0.785 – 1.141	0.580
Gender	-0.133	0.87	0.709 – 1.079	0.199
Hypertension	-0.047	0.95	0.796 – 1.142	0.670
Coronary Heart Disease	-0.056	0.95	0.592 – 1.511	0.860
Diabetes Mellitus	-0.137	0.87	0.697 – 1.088	0.264
Control Compliance	-0.258	0.77	0.605 – 0.988	0.038*
Smoking	-0.170	0.84	0.702 – 1.103	0.069

The results of the PWP-TT model analysis show that among the seven variables examined, only control compliance was found to be statistically significant as a predictor of recurrent events ($\beta = -0.258$; HR = 0.77; 95% CI 0.605–0.988; $p = 0.038$). Subjects with good control compliance had approximately 23% lower risk of experiencing a recurrent event compared to those with poor compliance.

In contrast, the remaining six variables did not show statistically significant associations in this model: age (HR = 0.94; 95% CI 0.785–1.141; $p = 0.580$), gender (HR = 0.87; 95% CI 0.709–1.079; $p = 0.199$), hypertension (HR = 0.95; 95% CI 0.796–1.142; $p = 0.670$), coronary heart disease (HR = 0.95; 95% CI 0.592–1.511; $p = 0.860$), diabetes mellitus (HR = 0.87; 95% CI 0.697–1.088; $p = 0.264$), and smoking (HR = 0.84; 95% CI 0.702–1.103; $p = 0.069$). The confidence intervals for all six of these variables crossed 1, indicating insufficient evidence to conclude any effect on the risk of recurrent events in this study population.

Although smoking showed an HR below 1, this result was not statistically significant and its confidence interval crossed 1, so it cannot be interpreted as a protective effect of smoking against recurrent events. These findings confirm that among the risk factors examined, control compliance was the only factor to play a statistically significant role in reducing the risk of recurrent events.

4. CONCLUSION

This study aimed to identify a more appropriate statistical approach for modeling recurrent stroke events through parameter estimation and hypothesis testing. The results showed that the Prentice–Williams–Peterson Total Time (PWP-TT) model provided the best fit based on AIC and BIC values, yielding the lowest values (4187.686 and 4219.691) among the models compared. The PWP-Gap Time (PWP-GT) model also demonstrated reasonable performance (AIC = 4252.759, BIC = 4285.292) and remains a viable alternative when the analytical focus is specifically directed toward the gap time between successive events. However, based on the comparison of AIC and BIC values, PWP-TT proved to be the more appropriate framework for this dataset, thereby fulfilling the study's objective of determining the best statistical approach for modeling recurrent stroke events. The proportional hazards assumption for the best-fitting model (PWP-TT) was formally tested using Schoenfeld residuals for each covariate, and all variables showed non-significant results (all $p > 0.05$), indicating no evidence of assumption violation and supporting the methodological validity of the hazard ratio interpretations reported.

Substantively, the analysis using the PWP-TT model showed that among the seven variables examined, only control compliance was found to have a statistically significant association with the risk of stroke recurrence ($\beta = -0.258$; HR = 0.77; 95% CI 0.605–0.988; $p = 0.038$), with subjects who adhered to control measures having approximately 23% lower risk of recurrence compared to those who did not. In contrast, age, gender, hypertension, coronary heart disease, and diabetes mellitus did not show significant associations in this model.

These findings indicate that, within the context and modeling of this study, classic cardiovascular risk factors did not show a statistically significant association with stroke recurrence risk, while patient adherence to post-stroke control was the only factor statistically shown to play a significant protective role. This does not mean that these classic risk factors are clinically unimportant, as the wide confidence intervals observed for several variables suggest limited statistical power rather than a definitive absence of effect. Practically, these findings support the need to strengthen monitoring and education programs for post-stroke control compliance as part of secondary prevention strategies, given that this factor is the only modifiable predictor showing a significant protective effect in this study.

