

Prognostic factors of first-ever stroke patients in suburban Malaysia by comparing regression models

Nadiah Wan-Arfah¹ , Mustapha Muzaimi² , Nyi Nyi Naing^{3*} , Vetriselvan Subramaniyan^{4,5} ,
Ling Shing Wong^{6**} , Siddharthan Selvaraj^{7,8} 

¹ Faculty of Health Sciences, Universiti Sultan Zainal Abidin, Kuala Terengganu, Terengganu, MALAYSIA

² Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, MALAYSIA

³ Faculty of Medicine, Medical Campus, Universiti Sultan Zainal Abidin, Jalan Sultan Mahmud, Kuala Terengganu, Terengganu, MALAYSIA

⁴ Pharmacology Unit, Jeffrey Cheah School of Medicine and Health Sciences, Monash University, Jalan Lagoon Selatan, Bandar Sunway, 47500 Selangor Darul Ehsan, MALAYSIA

⁵ Center for Transdisciplinary Research, Department of Pharmacology, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, 600077, INDIA

⁶ Faculty of Health and Life Sciences, INTI International University, Nilai, Negeri Sembilan, MALAYSIA

⁷ Faculty of Dentistry, AIMST University, Bedong, Kedah, MALAYSIA

⁸ Centre for Global Health Research, Saveetha Medical College and Hospital, Saveetha University, Chennai, Tamil Nadu, INDIA

*Corresponding Author: syedhatim@unisza.edu.my

**Corresponding Author: lingshing.wong@newinti.edu.my

Citation: Wan-Arfah N, Muzaimi M, Naing NN, Subramaniyan V, Wong LS, Selvaraj S. Prognostic factors of first-ever stroke patients in suburban Malaysia by comparing regression models. *Electron J Gen Med.* 2023;20(6):em545. <https://doi.org/10.29333/ejgm/13717>

ARTICLE INFO

Received: 14 Aug. 2023

Accepted: 30 Aug. 2023

ABSTRACT

Introduction: The aim of this study was to compare regression models based on the parameter estimates of prognostic factors of mortality in first-ever stroke patients.

Methods: A retrospective study among 432 first-ever stroke patients admitted to Hospital Universiti Sains Malaysia, Kelantan, Malaysia, was carried out. Patient's medical records were extracted using a standardized data collection sheet. The statistical analyses used for modelling the prognostic factors of mortality were Cox proportional hazards regression, multinomial logistic regression, and multiple logistic regression.

Results: A total of 101 (23.4%) events of death were identified and 331 patients (76.6%) were alive. Despite using three different statistical analyses, the results were very similar in terms of five major aspects of parameter estimates, namely direction, estimation, precision, significance, and magnitude of risk assessment. It was reported slightly better in Cox proportional hazards regression model, especially in terms of the precision of the results.

Conclusions: Given that this study had compared the findings from three different types of advanced statistical methods, this research has clearly yielded that with data of high quality, the selection of appropriate statistical method should not be a worrisome problem for researchers who may not be of expertise in the field of medical statistics.

Keywords: Cox proportional hazards regression, multinomial logistic regression, first-ever stroke, prognostic factors, health risk

INTRODUCTION

World health statistics reported that the leading causes of mortality among individuals around the globe were ischemic heart disease, lower respiratory infections followed by stroke. Years of life lost were calculated from the number of deaths at each age multiplied by the life expectancy for the age at which death occurred. For the last 12 years globally, the proportion of years of life lost due to non-communicable diseases has increased from 38% to 47%. Even though the overall recent trend has been successful in reducing mortality in communicable diseases, mortality due to non-communicable diseases has become the primary cause of death, with stroke as one of them [1].

Multivariable analysis is extremely important to statistically adjust the estimated effect of each variable in the

model and for more comprehensive statistical modelling. Cox proportional hazards regression model was widely used to determine the prognostic factors of mortality in stroke patients [2-6]. An alternative to Cox model reported in the literature was the binary logistic regression model [7-10]. However, there was no available information reported on the comparison of different statistical modelling used for the determining of prognostic factors of mortality in first-ever stroke patients.

Even though a few studies have been reported comparing multivariable statistical models towards an outcome, there are still very few similar studies reported in clinical settings. In addition, there is a major concern about collecting and analysing quality data on serious clinical conditions such as stroke to determine clinically important and plausible prognostic factors. There are always major questions whether which kind of outcome data to be collected and what multivariable analysis to be applied. There is a serious need to

prove that if reliable data with meaningful variables could be collected, then there should not be a serious thought on which analysis would be applied since it is hoped that results should be very similar towards common outcome such as mortality.

Modelling of data using different statistical analyses is compared in terms of five important major parameters, namely direction, estimation, precision, significance, and magnitude of the parameter estimates. Similar findings of these five points are hypothetically postulated with good data quality and data management. Data quality starts with the well-defined variables containing categorical and numerical variables to be included in the study. Data quality also involves the determination of the minimum required sample size and appropriate sampling methods. The validity and reliability of measurement tools also influence the quality of data. Data management involves data entry, data coding, selection of appropriate univariable and multivariable statistical tests and clinically, biologically, and statistically plausible data analyses.

