

MODELLING OF RISK FACTORS, CASE-FATALITIES, SURVIVAL AND FUNCTIONAL HEALTH STATUS FOR STROKE IN KELANTAN, MALAYSIA

Kamarul Imran Musa

Lancaster Medical School

Lancaster University

The thesis is submitted for the degree of Doctor of Philosophy

March 2017

Abstracts

MODELLING OF RISK FACTORS, CASE-FATALITIES, SURVIVAL AND FUNCTIONAL HEALTH STATUS FOR STROKE IN KELANTAN, MALAYSIA

Kamarul Imran Musa
Lancaster Medical School
Lancaster University

Submitted for the degree of Doctor of Philosophy
March 2017

Stroke is an important public health problem worldwide. It is a non-communicable disease of increasing importance in the ageing population. There are four major types of stroke: a) ischaemic stroke, b) primary intracerebral haemorrhage, c) subarachnoid haemorrhage and d) undetermined stroke (no computed tomography [CT], magnetic resonance imaging [MRI], autopsy or cerebral angiography).

The risk factors and fatalities for stroke vary worldwide and stroke accounts for about 9.7% of all deaths worldwide. Unfortunately, most stroke deaths occur in the less developed countries, where stroke research is scanty. Patients who survive stroke will have a wide range of functional limitations that affect their daily activities.

To date, only few reliable data are available for identifying the risk factors and understanding the stroke fatalities in the low- and middle-income countries. The effect of stroke on daily activities is also understudied. Measuring the effect of stroke on daily activities is important to evaluate the recovery process. Understanding the factors affecting daily activities post-stroke helps identify areas where intervention may benefit stroke survivors the most.

In this thesis, I have worked on the questions that will improve my knowledge and understanding of the natural history of stroke in the state of Kelantan, Malaysia. In this study, we posed questions to subjects from the population of interest to reflect our results across the greater Kelantan population to develop a model of risk factors for stroke, models of case-fatalities and survival to compare the characteristics and outcomes of two main types of stroke, i.e. ischaemic stroke and haemorrhagic stroke, and lastly to assess the longitudinal change in functional health status using the Barthel Index post-stroke.

This thesis includes four draft papers, in which several modelling data collection and data analysis strategies were applied to four datasets: one was provided by the hospitals, two were extracted by us and the final one was based on personal interviews with stroke survivors.

The first draft paper is based on an observational study using data from the records offices of two major hospitals in Kelantan. In this paper, I analysed and modelled the risk factors for stroke using a case-control study design. This dataset contained individual-level variables (patient variables from hospitals) and area-level variables supplied by the Department of Statistics, Malaysia, and we utilised logistic regression to model the risk factors for stroke. In the results, we showed the non-linear relationship between age and odds for stroke and the interaction of age with sex in the model.

In the second draft paper, we explored the important prognostic factors for in-hospital stroke fatalities. Using Cox proportional hazard regression, we found that the only two independent prognostic factors for stroke fatality in the hospitals were: a) stroke subtype and b) age.

To further investigate the different prognostic effect of stroke subtype on admission and on fatality, we recruited consecutive in-hospital stroke patients. In the analysis, we performed Cox proportional hazard regression to quantify the odds of stroke fatality for: a)

ischaemic stroke and b) haemorrhagic stroke. In this third draft paper, we showed the prognostic effect of stroke subtype on stroke fatality.

In the fourth draft paper, we recruited stroke patients and interviewed them on three occasions. In this longitudinal assessment, we assessed the functional health status of stroke patients until 3 months after hospital discharge. I conducted all interviews and assessed the functional outcome using the well-known Barthel Index. Considering the longitudinal format of the data, we used the linear mixed effect model to model the rate of change of the Barthel Index at the three measurement occasions.

We have identified several limitations in this PhD project and have taken several measures to minimize the biases caused by those limitations. The limitations include the need for us to do handsearching for data abstraction, potential informative censoring due to our study design and short follow-up times, limited generalizability of results, small sample sizes, missing observations, missing important variables (to be modelled as covariates), absence of residential coordinates and using data on arrival to Emergency department (no pre-arrival data). If the censoring mechanism provides significant information with time (T), numerical estimates from Kaplan-Meier and Cox proportional hazard regression are biased.

The new knowledge stemming from the stroke modelling and outcome assessment developed and analysed in this thesis could help improve our understanding of stroke in Kelantan. The thesis will also improve our understanding and knowledge of the natural history of the disease, i.e. the progression from risk factors to outcome (fatality or functional residuals) after stroke. In conclusion, our data and the four draft papers written based on this PhD project have added new stroke data and knowledge on the progression of stroke, which is understated in the Malaysian and Asian population in general and in Kelantan specifically.

Acknowledgements

There are many people who I like to say thanks:

To Dr Thomas J Keegan, who persistently and consistently agreed to read all the pages of this thesis and the draft manuscripts. His insight and curiosity helped this thesis be less abstruse and more readable but at the same time kept it well within a scientific scope. Silly mistakes, confusing sentences and ideas that sometimes veered off the main path did not deter him from improving this work. I am grateful and blessed that, throughout the years, our supervisor-supervisee relation has always been strong.

To Distinguished Professor Peter J Diggle, whose critical reviews, questions and suggestions guided me in writing this thesis. He stressed the importance of strong principles in performing research work and statistical analysis. He contributed significantly in linking the research questions, research objectives and the statistical results and, at the end, in knitting them together to make scientific sense. His ability to relate statistical methods and results to life problems and meanings is a trait every scientist should possess.

To my wife, Jue, who still cannot understand why I love to read almost all epidemiological and statistical books but who is always there for me. Her sacrifice and charm and insights have made this journey fun, wonderful and happy.

To my sons, Afif and Iman, who always question the methods of schooling. Both of you are gifts from Allah, and forever I will cherish you. Your thoughts and opinions will guide you in future years, and I know one thing: both of you will be people of significance.

To my late father, Tn Hj Musa Ibrahim and mother, Pn Hjh Napisah Hj Md Nor. They always believed in me and with their prayers made this a reality. To my father in law and my late mother in law, I would like to say 'Thanks,' too, for giving me space and time.