This study has several limitations, including the wide confidence intervals observed for some variables (e.g., gender, 95% CI: 0.709–1.079), which indicate a limited number of events in certain subgroups, resulting in less precise risk estimates for those variables. Further research with larger sample sizes and longer follow-up periods is needed to confirm the consistency of these findings and to explore interactions between control compliance and other risk factors in relation to stroke recurrence.

REFERENCES

[1] D. G. Kleinbaum and M. Klein, "Introduction to survival analysis," *Springer*, 2012, https://doi.org/10.1007/978-1-4419-6646-9_1.

[2] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, 2002.

[3] R. J. Cook and J. F. Lawless, *The Statistical Analysis of Recurrent Events*. Springer, 2007, https://doi.org/10.1007/978-0-387-69810-6.

[4] P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes: A large sample study," *The Annals of Statistics*, vol. 10, no. 4, pp. 1100–1120, 1982, https://doi.org/10.1214/aos/1176345976.

[5] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*. New York: Springer, 2000, https://doi.org/10.1007/978-1-4757-3294-8.

[6] R. L. Prentice, B. J. Williams, and A. V. Peterson, "On the regression analysis of multivariate failure time data," *Biometrika*, 1981.

[7] D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. John Wiley & Sons, 2008.

[8] D. O. Kleindorfer *et al.*, "2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack," *Stroke*, vol. 52, no. 7, 2021, https://doi.org/10.1161/STR.0000000000000375.

[9] V. L. Feigin *et al.*, "Global, regional, and national burden of stroke and its risk factors, 1990–2019," *The Lancet Neurology*, vol. 20, no. 10, pp. 795–820, 2021, https://doi.org/10.1016/S1474-4422(21)00252-0.

[10] V. L. Feigin, B. Norrving, and G. A. Mensah, "Global burden of stroke," *Circulation Research*, vol. 120, no. 3, pp. 439–448, 2017, https://doi.org/10.1161/CIRCRESAHA.116.308413.

[11] M. Irfan, M. Usman, S. Saidi, Warsono, D. Kurniasari, and Widiarti, "Survival analysis using cox proportional hazard regression approach in dengue hemorrhagic fever (dhf) case," *Journal of Physics: Conference Series*, vol. 1751, no. 1, p. 012011, 2021, https://doi.org/10.1088/1742-6596/1751/1/012011.

[12] K. M. Mohan *et al.*, "Risk and cumulative risk of stroke recurrence," *Stroke*, vol. 42, no. 5, pp. 1489–1494, 2011, https://doi.org/10.1161/STROKEAHA.110.602615.

[13] S. I. Lomo, Sugiyarto, and E. Darmawan, "A survival analysis with cox regression interaction model of type ii diabetes mellitus in indonesian," *Jurnal Profesi Medika*, vol. 15, no. 1, 2021, https://doi.org/10.33533/jpm.v15i1.2942.

[14] D. J. Ratnaningsih, A. Saefuddin, A. Kurnia, and I. W. Mangku, "Stratified-extended cox with frailty model for non-proportional hazard," *Thailand Statistician*, vol. 19, no. 1, pp. 209–228, 2021.

[15] L. D. A. F. Amorim and J. Cai, "Modelling recurrent events: A tutorial for analysis in epidemiology," *International Journal of Epidemiology*, vol. 44, no. 1, pp. 324–333, 2015.

[16] P. Peduzzi, J. Concato, A. R. Feinstein, and T. R. Holford, "Importance of events per independent variable in proportional hazards regression analysis ii," *Journal of Clinical Epidemiology*, vol. 48, no. 12, pp. 1503–1510, 1995, https://doi.org/10.1016/0895-4356(95)00048-8.

[17] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B*, vol. 34, no. 2, pp. 187–202, 1972, https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

[18] P. M. Grambsch and T. M. Therneau, "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, vol. 81, no. 3, pp. 515–526, 1994, https://doi.org/10.1093/biomet/81.3.515.