This study was an eye opener to the researchers that they should not over-worry about the analysis related to outcome and to prove that even by using different statistical methods, the direction, estimation, precision, significance, and magnitude of parameters are similar, provided data is of good and reliable quality. Researchers can choose the statistical test based on the available data to answer the research questions. Another justification for this study was to highlight which analysis provides more reliable, informative, and best-estimated results. Our study was conducted to compare the parameter estimates of prognostic factors of mortality in first-ever stroke patients using three different multivariable regression methods, including Cox proportional hazards regression, multinomial logistic regression, and multiple logistic regression.

METHODS

A retrospective study among 432 first-ever stroke patients receiving care at Hospital Universiti Sains Malaysia, a 700-bed hospital servicing a predominantly rural area in Northeast Malaysia was conducted. Data from medical records were reviewed, and related information was extracted using a standardized data collection sheet. The inclusion criteria of the participants were the individuals who were clinically diagnosed as first-ever stroke aged more than 18 years old as confirmed by computed tomography scan or magnetic resonance imaging and neurological examination during admission. Participants with recurrent stroke or having any neurological deficits secondary to an infection, epilepsy, tumor, or traumatic causes were excluded from the study.

Power and sample size calculation software version 3.1 (11-12) was used to determine the minimum required sample size. The determination of sample size by Cox regression was calculated based on variable types of strokes. The parameters required for calculation of sample size were level of significance (α) of 0.05 and with pre-determined power ($1-\beta$) of 0.80, the detectable hazard ratio of those with subarachnoid haemorrhage to those with cerebral infarct was decided by the researcher based on clinical expert opinion (hazards ratio [HR]=2.2), the median survival time of stroke patients with cerebral infarct were obtained from the literature [13]. $m_1=84$, the ratio of stroke patients with cerebral infarct to those with subarachnoid haemorrhage was obtained from the literature

[13] ($m=1,318/238=5.54$), the accrual time (A) during which the patients were follow-up was 84 months and with an additional follow-up time (F) of 12 months. The predetermined sample size was 453 patients after adding 10% to the final figure in anticipation of non-readable records and missing data. Systematic random sampling was applied from the list of stroke patients within the study period.

A standardized data collection proforma was designed and verified by another researcher to record all the related information from patient's medical records, namely demographic characteristics, past medical history, clinical characteristics, medications prior to stroke and symptoms and signs of first-ever stroke patients. The dependent variables in this study were determined according to the chosen statistical analysis. For Cox regression, the dependent variable was time to time event, which was the survival time of first-ever stroke patients, measured in days. The survival time was defined as the time interval between the time of diagnosis of stroke and the time of death due to stroke, whereas death was the event of interest. The censored observations were patients who did not experience the event of the study, who were still alive at the end of the study period and who were lost to follow-up during the study period. For multinomial logistic regression, the dependent variable was the status of the patients, which was divided into three levels: alive without neurological deficit, alive with neurological deficit and dead. On the other hand, for multiple logistic regression, the dependent variable was dichotomous, which was alive and dead.

Statistical Analysis

The data were entered and analysed by using Stata/SE version 11 software and IBM SPSS statistics version 22 software. The statistical tests used in this study were Cox proportional hazards regression, multinomial logistic regression, and multiple logistic regression. The comparison of three statistical models was based on the five major parameters, namely direction, estimation, precision, significance, and magnitude of parameter estimates based on different measures of dependent outcome variables and the type of regression method.

The first point direction of the regression coefficient was related to the risk estimates. A positive regression coefficient indicated positively related to the mortality, and a negative regression coefficient indicated protective towards mortality. The second point was estimation with a 95% confidence interval (CI) from sample statistics to population parameters. The third point was the width of CI, known as precision. The next point was the significance of the statistical analysis based on hypothesis formulation and testing, looking at p-value. Lastly, the magnitude of the risk, namely HR, relative risk ratio (RRR) and odds ratio (OR), if these risk estimates were away from the null hypothesis with higher magnitude or away from the null hypothesis with lesser magnitude. Some variables had the same findings of all parameters, which indicated not being able to reject the null hypothesis.

Cox proportional hazards regression was a regression model, which involved the time to an event as the outcome of interest or known as survival time. After univariable analysis, the variables were chosen for the multivariable analysis based on pre-determined criteria. The preliminary main effect model was obtained upon completion of the variable's selection. This was followed by checking linearity of continuous variables, then checking two-way interaction and multicollinearity to

obtain the preliminary final model. The preliminary final model was then checked for its specification error. The next step was checking the proportional hazards assumptions (by using a hazard function plot, a log-minus-log plot, a scatterplot of scaled Schoenfeld residuals, scaled and unscaled Schoenfeld residuals test ($r^*k[\beta]=V^{-1}[\beta, tk]rk[\beta]$) and C-statistics. Then regression diagnostic was performed including Cox-Snell residual, martingale residual, deviance residual and influential analysis [14]. From influential analysis, any potential influential outlier was detected. The final model was achieved after remedial measures were performed by calculating the per cent changes in the regression coefficient. If the percentage were equal and more than 20%, the outlier was considered an influential outlier. The results were expressed based on determined variables, adjusted regression coefficient (b), adjusted HR with 95% CI, Wald statistic and its corresponding p-values.