To all the people at CHICAS who have been wonderful and supportive. Many of you are talking notations and formulae, but your willingness to hear applied matters impresses me.

To the Malaysian community in Lancaster who have been very accommodating and wonderful. The time working and being with all of you will become some of the fondest memories in our lives.

Declaration

I declare that this thesis is my own work and has not been submitted in any form for the award of a higher degree elsewhere.

Kamarul Imran Musa
Student ID: 30498505

List of Draft Papers

This thesis includes the following four draft papers:

1. Kamarul Imran Musa, Peter J Diggle, Thomas J Keegan. *Individual and Area-level Risk Factors for Stroke in Kelantan, Malaysia: A Hospital-based Case-control Study.*
2. Kamarul Imran Musa, Peter J Diggle, Thomas J Keegan. *Glasgow Coma Scale and Age as Independent Prognostic Factors for In-hospital Stroke Fatality.*
3. Kamarul Imran Musa, Juhara Haron, Peter J Diggle, Thomas J Keegan. *Comparing Cerebral Infarction and Haemorrhagic Stroke in Asia: Case-fatality and Prognostic Effect for In-hospital Survival.*
4. Kamarul Imran Musa, Peter J Diggle, Thomas J Keegan. *Barthel Index Scores Over Three Months Are Related to Age and Stroke Subtype in Asian Stroke Patients.*

Table of Contents

Abstracts	i
Acknowledgements.....	iv
Declaration.....	vi
List of Draft Papers.....	vii
List of Figures	xv
List of Tables	xviii
List of Abbreviations	xxi
Chapter 1 Introduction.....	1
1.1 Stroke as a public health problem.....	1
1.2 Definition of stroke	3
1.3 Classification of stroke	4
1.3.1 Ischaemic stroke (IS).....	5
1.3.2 Haemorrhagic stroke (HS)	6
1.3.2.1 Primary intracerebral haemorrhage (PICH).....	7
1.3.2.2 Subarachnoid haemorrhage (SAH)	7
1.3.2.3 Other types of haemorrhagic stroke.....	10
1.4 Coding of stroke	11
1.5 Epidemiology of stroke.....	12
1.5.1 Incidence, prevalence and burden for stroke.....	13
1.5.2 Proportional frequency of stroke types	15
1.5.2.1 Proportional frequency for ischaemic stroke (IS)	16
1.5.2.2 Proportional frequency for haemorrhagic stroke (HS).....	17
1.5.3 Stroke trends	18

1.6	Demography of Malaysia and Kelantan and stroke as a public health problem	18
1.6.1	Demography of Malaysia.....	19
1.6.2	Demography of Kelantan, Malaysia	19
1.6.3	Stroke burden in Malaysia.....	20
1.6.4	Health care system in Malaysia	21
1.6.5	Stroke services in Kelantan, Malaysia	23
1.6.5.1	Hospital Universiti Sains Malaysia (HUSM)	24
1.6.5.2	Hospital Raja Perempuan Zainab (HRPZ) II.....	28
1.7	Risk factors for stroke	28
1.7.1	Non-modifiable risk factors	29
1.7.2	Modifiable risk factors.....	30
1.7.3	Risk factors for ischaemic stroke	31
1.7.4	Risk factors for haemorrhagic stroke	32
1.8	Outcomes of stroke.....	33
1.8.1	Stroke mortality and stroke case-fatality	34
1.8.2	Predictors for stroke mortality and stroke case-fatality	37
1.8.3	Predictors of functional health status after stroke	38
1.9	Modelling binary outcome: Logistic regression model	39
1.9.1	Logistic regression model	39
1.9.2	Model fitting for a logistic regression	40
1.9.3	Model checking for a logistic regression model	41
1.10	Modelling time-to-event outcome: Survival analysis and the Cox proportional hazard regression	42
1.10.1	Descriptive method for survival analysis	43
1.10.2	The Cox proportional hazard regression	44

1.10.3	Model fitting for the Cox proportional hazard regression.....	46
1.10.4	Censoring mechanism and informative censoring	47
1.10.5	Implication of informative censoring: An example in stroke setting	48
1.10.6	Methods to handle informative censoring	49
1.10.7	Residuals and proportionality assumption for the Cox proportional hazard regression	51
1.11	Modelling longitudinal data analysis: Linear mixed models	52
1.11.1	Naïve analysis for the correlated data.....	52
1.11.2	Multi-level models for longitudinal data	53
1.11.3	Model checking for the linear mixed models	54
1.12	Issues in model building	55
1.12.1	Effect of the third variable.....	55
1.12.1.1	Confounders.....	55
1.12.1.2	Interaction.....	56
1.12.2	Modelling the functional form of continuous covariates	57
1.12.2.1	The functional forms of continuous covariates	57
1.12.2.2	Fractional polynomials (FP) and other methods.....	58
1.12.2.3	Software to model the functional form of the continuous covariates.....	59
1.12.3	Variable selection	60
1.12.4	Goals of model building.....	61
1.13	Thesis overview	62
1.13.1	Draft Paper 1	63
1.13.2	Draft Paper 2	64
1.13.3	Draft Paper 3	65
1.13.4	Draft Paper 4.....	66

Chapter 2	Individual and Area-level Risk Factors for Stroke in Kelantan, Malaysia: A Hospital-based Case-control Study	68
2.1	Abstract	69
2.2	Introduction	70
2.3	Methods	72
2.3.1	Study design and setting	72
2.3.2	Selection of cases and controls	72
2.3.3	Statistical analysis	74
2.3.4	Ethical approval	77
2.4	Results	77
2.5	Discussion	84
2.6	Conclusion	90
2.7	Acknowledgements	91
2.8	Funding	91
Chapter 3	Glasgow Coma Scale and Age as Independent Prognostic Factors for In-hospital Stroke Fatality	92
3.1	Abstract	93
3.2	Introduction	94
3.3	Methods	96
3.3.1	Study design and data source	96
3.3.2	Case definition	96
3.3.3	Variables	97
3.3.4	Statistical analysis	98
3.3.5	Ethical approval	101
3.4	Results	101