Multinomial logistic regression was the estimation of the relationship between a polytomous dependent variable and more than one independent variable or covariates. Univariable multinomial logistic regression was performed for each logit function to screen for the important prognostic factors. From the univariable analysis, the variables with p-values less than 0.25 and clinically important were selected for multivariable analysis, which adjusted for confounders. Here the preliminary main effect model was achieved. Then, the linearity of the continuous variables for separate binary models or for each logit function was fitted. The next step was checking interaction and multicollinearity between independent variables to obtain the preliminary final model. Then, the overall fit of the model for separate logit functions was assessed by using four methods, namely the Hosmer-Lemeshow test, Pearson chi-square test, classification table and area under receiver operating characteristic (ROC) curve. Several plots have been suggested in the regression diagnostic to identify the influential covariate patterns for each logit function. The plots consisted of leverage (h), delta Chi-square (ΔX^2), delta deviance (ΔD), and Pregibon delta beta ($\Delta \beta$) versus the estimated logistic probability [15]. The covariate patterns identified as influential factors, which were then further assessed for their changes in the regression coefficient. The changes that are equal, and more than 20% indicated that the covariate pattern was influential to the model. The final model was expressed based on determined variables, adjusted regression coefficient (b), adjusted RRR with 95% CI, Wald statistics and its corresponding p-values.

Multiple logistic regression is the estimation of the relationship between a dichotomous dependent variable and more than one independent variable or covariates, either numerical or categorical variables. Simple logistic regression was performed to screen for the important prognostic factors. Any variable with a p-value less than 0.25 and biologically plausible was recommended to be selected for the multivariable model. The preliminary main effect model was achieved after performing the variables selection. Then, checking linearity of continuous variables followed by checking interaction and multicollinearity were performed. The model was then termed as the preliminary final model. The preliminary final model was then checked for its specification error. Then, the overall fit of the model was assessed by using four methods, namely the Hosmer-Lemeshow test, Pearson chi-square test, classification table and area under ROC curve. Several plots have been suggested in the regression diagnostic

to identify the influential covariate patterns. The plots consisted of leverage (h), delta Chi-square (ΔX^2), delta deviance (ΔD), and Pregibon delta beta ($\Delta \beta$) versus the estimated logistic probability. The covariate patterns identified as influential factors were then further assessed for its changes in the regression coefficient. The changes that are equal, and more than 20% indicated that the covariate pattern was influential to the model. The final model was expressed based on determined variables, adjusted regression coefficient (b), adjusted OR with 95% CI, Wald statistic and its corresponding p-values.

RESULTS

The comparison of statistical modelling using three different statistical analyses is shown in **Table 1**. There were 12 significant prognostic factors detected when using Cox regression, 11 factors when using multinomial logistic regression and nine factors when using multiple logistic regression. The male was identified as a significant variable for all three regression models, where it was identified as a protective factor towards mortality.

The second variable identified as a significant prognostic factor towards mortality was fasting blood sugar. The third significant prognostic factor towards mortality was marital status with four levels; never married (as a reference), married, widowed, and divorced. Being married was a protective factor towards mortality in all three models. On the other hand, being widowed was a protective factor in Cox regression and multiple logistic regression. Being divorced was not observed as significant determinants in all three models.

Other significant prognostic factors towards mortality were diastolic blood pressure and systolic blood pressure; however, it was identified only in Cox regression. The next prognostic factor towards mortality was urea. Another factor was rheumatic heart disease, which was identified in two models, Cox regression and multiple logistic regression. Another significant factor for all models was smoking status with three levels; never smoke (as a reference), ever-smoker and current smoker. Having seizure/fit was protective determinants towards mortality and was observed in Cox regression and multiple logistic regression. Glasgow coma scale was also reported to be a significant prognostic factor towards mortality for all three models.

Other prognostic factors were the usage of aspirin and age at the time of diagnosis; however, these variables were only identified in Cox regression model. Final diagnosis with three levels (ischaemic stroke [as a reference], intracerebral haemorrhage and subarachnoid haemorrhage) was identified as a prognostic factor in multinomial logistic regression and multiple logistic regression models but not in Cox regression model. However, only subarachnoid haemorrhage was observed as a significant prognostic factor towards mortality. Atrial fibrillation and paresis at any site were identified as other determinants towards mortality but were only identified in multinomial logistic regression but not in other two models.

DISCUSSION

Worldwide, stroke comes second after ischaemic heart disease as the leading cause of death over the past decade [16].

Table 1. Comparison of findings of prognostic factors of first-ever stroke patients using three different advanced statistical modelling approaches

Variable	Cox regression			Multinomial logistic regression ^a			Multiple logistic regression		
	Adj. b	Adj. HR (95% CI)	p	Adj. b	Adj. RRR (95% CI)	p	Adj. b	Adj. OR (95% CI)	p
Gender									
Female	0	1		0	1		0	1	
Male	-1.245	0.288 (0.140, 0.595)	0.001	-1.516	0.220 (0.075, 0.642)	0.006	-1.020	0.361 (0.141, 0.923)	0.033
Fasting blood sugar	0.089	1.093 (1.045, 1.143)	<0.001	0.145	1.156 (1.063, 1.257)	0.001	0.118	1.125 (1.051, 1.205)	0.001
Marital status									
Never married	0	1		0	1		0	1	
Married	-2.205	0.110 (0.045, 0.270)	<0.001	-1.456	0.233 (0.060, 0.899)	0.035	-1.636	0.195 (0.065, 0.588)	0.004
Widowed	-2.152	0.116 (0.036, 0.374)	<0.001	-1.467	0.231 (0.040, 1.319)	0.099	-1.895	0.150 (0.036, 0.630)	0.010
Divorced	-1.763	0.171 (0.021, 1.432)	0.103	-0.762	0.467 (0.026, 8.271)	0.603	0.613	1.846 (0.109, 31.307)	0.613
Diastolic blood pressure	0.024	1.024 (1.010, 1.038)	0.001	-	-	-	-	-	-
Urea	0.030	1.030 (1.001, 1.060)	0.043	0.063	1.065 (1.006, 1.128)	0.030	0.091	1.095 (1.043, 1.150)	<0.001
Systolic blood pressure	-0.016	0.984 (0.975, 0.994)	0.001	-	-	-	-	-	-
Rheumatic heart disease	1.766	5.85 (1.982, 17.256)	0.001	-	-	-	2.592	13.362 (2.291, 77.921)	0.004
Smoking status									
Never smoke	0	1		0	1		0	1	
Ever-smoker	1.482	4.40 (1.921, 10.089)	<0.001	1.817	6.155 (1.823, 20.785)	0.003	1.390	4.015 (1.397, 11.541)	0.010
Current smoker	1.442	4.230 (2.051, 8.726)	<0.001	1.687	5.405 (1.817, 16.080)	0.002	1.559	4.753 (1.844, 12.247)	0.001
Seizure/ fit	-0.983	0.374 (0.189, 0.740)	0.005	-	-	-	-1.197	0.302 (0.119, 0.766)	0.012
Glasgow coma scale	-0.311	0.733 (0.687, 0.782)	<0.001	-0.495	0.609 (0.540, 0.688)	<0.001	-0.337	0.714 (0.654, 0.780)	<0.001
Usage of aspirin	-0.572	0.564 (0.342, 0.932)	0.025	-	-	-	-	-	-
Age at time of diagnosis	0.041	1.042 (1.023, 1.062)	<0.001	-	-	-	-	-	-
Hemoglobin	-	-	-	0.069	1.072 (0.931, 1.234)	0.334	-	-	-
Types of strokes									
Ischemic stroke	-	-	-	0	1		0	1	
Intracerebral hemorrhage	-	-	-	0.591	1.805 (0.839, 3.884)	0.131	0.344	1.411 (0.757, 2.628)	0.278
Subarachnoid hemorrhage	-	-	-	1.656	5.237 (1.471, 18.641)	0.011	1.328	3.773 (1.469, 9.690)	0.006
Atrial fibrillation	-	-	-	1.464	4.322 (1.027, 18.193)	0.046	-	-	-
Speech deficit	-	-	-	0.544	1.723 (0.860, 3.445)	0.125	-	-	-
Paresis at any site	-	-	-	1.553	4.725 (2.004, 11.139)	<0.001	-	-	-

Note. ^aMultinomial logistic regression for second logit function (death vs. alive without neurological deficit); b: Regression coefficient; CI: Confidence interval; HR: Hazard ratio; RRR: Relative risk ratio; & OR: Odds ratio

In Malaysia, stroke represents one of the ten principal causes of hospitalization and death in Malaysian hospitals [17]. This current study was a retrospective study, which aimed to determine prognostic factors of mortality among first-ever stroke patients receiving care at a 700-bed hospital servicing a predominantly rural area in Northeast Malaysia by using three different regression models. These three models were then compared based on the direction, estimation, precision, significance, and magnitude of determinants of parameters.

Despite using three different statistical analysis approaches, namely Cox regression, multinomial logistic regression and multiple logistic regression, the results were the

same in terms of five major parameters based on inferential statistics. First was direction of risk. The second was based on the direction estimation with a 95% CI from sample statistics to the population parameter. Third was the width of CI, known as precision. Fourth was the significance of the statistical analysis based on hypothesis formulation and testing, looking at p-value and the level of significance. Lastly was the magnitude of the risk, whether away or not from the null hypothesis.

The summary of the comparison of three different statistical modelling is shown in **Table 2**.

Table 2. Summary of comparison of three different statistical modelling

Variable	Aspects	Cox regression	Multinomial logistic regression	Multiple logistic regression
Gender: Male	Direction	Negative	Negative	Negative
	Estimation	Less than one	Less than one	Less than one
	Precision	Narrowest	Narrow	Narrow
	Significance	More significant	Significant	Significant
	Magnitude	Smaller	Smallest	Small
Fasting blood sugar	Direction	Positive	Positive	Positive
	Estimation	More than one	More than one	More than one
	Precision	Narrowest	Narrow	Narrow
	Significance	More significant	Significant	Significant
	Magnitude	Smallest	Small	Smaller
Marital status: Married	Direction	Negative	Negative	Negative
	Estimation	Less than one	Less than one	Less than one
	Precision	Narrowest	Narrow	Narrow
	Significance	More significant	Significant	Significant
	Magnitude	Smallest	Small	Smaller

Table 2 (Continued). Summary of comparison of three different statistical modelling