3.5	Discussion	105
3.6	Conclusion.....	110
3.7	Acknowledgements.....	111
Chapter 4 Comparing Cerebral Infarction and Haemorrhagic Stroke in Asia:		
Case–Fatality and Prognostic Effect for In-hospital Survival		
		112
4.1	Abstract.....	113
4.2	Introduction	114
4.3	Methods.....	115
4.3.1	Study venue	115
4.3.2	Case definition	116
4.3.3	Variables.....	117
4.3.4	Statistical analysis	117
4.3.5	Ethical approval.....	119
4.4	Results.....	119
4.4.1	Stroke fatality.....	120
4.4.2	Prognostic factors between ischaemic and haemorrhagic stroke	121
4.5	Discussion	122
4.6	Conclusion.....	132
4.7	Acknowledgements.....	132
Chapter 5 Barthel Index Scores over Three Months are Related to Age and		
Stroke Subtype in Asian Stroke Patients		
		133
5.1	Abstract.....	134
5.2	Introduction	135
5.3	Methods.....	136
5.3.1	Patients.....	136

5.3.2	Barthel Index.....	137
5.3.3	Consent and interview	137
5.3.4	Measurement occasions and covariates	139
5.3.5	Statistical analysis	139
5.3.6	Ethical approval.....	140
5.4	Results.....	141
5.5	Discussion	144
5.6	Conclusion.....	151
5.7	Acknowledgements.....	152
Chapter 6	Conclusion	153
6.1	Limitations and strengths.....	154
6.2	Experience gained from collaboration.....	160
6.3	Further research.....	162
6.4	Closing remarks	163
Chapter 7	Appendices	165
	Appendix A: Additional results for Chapter 2	165
	Appendix B: Additional results for Chapter 3	169
	Appendix C: Additional results for Chapter 4	174
	Appendix D: Additional results for Chapter 5	182
	Appendix E: Data abstraction form	189
	Appendix F: Interview Form (Malay version)	193
	Appendix G: Interview Form (English version)	205
	Appendix H: Methods to deal with informative censoring	216
	Gamma imputation method.....	216
	Risk score multiple imputation method.....	218

Software package for multiple imputation in time-to-event data	219
Assessing non-informative censoring in our data.....	220
Chapter 8 References	225

List of Figures

Figure 1-1 Signs of cerebral ischaemic on CT images. Hypo-attenuation of brain tissue is due to the increased brain water content inside the brain. Swelling of the gyri is due to the brain oedema. Hyperdense middle-cerebral artery (MCA) sign is a result of thrombus or embolus in the MCA. Blurred basal ganglia in MCA infarct and in insular ribbon sign refers to hypodensity and swelling of the insular cortex (Ahmed Abd Rabou and Frank Gallard, 2012). 6

Figure 1-2 (a) Non-contrast CT brain and (b) MRI brain of a patient with a basal ganglia haemorrhage. This location is typical for a hypertensive intracerebral haemorrhage (ICH). (c) Non-contrast CT brain of a different patient demonstrating a large, acute left parietal intraparenchymal ICH. The location and appearance of the lesion is suggestive of amyloid angiopathy. (d) CT brain of another patient, showing the classic appearance of a subarachnoid haemorrhage (Trevor and Fedi, 2013).8

Figure 1-3 Subarachnoid haemorrhage (SAH) on CT images. CT scan is the first line of investigation for haemorrhagic stroke. This image shows positive sign for hyperdense appearance of extravasated blood in the basal cistern. False positive SAH could be due to generalized brain oedema. In this figure, sedimentation in the left occipital horn is the only positive finding (van Gijn and Rinkel, 2001).9

Figure 1-4 Subarachnoid haemorrhage (SAH) on CT images appears as a high-attenuating, amorphous substance that fills the normally dark, CSF-filled subarachnoid spaces around the brain, appear white in acute haemorrhage, most evident in the largest subarachnoid spaces, such as the supra-sellar cistern and Sylvian fissures (van Gijn and Rinkel, 2001).9

Figure 1-5 Extradural haemorrhage on CT scan images (Hacking and Gaillard, 2012b).10

Figure 1-6 Acute subdural haemorrhage on CT scan images (Hacking and Gaillard, 2012a).
.....10

Figure 1-7 Map of Malaysia and the population density by state, Malaysia for year 2010 (Department of Statistics, 2015)	25
Figure 1-8 Age-standardized stroke mortality per 100,000 people in 2010 (Feigin et al., 2014). Malaysia – comprised of Peninsular Malaysia and East Malaysia in the Borneo – is located inside the box.	26
Figure 1-9 Worldwide geographic distribution of relative mortality from stroke (Kim and Johnston, 2011).	36
Figure 2-1 The estimated log-odds from the final model (n=3118) are plotted against the variable age. The estimated age effects for males (circles) and females (crosses) shows that the log odds for stroke in men are lower than women at ages up to 35 years, slightly higher between ages 40 years and 60 years, and increasingly lower starting at the age of 65 years.	82
Figure 2-2 The adjusted odds ratios (OR) from the final model (n=3118) are plotted against the variable age. Though the log odds show different age effect for men and women in throughout the age as shown in Figure 2-1, but based the 95% confidence intervals (CI) for the odds ratio, they only reached statistical significance at the age of 75 years and beyond (as 95% CIs are below 1).	82
Figure 2-3 Sample variogram and envelopes generated by simulating the semivariance against the distance (in metre). The variances come from the final fitted multiple logistic model. The variogram was based on observed residuals generated by the logistic model and the ordinates were taken from the centroid coordinates of the sub-sub districts ('mukim kecil'), n=286.....	84
Figure 4-1 Kaplan-Meier survival estimates and curves of overall survival for the cerebral infarction (CI) group and haemorrhagic stroke (HS) group. The number at risk shows the	

numbers of patients still alive, and therefore at risk for stroke death at different period of time (days). The survival probability indicates the Kaplan-Meier probabilities (conditional probabilities) that a patient survives so many days or longer. The median survival time for CI was 28 days and for HS was 14 days – meaning that a patient with CI has a 50% probability of surviving 28 days or longer and for a patient with HS has a 50% probability of surviving 14 days or longer. The log rank test (results not shown) gave the p-value <0.001.....125

Figure 5-1 The cumulative proportions and the baseline Barthel Index scores for all patients (n=98) and for patients who died after baseline measurement up to 3 months post baseline (n=13). The cumulative proportions for Barthel Index were compared using the two-sample Kolmogorov-Smirnov test (2-sided test, D = 0.4027, p-value =0.0484)144

List of Tables

Table 1-1 The International Classification of Disease 10 th Revision (ICD-10), Chapter IX (Diseases of circulatory system), Blocks I60-I69 (Cerebrovascular diseases).....	13
Table 1-2 Demography of Malaysia for year 2010, 2011, 2012 and 2013.	19
Table 1-3 Demography of Kelantan, Malaysia for year 2010, 2011, 2012 and 2013.....	27
Table 2-1 Diagnostic categories for controls.....	74
Table 2-2 Individual and area-level variables of cases and controls. The outcome was coded as a binary variable either as a case (has stroke) or a control (does not have stroke).....	79
Table 2-3 Simple logistic regression results showing the estimated regression coefficients, respective 95% confidence intervals, p-values and the log-likelihood values for each variable.	81
Table 2-4 The final model developed from multiple logistic regression analysis (n=3118) showing the estimated odds ratios, respective 95% confidence intervals and p-values.....	81
Table 3-1 On-admission numerical variables for all patients (regardless of survival status at discharge) and for patients based on survival status at discharge, n=226	103
Table 3-2 On-admission categorical variables for all patients (regardless of survival status at discharge) and for patients based on survival status at discharge, n=226	106
Table 3-3 Numerical variables, crude hazard ratios (HR) with 95% confidence intervals (CI) and the respective p-values. The outcome variable was a time-to-event variable; where the event was status at discharge (dead or alive) and the time was the numbers of days of hospitalization.	106

Table 3-4 Categorical variables, crude hazard ratios (HR) with 95% confidence intervals (CI) and the respective p-values. The outcome variable was a time-to-event variable; where the event was status at discharge (dead or alive) and the time was the numbers of days of hospitalization107

Table 3-5 Two prognostic models for in-hospital stroke fatality. The best model (Model 1, n=225) contains two independent predictors: age (years) and Glasgow Coma Scale (range of score is from 3 to 15). The alternative model (Model 2, n=226) has age (years) and stroke subtypes (haemorrhagic stroke vs non-haemorrhagic stroke) as the independent predictors. The parameters shown are the adjusted hazard ratios (HR), standard error (SE), lower and upper 95 confidence intervals (CI) for adjusted hazard ratios (HR) and the respective p-values107

Table 4-1 The means and standard deviations (SD) for age, length of stay (LOS), Glasgow Coma Scale (GCS), systolic blood pressure (SBP in mmHg) and diastolic blood pressure (DBP in mmHg) stratified for sex and stroke subtype123

Table 4-2 The frequencies and percentages of the type of referral for stroke patients to the Emergency Department, Hospital Universiti Sains Malaysia (HUSM) for different stroke subtypes124

Table 4-3 The overall stroke case fatality and the 3 days, 7 days and 14 days stroke fatalities for patients with cerebral infarction (CI) and haemorrhagic stroke (HS).....124

Table 4-4 The crude hazard ratios (HR) and adjusted HR (adjusted for age then sex), standard errors and 95% confidence intervals for HR estimated using the Cox proportional hazard regression model. The haemorrhagic stroke (HS) was compared against the cerebral infarction (CI as the baseline group). In all comparisons, p-values yielded were less than 0.001.125

Table 4-5 The crude hazard ratios (HR) and adjusted HR (adjusted for age then sex), standard errors and 95% confidence intervals for HR estimated using the Cox proportional hazard regression model. The haemorrhagic stroke (HS) was compared against the cerebral infarction (CI as the baseline group). All p-values were less than 0.050.126

Table 5-1 Patient characteristics and distribution of patients at baseline, one month and three months' post baseline.....143

Table 5-2 The Barthel Index score at baseline, 1-month post-baseline and 3 months' post-baseline by sex, major stroke subtype, and vital status.....143

Table 5-3 The estimated regression parameters from univariate linear regression. The outcomes are Barthel Index scores at baseline, 1-month post-baseline and 3 months' post-baseline.148

Table 5-4 Results from linear mixed models performed using maximum likelihood estimation. The covariates and the assigned model are shown in the first column. The regression parameters estimated from these main effect models are shown with the p-values and log-likelihood.149

Table 5-5 Results from linear mixed models performed using maximum likelihood estimation. The covariates contained the 2-way interaction term from the best model in table 4 (model 4). The regression parameters estimated from these main effect models are shown with the p-values and log-likelihood.....150

List of Abbreviations

AASP	-	Asian Acute Stroke Advisory Panel
AIC	-	Akaike Information Criterion
BI	-	Barthel Index
BIC	-	Bayesian Information Criterion
CI	-	Confidence Intervals
CT	-	Computed-Tomography
DALY	-	Daily-Adjusted-Life-Years
DBP	-	Diastolic Blood Pressure
dof	-	degree of freedom
EMR	-	Emergency Medical Records
FP	-	Fractional Polynomials
FP2	-	Fractional Polynomials (2 degrees)
FP3	-	Fractional Polynomials (3 degrees)
GCS	-	Glasgow Coma Scale
HDL	-	High-density Lipoprotein
HR	-	Hazard Ratio
HREC	-	Human Research Ethics Committee
HRPZ	-	Hospital Raja Perempuan Zainab
HS	-	Haemorrhagic stroke
HUSM	-	Hospital Universiti Sains Malaysia
ICD-10	-	International Classification of Diseases 10 th Revision
ICH	-	Intra-Cerebral haemorrhage
IS	-	Ischaemic stroke
LDL	-	Low-density Lipoprotein
LL	-	Likelihood ratio
LR	-	Likelihood Ratio
MCA	-	Middle Cerebral Artery
MFP	-	Multivariable Fractional Polynomials
MREC	-	Medical and Research Ethics Committee
MRI	-	Magnetic Resonance Imaging
NHANES	-	National Health And Nutritional Examination Survey
NIHSS	-	National Institute of Health Stroke Scale
OR	-	Odds Ratio
PACS	-	Picture Archiving and Information System