Variable	Aspects	Cox regression	Multinomial logistic regression	Multiple logistic regression
Marital status: Widowed	Direction	Negative	Negative	Negative
	Estimation	Less than one	Less than one	Less than one
	Precision	Narrowest	Narrow	Narrow
	Significance	More significant	Not significant	Significant
	Magnitude	Smaller	Not interpretable	Small
Marital status: Divorced	Direction	Negative	Negative	Positive
	Estimation	Less than one	Less than one	More than one
	Precision	Narrowest	Narrow	Narrow
	Significance	Not significant	Not significant	Not significant
	Magnitude	Not interpretable	Not interpretable	Not interpretable
Diastolic blood pressure	Direction	Positive	-	-
	Estimation	More than one	-	-
	Precision	Narrow	-	-
	Significance	Significant	-	-
	Magnitude	Interpretable	-	-
Level of urea	Direction	Positive	Positive	Positive
	Estimation	More than one	More than one	More than one
	Precision	Narrowest	Narrow	Narrow
	Significance	Significant	More significant	Significant
	Magnitude	Smallest	Smaller	Small
Systolic blood pressure	Direction	Negative	-	-
	Estimation	Less than one	-	-
	Precision	Narrow	-	-
	Significance	Significant	-	-
	Magnitude	Interpretable	-	-
Rheumatic heart disease	Direction	Positive	-	Positive
	Estimation	More than one	-	More than one
	Precision	Narrower	-	Narrow
	Significance	More significant	-	Significant
	Magnitude	Smaller	-	Small
Smoking status: Ever-smoker	Direction	Positive	Positive	Positive
	Estimation	More than one	More than one	More than one
	Precision	Narrowest	Narrow	Narrow
	Significance	More significant	Significant	Significant
	Magnitude	Smaller	Small	Smallest
Smoking status: Current smoker	Direction	Positive	Positive	Positive
	Estimation	More than one	More than one	More than one
	Precision	Narrowest	Narrow	Narrow
	Significance	More significant	Significant	Significant
	Magnitude	Smallest	Small	Smaller
Seizure/fit	Direction	Negative	-	Negative
	Estimation	Less than one	-	Less than one
	Precision	Narrower	-	Narrow
	Significance	More significant	-	Significant
	Magnitude	Small	-	Smaller
Glasgow coma scale	Direction	Negative	Negative	Negative
	Estimation	Less than one	Less than one	Less than one
	Precision	Narrow	Narrow	Narrow
	Significance	Significant	Significant	Significant
	Magnitude	Small	Smallest	Smaller
Usage of aspirin	Direction	Negative	-	-
	Estimation	Less than one	-	-
	Precision	Narrow	-	-
	Significance	Significant	-	-
	Magnitude	Interpretable	-	-
Age at time of diagnosis	Direction	Positive	-	-
	Estimation	More than one	-	-
	Precision	Narrow	-	-
	Significance	Significant	-	-
	Magnitude	Interpretable	-	-
Hemoglobin	Direction	-	Positive	-
	Estimation	-	More than one	-
	Precision	-	Narrow	-
	Significance	-	Significant	-
	Magnitude	-	Interpretable	-

Table 2 (Continued). Summary of comparison of three different statistical modelling

Variable	Aspects	Cox regression	Multinomial logistic regression	Multiple logistic regression
Intracerebral hemorrhage	Direction	-	Positive	Positive
	Estimation	-	More than one	More than one
	Precision	-	Narrow	Narrower
	Significance	-	Not significant	Not significant
	Magnitude	-	Not interpretable	Not interpretable
Subarachnoid hemorrhage	Direction	-	Positive	Positive
	Estimation	-	More than one	More than one
	Precision	-	Narrow	Narrower
	Significance	-	Significant	Significant
	Magnitude	-	Smaller	Smallest
Atrial fibrillation	Direction	-	Positive	-
	Estimation	-	More than one	-
	Precision	-	Narrow	-
	Significance	-	Significant	-
	Magnitude	-	Interpretable	-
Speech deficit	Direction	-	Positive	-
	Estimation	-	More than one	-
	Precision	-	Narrow	-
	Significance	-	Not significant	-
	Magnitude	-	Not interpretable	-
Paresis at any site	Direction	-	Positive	-
	Estimation	-	More than one	-
	Precision	-	Narrow	-
	Significance	-	Significant	-
	Magnitude	-	Interpretable	-

In the current study, the male was a protective factor towards mortality for all three regression models. The direction of the regression coefficient for the three models was negatively related to the mortality, given the estimation of the hazard's ratio, RRR, and OR, all less than one. The width of the 95% CI was narrower in Cox regression compared to the other two regression models, indicating it was more precised in Cox regression. p-value was smaller, showing highly significant in Cox regression. All three measurements of risk were away from the null hypothesis with lesser magnitude.

Since OR was the estimation of true risk, the value of OR was a bit larger compared to the other two measurements of risk. Overall, for variable gender, the results were not much different among the three statistical analyses, even though it was slightly better in Cox regression.

On the other hand, for fasting blood sugar, its regression coefficient was positively related to the mortality; with the values of estimation of HR, RRR, and OR were more than one. The width of the 95% CI was not much different between the three statistical analyses, even though it was slightly précised in Cox regression. The significance of p-value also was highly significant in the model using Cox regression. The risk assessments were away from the null hypothesis with a higher magnitude, and the magnitude of the hazard ratio was a bit lower in Cox regression

Being married was a protective factor towards mortality. The regression coefficient of married was negatively related to the mortality, with the values of risk assessment less than one. The width of the 95% CI was narrower in the model using Cox regression, showing precision was better in this model. Even though p-value was significant for all three analyses, it was highly significant in Cox regression. The risk estimates were away from the null hypothesis with a lesser magnitude. The magnitude of the hazard ratio was lower compared to RRR and OR for the married variable.