PICH	-	Primary Intra-Cerebral Haemorrhage
REGARDS	-	The REasons for Geographic And Racial Differences Study
ROC	-	Receiver-Operating-Characteristics
SAH	-	Subarachnoid haemorrhage
SBP	-	Systolic Blood Pressure
SD	-	Standard Deviation
SE	-	Standard Error
SEA	-	South-East Asian
SSS	-	Scandinavian Stroke Scale
TIA	-	Transient Ischaemic Attack
TWC	-	Total White Count
WHO	-	World Health Organization

Chapter 1 Introduction

In this chapter, I begin by introducing stroke as an important public health problem and give the definition of stroke. Next, I classify stroke and describe stroke according to the World Health Organisation (WHO) International Classification of Diseases 10th Revision (ICD-10). Following that, I describe the epidemiology of stroke, globally and regionally, including in Malaysia. Next, I discuss the common risk factors and the outcomes after stroke, which all motivated this work. In the following four sections, I outline the foundations of the epidemiological and statistical methods I have applied in my study. These foundations include the modelling of binary outcome, time-to-event outcome and longitudinal data. I also briefly discuss important issues in model building in epidemiology such as the presence of confounders, interaction and the choice of functional form for numerical covariates. Finally, I present a summary of the thesis, including 4 draft papers planned for submission stemming from this work.

1.1 Stroke as a public health problem

Cardiovascular diseases are important public health problems in the 21st century. This is largely due to an epidemiological phenomenon known as epidemiologic transition. In epidemiologic transition, the process of disease transition in the population from a population infected with largely infectious diseases (communicable) to a population burdened with non-infectious (non-communicable) diseases takes place (Omran, 1971, Omran, 1983, Omran, 2005, Gaziano, 2010). Due to epidemiologic transition, countries in Southeast Asia (SEA), including Malaysia, previously known for their communicable diseases, are now facing an epidemic of non-communicable diseases (Dans et al., 2011).

Due to falling fertility rates and prolonged life expectancy, the number of people aged 65 or older is projected to grow to 1.5 billion in 2050, with most of the increase in developing countries. This aging, as well as diet and changes in lifestyle, drive the rise in chronic non-communicable diseases in today's developing countries (World Health Organization, 2011a).

Non-communicable diseases, also known as chronic diseases, affect the low- and middle-income countries disproportionately. Of all non-communicable diseases, cardiovascular diseases account for most non-communicable deaths. With more than 150 known causes, stroke is an example of a heterogeneous cardiovascular disease (Amarenco et al., 2009){Amarenco, 2009 #554} which causes major morbidity and mortality in both the developed and developing countries (Lindley, 2008, Venketasubramanian and Chen, 2008, Venketasubramanian et al., 2015, Liu et al., 2001, Thorvaldsen et al., 1997, Feigin et al., 2003, Feigin, 2007, Feigin et al., 2009, Feigin et al., 2014, Krishnamurthi et al., 2014) .

Based on the WHO data, it was estimated that there were 15.3 million strokes worldwide in 2002, causing 10% (5.5 million) of deaths in the same year (Johnston et al., 2009). In the developed countries, stroke is the third most common cause of death after heart disease and cancer (Lindley, 2008). However, more stroke deaths (85%) occur in the low- and middle-income countries (Lopez et al., 2006).

The Global Burden of Diseases, Injuries, and Risk Factors Study 2010 (GBD 2010) also provides estimates for difference disease burden including stroke. GBD 2010 searched databases including Medline, Embase, Scopus, Pubmed and many others to estimate the global and regional burden of stroke during 1990 and 2010. GBD 2010 used an analytical technique (DisMod-MR) to calculate these estimates: stroke incidence, prevalence, mortality and disability-adjusted life-years (DALYs) lost by age group and country income level (high-income and low-income and middle-income) for 1990, 2005, and 2010 (Krishnamurthi et al., 2014, Feigin et al., 2014, Krishnamurthi et al., 2013). Unfortunately, there were limited data

sources with questionable quality from low-income and middle-income countries (Fuentes and Tejedor, 2014, Feigin et al., 2014). Of 14 of 61 articles or 23% (that provided data to GBD 2010) were rated as high-quality (Fuentes and Tejedor, 2014).

GBD 2010 shows that the stroke mortality has increased in the low- to middle-income countries (Feigin et al., 2014) but has not been shown as such: based on a review of 56 population-based studies, stroke case–fatality only differed by –0.6% between 1980 and 2008 in the low- to middle-income countries (Feigin et al., 2009). Among those who have survived stroke, disabilities such as limb weakness, spasticity, pain, psychological issues, language impairment and functional limitations set in, and are frequently profound (Norrving and Kissela, 2013).

This thesis reports on stroke research in Malaysia, and we impress on the reader how more stroke studies must be done in Malaysia. The reasons are: a) the great impact of stroke in the low- to middle-income countries such as Malaysia, b) lack of Malaysian stroke data (Malaysia is categorised as a country with moderate epidemiological data (Asian Acute Stroke Advisory Panel., 2000), and c) the need to improve understanding of stroke risk factors and outcome to help in stroke prevention and treatment (Johnston et al., 2009).

1.2 Definition of stroke

Stroke is defined as: a clinical syndrome which is characterised by ‘rapidly developing clinical symptoms and/or signs of focal, and at times global (applied to patients in deep coma and those with subarachnoid haemorrhage), loss of cerebral function, with symptoms lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin’ (Lindley, 2008, Hatano, 1976, O'Donnell et al., 2010a, World Health Organization, 2014). Stroke is also known as ‘brain attack’ (Lindley, 2008) or ‘cerebrovascular accidents’ (Adams,

2007), which indicate that the major pathological events in stroke occur in the first 24 hours (Lindley, 2008).

The definition excludes transient ischaemic attacks and stroke-mimicking symptoms secondary to trauma, shock and hypertensive encephalopathies (World Health Organization, 2014). Given the non-adherence to the definition, transient episodes of cerebral ischaemia (less than 24 hours) and cerebrovascular lesions discovered at autopsy without having shown clinical manifestations before death are not registered as stroke (Hatano, 1976, Good, 1990). The definition implies that stroke is a clinical diagnosis—without the need for imaging—but with the wider availability of brain imaging methods, radiological diagnosis is almost always requested by clinicians to help ascertain the final diagnosis of stroke.