The regression coefficient of the widowed variable was negatively related to the mortality with the values of risk assessment less than one. The width of the 95% CI was narrower in the model using Cox regression, showing precision was better in this model. The significance of p-value was observed only in models using Cox regression and multiple logistic regression but not in multinomial logistic regression. The risk estimates were away from the null hypothesis with lesser magnitude in the two models. The magnitude of the hazard ratio was a bit smaller compared to OR. However, the risk estimate in multinomial logistic regression not being able to reject the null hypothesis. Thus, the magnitude of RRR was not interpretable.

For divorced variables, the regression coefficient was negatively related to the mortality, with the values of risk assessment less than one for the model using Cox regression and multinomial logistic regression. However, it was a contrast to the model using multiple logistic regression, where the regression coefficient was positively related to the mortality with the value of OR of more than one. The width of the 95% CI was narrower in Cox regression. In terms of the significance of p-value, all models gave insignificant results. The risk estimates were not able to reject the null hypothesis. Therefore, the magnitude of the risk assessment could not be interpretable.

Diastolic blood pressure, which was identified only in Cox regression, the regression coefficient was positively related to the mortality with a hazard ratio of more than one. The width of the 95% CI was relatively narrow. p-value showed the variable was highly significant. The hazard ratio was away from the null hypothesis with a higher magnitude; thus, the magnitude of the hazard ratio could be interpretable.

Next was the level of urea; its regression coefficient was positively related to mortality with the risk assessment of more than one. The width of the 95% CI was relatively narrow in Cox regression compared to the other two analyses. The

significance of p-value was observed in all three models. However, it was highly significant in the multiple logistic regression model. The three risk estimates were away from the null hypothesis with a higher magnitude. The magnitude of the hazard ratio was a bit smaller compared to the other two risk assessments.

Another prognostic factor was systolic blood pressure; however, it was identified only in Cox regression. The regression coefficient was negatively related to the mortality with a hazard ratio of less than one. The width of the 95% CI was relatively narrow. p-value shows the variable was highly significant. The hazard ratio was away from the null hypothesis with a lesser magnitude; thus, the magnitude of the hazard ratio could be interpretable.

The next factor was rheumatic heart disease, which was identified in two models, Cox regression and multiple logistic regression. The regression coefficient was positively related to the mortality with the value of risk assessment more than one. The width of the 95% CI was relatively wide for both models, especially in the multiple logistic regression model. The significance of p-value was highly significant in survival analysis compared to multiple logistic regression. Both risk estimates were away from the null hypothesis with a higher magnitude, and the magnitude of the hazard ratio was much lower compared to OR.

Another significant factor was smoking status. The regression coefficient of ever-smoker was positively related towards mortality with the estimation of risk assessment of more than one. The width of the 95% CI was relatively wide for the multinomial logistic regression model compared to the other two models. In terms of the significance of p-value, Cox regression gave the smallest p-value. The risk estimates of the three models were away from the null hypothesis with a higher magnitude. The magnitude of risk was not much different in the model using Cox regression and multiple logistic regression.

On the other hand, for current smokers, the regression coefficient was also positively related to the mortality with the estimation of risk assessment more than one. The width of the 95% CI was observed to be wider for the multinomial logistic regression model and narrower for Cox regression model, showing precision was better in this model. The significance of p-value was highly significant for Cox regression. The risk estimates were away from the null hypothesis with a higher magnitude for all three models. The magnitude of the hazard ratio was a bit lower, followed by OR and RRR.

Another prognostic factor towards mortality was seizure/fit, which was observed in Cox regression and multiple logistic regression. The direction of the regression coefficient was negatively related to the mortality with the estimation of risk assessment less than one. The width of the 95% CI was the same for these two models. Cox regression reported a smaller p-value compared to multiple logistic regression. The risk estimates for both models were away from the null hypothesis with lesser magnitude. The magnitude of risk assessment for both models was not much different.

Glasgow coma scale was also reported to be a significant prognostic factor towards mortality for all three models. The direction of the regression coefficient was negatively related towards mortality, with the risk assessment value less than one. The width of the 95% CI for all the models was not much different and relatively narrow. p-value was reported to be

highly significant for all models. The risk estimates were away from the null hypothesis for all models with lesser magnitude. The results for the magnitude of risk assessment were similar in all models.

Another prognostic factor was the usage of aspirin; however, it was only identified in Cox regression model. The regression coefficient was negatively related towards the mortality with a hazard ratio of less than one. The width of the 95% CI was relatively narrow. p-value shows the variable was statistically significant. The hazard ratio was away from the null hypothesis with a lesser magnitude; thus, the magnitude of the hazard ratio could be interpretable.

Another factor in Cox regression was age at the time of diagnosis. The direction of the regression coefficient was positively related to the mortality with a hazard ratio of more than one. The width of the 95% CI was relatively narrow. p-value was highly significant, and the risk estimate was away from the null hypothesis with a higher magnitude. Thus, the magnitude of the hazard ratio could be interpretable.

Types of diagnosis were identified as a prognostic factor in multinomial logistic regression and multiple logistic regression models. For intracerebral haemorrhage, the direction of the regression coefficient was positively related towards mortality with the estimation of risk assessment more than once. The width of the 95% CI was relatively narrow for both models. In terms of the significance of p-value, both models gave insignificant results. The risk estimates for both models were not able to reject the null hypothesis. Therefore, the magnitude of risk assessment could not be interpretable.

The direction of the regression coefficient for subarachnoid haemorrhage was positively related towards mortality with the estimate of risk assessment of more than one. The 95% CI was more precise in multiple logistic regression as the width was narrower compared to the multinomial logistic regression. Multiple logistic regression gave a smaller p-value. Both risk estimates were away from the null hypothesis with a higher magnitude. OR gave a smaller magnitude compared to RRR.