1.3 Classification of stroke

Early studies required necropsy results to confirm the diagnosis and classification of stroke. With the advancement of brain imaging methods such as computed tomography (CT) scan, magnetic resonance imaging (MRI), MR angiography (MRA) and CT angiography (CTA), it has become possible to classify stroke in living patients and so to an extent identify the type of vascular lesion, location in the brain, stroke mechanism and stroke outcome (Caplan, 2011).

Generally, clinicians and stroke scientists categorise stroke into four major types: a) ischaemic stroke, b) primary intracerebral haemorrhage (PICH), c) subarachnoid haemorrhage (SAH) and d) undetermined stroke (Feigin et al., 2003). This rather crude classification is still being used in daily clinical practice in our setting in Malaysia.

The primary distinction between stroke types almost always requires clinical assessment and neuroimaging (CT or MRI) (O'Donnell et al., 2010a, Caplan, 2011). The distinction—stroke subtype classification—is useful in: a) daily clinical practice, and b)

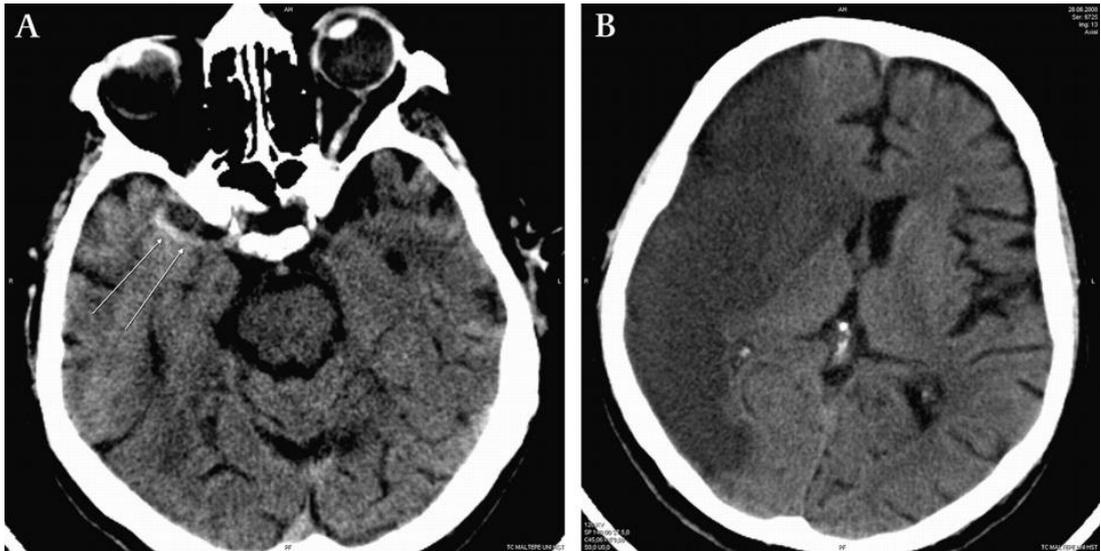
epidemiological and genetic studies, and should be able to classify (at least): a) ischaemic stroke (IS), b) haemorrhagic stroke (HS) and c) subarachnoid haemorrhage (SAH) (Amarenco et al., 2009). The classification for stroke also needs to be reliable (Department of Statistics, 2015).

1.3.1 Ischaemic stroke (IS)

Ischaemic stroke is the most common type of stroke (Shiber et al., 2010, Feigin et al., 2014, Feigin et al., 2003, Feigin et al., 2009). In ischaemic stroke, blood supply to the brain diminishes mainly due to one of these medical conditions: a) atherosclerosis (80% of ischemic stroke), b) stenosis of the small intracranial arteries, c) emboli from the heart or d) haematological causes, dysplasia or vasculitis (less common causes) (Good, 1990, Lindley, 2008).

In atherosclerosis, the atherosclerotic process activates platelet activation, which then forms a clot. This clot—a thrombi or emboli—develops *in situ* but some pieces of the clot may break away then migrate to the brain through a process known as embolization (Adams, 2007). Such a clot can occlude the small blood vessels, causing hypoxia (lack of oxygen). If this clot occurs in the brain, the oxygen-deprived brain tissue dies—a process known as cerebral infarction—and a patient with this condition will show the signs and symptoms of stroke.

CT scan images remains the mainstay of imaging in acute stroke and ischaemic stroke; CT scans show the hyperdense segment of a vessel, loss of grey–white matter differentiation and hypo-attenuation of the deep nuclei, as shown in Figure 1-1.



In

Figure 1-1 Signs of cerebral ischaemic on CT images. Hypo-attenuation of brain tissue is due to the increased brain water content inside the brain. Swelling of the gyri is due to the brain oedema. Hyperdense middle-cerebral artery (MCA) sign is a result of thrombus or embolus in the MCA. Blurred basal ganglia in MCA infarct and in insular ribbon sign refers to hypodensity and swelling of the insular cortex (Ahmed Abd Rabou and Frank Gallard, 2012).

1.3.2 Haemorrhagic stroke (HS)

In haemorrhagic stroke, a blood vessel ruptures and bleeds into the brain, the spinal cord or the adjacent structures, causing cell damage in the surrounding area (Torpy et al., 2010, Bernardini and DeShaies, 2001). Haemorrhagic stroke more often leads to a worse outcome than ischaemic stroke (Trevor and Fedi, 2013, Bernardini and DeShaies, 2001) because the bleeding inside the brain (haemorrhage) can progress for several hours with expansion of the haematoma (blood clot) (Adams, 2007, Brott et al., 1997, Kazui et al., 1996). This accumulation of blood inside a confined area, in this case, the brain, causes neuronal injury via hypoperfusion and ischaemia in the adjacent areas in the brain. Eventually, neuronal death occurs as a result of apoptotic (normal cell deaths) and necrotic (cell deaths due to injury) events and oedema (swelling) due to the accumulation of proteins within the clot, and as well as

disruption of the blood–brain barrier secondary to ischaemia (Adams, 2007, Kazui et al., 1996, Qureshi et al., 2001).

The three common subtypes of haemorrhagic stroke are: a) Primary intra-cerebral haemorrhage (PICH), b) Subarachnoid haemorrhage (SAH) and c) other intracranial haemorrhage such as extradural or subdural haemorrhage (Krishnamurthi et al., 2014, Bernardini and DeShaies, 2001).

1.3.2.1 Primary intracerebral haemorrhage (PICH)

In PICH, haemorrhagic stroke is caused by an arterial blood vessel in the brain that leaks or bursts (Lindley, 2008, Bernardini and DeShaies, 2001, Trevor and Fedi, 2013). In PICH, the bleeding from the artery spreads along the planes of white matter, leaving areas of intact neural tissues surrounding the haematoma (Adams, 2007, Qureshi et al., 2001).

1.3.2.2 Subarachnoid haemorrhage (SAH)

With SAH, bleeding occurs primarily in the subarachnoid space and the clinical hallmark is a history of severe and sudden headaches (van Gijn and Rinkel, 2001). In SAH, CT scanning is the first line of investigation, and SAH is characterised by the hyperdense appearance of the extravasated blood (van Gijn and Rinkel, 2001). Localised clots also form as a result of the trabeculations of tissue between the innermost pia and the arachnoid layers of the meninges (Adams, 2007). SAH may be pure subarachnoid bleeding though it may also arise from PICH (Lindley, 2008, van Gijn and Rinkel, 2001).

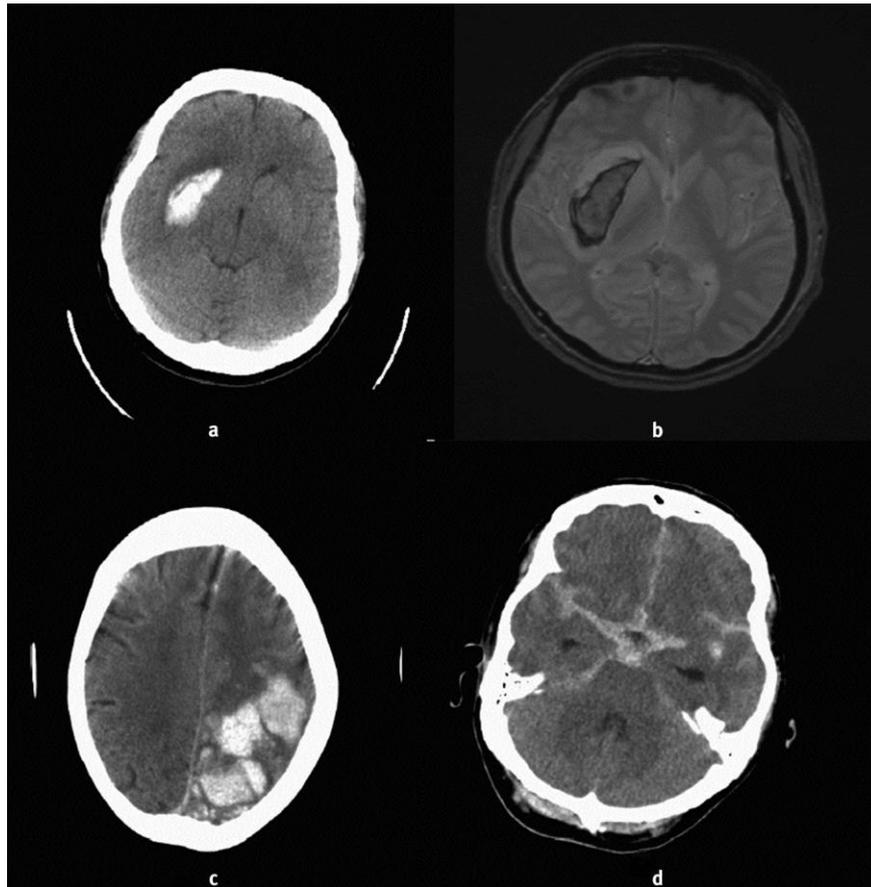


Figure 1-2 (a) Non-contrast CT brain and (b) MRI brain of a patient with a basal ganglia haemorrhage. This location is typical for a hypertensive intracerebral haemorrhage (ICH). (c) Non-contrast CT brain of a different patient demonstrating a large, acute left parietal intra-parenchymal ICH. The location and appearance of the lesion is suggestive of amyloid angiopathy. (d) CT brain of another patient, showing the classic appearance of a subarachnoid haemorrhage (Trevor and Fedi, 2013).

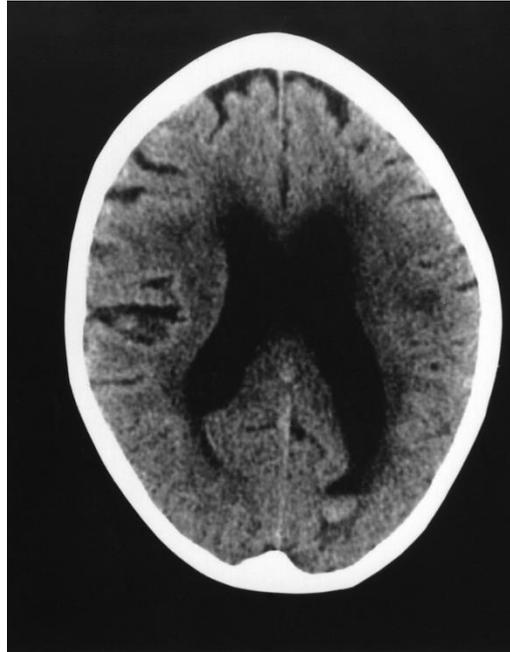


Figure 1-3 Subarachnoid haemorrhage (SAH) on CT images. CT scan is the first line of investigation for haemorrhagic stroke. This image shows positive sign for hyperdense appearance of extravasated blood in the basal cistern. False positive SAH could be due to generalized brain oedema. In this figure, sedimentation in the left occipital horn is the only positive finding (van Gijn and Rinkel, 2001).