Another prognostic factor towards mortality was atrial fibrillation but was only identified in multinomial logistic regression. The direction of the regression coefficient was positively related to the mortality with an estimate of an RRR of more than one. The width of the 95% CI was relatively wide. The variable gave the significant result of p-value; the risk estimate was away from the null hypothesis with a higher magnitude. Therefore, the magnitude of risk assessment was interpretable.

Paresis at any site was identified as another prognostic factor in multinomial logistic regression but not in the other two models. The direction of the regression coefficient was positively related towards mortality with risk assessment more than once. The width of the 95% CI was relatively acceptable. p-value was highly significant, and RRR was away from the null hypothesis with a higher magnitude; therefore, it was interpretable.

Comparison of each variable in three different regression models yielded more or less similar results, even though it was slightly better in Cox regression, especially in terms of precision of the results.

Few studies reported that Cox proportional hazards model was better than other comparison model. A retrospective study among the United States elderly aimed to identify factors associated with long-stay nursing homes by comparing

multivariable logistic regression and Cox proportional hazards model [18].

There were nine significant variables in logistic regression, while only eight significant variables in Cox model. The findings reported hazard ratio from Cox model and OR from logistic regression were similar. However, standard error was found to be smaller in Cox model, indicating results from this model were more precise than from the logistic regression model. This study only reported the comparison based on the precision, which was measured using standard error instead of using the width of CI like in the current study.

The width of CI depended partly on a standard error, which included both standard deviation and sample size [19]. Another prospective study among patients admitted to the intensive care unit comparing three logistic regression models reported that models using dichotomous scales were more appropriate when compared to those using ordinal or quantitative as these models failed to fulfil the model assumptions and estimated biased coefficients.

The extended Cox regression model was the most valid model as it fulfilled all the assumptions, estimated unbiased coefficients, and obtained a precise assessment of risk. However, this study did not report any findings on a comparison based on five aspects as in the current study [20].

A prospective study in a university teaching hospital in France among neurosurgical patients using logistic and Cox models to assess the risk factors for surgical site infections revealed only one significant risk factor for surgical site infection. The parameter estimates, including risk estimates (HR for Cox model and OR for the logistic model) and its 95% CI, were of similar findings. The results of this study reported only three aspects of the parameter, namely estimation, precision, and magnitude of the risk [21].

The study from the project to prevent falls in veterans compared the performance of eight regression models for analysing the risk of falling, focusing on the effect of physical inactivity in older veterans. Three of the eight regression models treated falling as a non-recurrent event, and the parameter estimates were OR by logistic regression, risk ratio by modified Poisson regression and HR by Cox proportional hazards regression.

It was reported that the magnitude of the point estimates was different among the three models; the risk ratio was the smallest, followed by the hazard's ratio, and OR gave the largest value. CI also reported similar results; CI for risk ratio gave the most precise results compared to the two-point estimates [22-25]. This study gave similar results as in the current study in terms of Cox proportional hazards regression and logistic regression on the aspects of estimation, precision, and magnitude of risk assessment.

CONCLUSIONS

Given that, this study had compared three types of statistical analyses frequently used in the literature to analyse stroke data. Our study has clearly yielded that with reliable available data towards a common outcome of interest, different multivariable regression methods provide comparable findings with emphasis on five aspects: direction, estimation, precision, significance, and magnitude, which appeared clinically and statistically plausible.

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that ethics approval was obtained from Human Ethics Committee of Universiti Sains Malaysia USM/KK/PPP/JEPeM (256.4[2.9]).

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

1. WHO. World health statistics 2020. World Health Organization; 2020. Available at: <https://www.who.int/data/gho/publications/world-health-statistics> () Accessed: 13 August 2023.
2. Castro HH, Alencar AP, Benseñor IM, Lotufo P A, Goulart AC. Multimorbidities are associated to lower survival in Ischaemic stroke: Results from a Brazilian stroke cohort (EMMA study). *Cerebrovasc Dis.* 2017;44(3-4):232-9. <https://doi.org/10.1159/000479827> PMID:28848194
3. Peng Y, Ngo L, Hay K, Alghamry A, Colebourne K, Ranasinghe I. Long-term survival, stroke recurrence, and life expectancy after an acute stroke in Australia and New Zealand from 2008-2017: A population-wide cohort study. *Stroke.* 2022;53(8):2538-48. <https://doi.org/10.1161/STROKEAHA.121.038155> PMID:35418238
4. Ekker MS, Boot EM, Singhal AB, et al. Epidemiology, aetiology, and management of ischaemic stroke in young adults. *Lancet Neurol.* 2018;17(9):790-801. [https://doi.org/10.1016/S1474-4422\(18\)30233-3](https://doi.org/10.1016/S1474-4422(18)30233-3) PMID:30129475
5. Barlas RS, Honney K, Loke YK, et al. Impact of hemoglobin levels and anemia on mortality in acute stroke: Analysis of UK regional registry data, systematic review, and meta-analysis. *J Am Heart Assoc.* 2016;5(8):e003019. <https://doi.org/10.1161/JAHA.115.003019> PMID:27534421 PMID:PMC5015269
6. Avan A, Digaleh H, Di Napoli M, et al. Socioeconomic status and stroke incidence, prevalence, mortality, and worldwide burden: An ecological analysis from the global burden of disease study 2017. *BMC Med.* 2019;17(1):191. <https://doi.org/10.1186/s12916-019-1397-3> PMID:31647003 PMID:PMC6813111
7. Kooi CW, Peng HC, Aziz ZA, Looi I. A review of stroke research in Malaysia from 2000-2014. *Med J Malaysia.* 2016;71(Supplement 1):58-69.
8. Bolayir A, Gokce SF, Cigdem B, et al. Monocyte/high-density lipoprotein ratio predicts the mortality in ischemic stroke patients. *Neurol Neurochir Pol.* 2018;52(2):150-5. <https://doi.org/10.1016/j.pjnns.2017.08.011> PMID:28864326
9. Aminov A, Rogers JM, Middleton S, Caeyenberghs K, Wilson PH. What do randomized controlled trials say about virtual rehabilitation in stroke? A systematic literature review and meta-analysis of upper-limb and cognitive outcomes. *J Neuroeng Rehabil.* 2018;5(1):29. <https://doi.org/10.1186/s12984-018-0370-2> PMID:29587853 PMID:PMC5870176
10. Russell JB, Charles E, Conteh V, Lisk DR. Risk factors, clinical outcomes and predictors of stroke mortality in Sierra Leoneans: A retrospective hospital cohort study. *Ann Med Surg (Lond).* 2020;60:293-300. <https://doi.org/10.1016/j.amsu.2020.10.060> PMID:33204420 PMID:PMC7649580

11. Ryan TP. Sample size determination and power. Toronto, ON: John Wiley & Sons; 2013. <https://doi.org/10.1002/9781118439241>
12. Naing NN. A practical guide on determination of sample size in medical and health sciences research. Kelantan, Malaysia: Pustaka Aman Press; 2014.
13. Singh RJ, Chen S, Ganesh A, Hill MD. Long-term neurological, vascular, and mortality outcomes after stroke. *Int J Stroke*. 2018;13(8):787-96. <https://doi.org/10.1177/1747493018798526> PMID:30160619
14. Hosmer DW, Lemeshow S, May S. Applied survival analysis: Regression modeling of time-to-event data. Hoboken, NJ: John Wiley & Sons, Inc; 2008. <https://doi.org/10.1002/9780470258019>
15. Hosmer DW, Lemeshow S. Applied logistic regression. New York City, NY: John Wiley & Sons, Inc; 2000. <https://doi.org/10.1002/0471722146> PMID:10886529
16. Sargolzaie N, Kiani M, Salari M, Khosravi M. Top ten causes of death and life expectancy in Zahedan (South-East Iran) in 2014. *Glob J Health Sci*. 2016;9(6):135. <https://doi.org/10.5539/gjhs.v9n6p135>
17. Khoo S, Poh BK, Suhaimi SA, Chong KH, Ramirez Varela A. Physical activity promotion in Malaysia: Challenges and opportunities. *Front Public Health*. 2020;8:536239. <https://doi.org/10.3389/fpubh.2020.536239> PMID:33194945 PMCID:PMC7652762
18. Rolden HJA. Population ageing and health care expenditure [Doctoral dissertation]. Leiden: Leiden University; 2016.
19. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *Eur J Epidemiol*. 2016;31:337-50. <https://doi.org/10.1007/s10654-016-0149-3> PMID:27209009 PMCID:PMC4877414
20. Brown KA, Daneman N, Stevens VW, et al. Integrating time-varying and ecological exposures into multivariate analyses of hospital-acquired infection risk factors: A review and demonstration. *Infect Control Hosp Epidemiol*. 2016;37(4):411-9. <https://doi.org/10.1017/ice.2015.312> PMID:26880280
21. Fang C, Zhu T, Zhang P, Xia L, Sun C. Risk factors of neurosurgical site infection after craniotomy: A systematic review and meta-analysis. *Am J Infect Control*. 2017;45(11):e123-34. <https://doi.org/10.1016/j.ajic.2017.06.009> PMID:28751035
22. Ge ML, Simonsick EM, Dong BR, Kasper JD, Xue QL. Frailty, with or without cognitive impairment, is a strong predictor of recurrent falls in a US population-representative sample of older adults. *J Gerontol A Biol Sci med Sci*. 2021;76(11):e354-60. <https://doi.org/10.1093/gerona/glab083> PMID:33721909 PMCID:PMC8562403
23. Fekete M, Szarvas Z, Fazekas-Pongor V, et al. Nutrition strategies promoting healthy aging: From improvement of cardiovascular and brain health to prevention of age-associated diseases. *Nutrients*. 2022;15(1):47. <https://doi.org/10.3390/nu15010047> PMID:36615705 PMCID:PMC9824801
24. Hartley P, Forsyth F, O'Halloran A, Kenny RA, Romero-Ortuno R. Eight-year longitudinal falls trajectories and associations with modifiable risk factors: Evidence from The Irish longitudinal study on ageing (TILDA). *Age Ageing*. 2023;52(3):afad037. <https://doi.org/10.1093/ageing/afad037> PMID:36995137 PMCID:PMC10061938
25. Guo X, Pei J, Ma Y, et al. Cognitive frailty as a predictor of future falls in older adults: A systematic review and meta-analysis. *J Am Med Dir Assoc*. 2022;24(1):38-47. <https://doi.org/10.1016/j.jamda.2022.10.011> PMID:36423679