Figure 1-4 Subarachnoid haemorrhage (SAH) on CT images appears as a high-attenuating, amorphous substance that fills the normally dark, CSF-filled subarachnoid spaces around the brain, appear white in acute haemorrhage, most evident in the largest subarachnoid spaces, such as the supra-sellar cistern and Sylvian fissures (van Gijn and Rinkel, 2001).

1.3.2.3 Other types of haemorrhagic stroke

Other types of haemorrhagic stroke include: a) extradural haemorrhage and b) subdural haemorrhage.



Figure 1-5 Extradural haemorrhage on CT scan images (Hacking and Gaillard, 2012b).



Figure 1-6 Acute subdural haemorrhage on CT scan images (Hacking and Gaillard, 2012a).

1.4 Coding of stroke

the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10) is the global health information standard for mortality and morbidity statistics of all diseases (World Health Organization, 2012, World Health Organization, 2011b). It facilitates the storage, retrieval, analysis and interpretation of health data and their comparison among the member states (World Health Organization, 2010).

The ICD-10 is the most widely used nosology (systematic classification of diseases). In a clinical setting, trained medical coders code diagnoses according to the ICD-10 using a computerised software programme (called 'a grouper') (O'Malley et al., 2005). It has been proven useful in the coding of stroke (Kokotailo and Hill, 2005).

The ICD-10 places stroke in Chapter IX inside Blocks I60–I69 (the cerebrovascular diseases blocks). Blocks I60–I69 contain more detailed codes for cerebrovascular diseases (stroke): a) Block I60 for SAH, b) Block I61 for intracerebral haemorrhage, c) Block I62 for other non-traumatic intracranial haemorrhage, d) Block I63 for cerebral infarction, e) Block I64 for stroke (not specified as haemorrhagic or infarction), f) Block I65 for occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction, g) Block I66 for occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction, h) Block I67 for other cerebrovascular diseases, i) Block I68 for cerebrovascular disorders in diseases classified elsewhere, and j) Block I69 for sequelae of cerebrovascular disease (World Health Organization, 2015).

The ICD-10 shows good coverage, with the percentage of correct coding ranging between 91% and 100% depending on stroke subtype. The limitations of ICD-10 include generalisation of results and misinterpretation of doctors' clinical diagnosis by the coders (Kokotailo and Hill, 2005). A validation study supported the use of routinely collected administrative data like ICD-10 for stroke diagnosis in a clinical setting (Aboa-Eboule et al.,

2013) provided the coding is restricted to acute stroke diagnosis. This is because in acute stroke setting, the incidence estimate based on the ICD shows close agreement with the true stroke incidence rate using a stroke register (Ellekjaer et al., 1999).

1.5 Epidemiology of stroke

In the 21st century, the importance of public health diseases such as cardiovascular disease, including stroke, have increased for a number of reasons, including: a) the decline of the worldwide scourge of infectious disease, b) better public health programmes that have improved health surveillance and c) economic advances that have led to increased life expectancy (Adams, 2007).

Although there has been a 42% decrease in stroke incidence in the high-income countries, stroke incidence in the low- to middle-income countries has increased by more than 100% (Feigin et al., 2009). The trend is generalised because studies suggest that the geographical variations in stroke incidence and prevalence are small (Feigin et al., 2003). While the geographical variation of stroke incidence is small worldwide, the burden of stroke shows larger geographical variation (Feigin et al., 2014, Krishnamurthi et al., 2014). Unfortunately, most stroke burden is carried by the low- to middle-income countries (Feigin et al., 2014) and the less developed countries (Norrving and Kissela, 2013). In addition, the early stroke case–fatality rate in the low- to middle-income countries (in 2000–2008) was 25% higher than that in the high-income countries (Feigin et al., 2009).

Despite the alarming threat of stroke as a major public health problem in Asia, stroke epidemiology is not well studied in this region. In Asia, only a few countries such as China, Taiwan and Japan are actively engaged in stroke studies (Burke and Venketasubramanian, 2006). There was a call for more studies on stroke prevalence especially: a) with a study design

that utilises door-to-door surveying, b) with a focus on people aged over 85 years old, and c) on the verification of stroke type and mortality data (Feigin et al., 2003, Feigin et al., 2014).

Table 1-1 The International Classification of Disease 10th Revision (ICD-10), Chapter IX (Diseases of circulatory system), Blocks I60-I69 (Cerebrovascular diseases).

Block	Stroke subtypes
I60	Subarachnoid haemorrhage
I61	Intracerebral haemorrhage
I62	Other non-traumatic intracranial haemorrhage, Subdural haemorrhage (acute) (non-traumatic), Non-traumatic extradural haemorrhage
I63	Cerebral infarction
I64	Stroke, not specified as haemorrhage or infarction
I65	Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
I66	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
I67	Other cerebrovascular diseases
I68	Cerebrovascular disorders in diseases classified elsewhere
I69	Sequelae of cerebrovascular disease

Source: (World Health Organization, 2015)

1.5.1 Incidence, prevalence and burden for stroke

The current incidence and prevalence of stroke has led people to describe the situation as an epidemic (deVeber, 2003). A review of 56 population-based studies from 28 countries showed noticeable variation in the crude and age-adjusted stroke incidence (Feigin et al., 2009). Among the high-income countries, the total crude incidence rate (per 100,000 person-years) ranged from 125 to 460 in 1970–1979, 156 to 466 in 1980–1989, 131 to 451 in 1990–1999 and 112 to 223 in 2000–2008. Among the low- to middle-income countries, the total crude incidence rate for stroke was 15–50 per 100,000 population in 1970–1979; 202–217 per 100,000 population in 1980–1989; 167–281 per 100,000 population in 1990–1999 and 73–165 per 100,000 population in 2000–2008 (Feigin et al., 2009).

Between 1990 and 2010, the overall age-standardised stroke incidence in the low- and middle-income countries increased by 12% (from -3% to 22%) but the mortality rates for all

stroke types decreased significantly at 20% (from 15% to 30%) (Feigin et al., 2014). In the high-income countries, the age-adjusted stroke incidence rates decreased by 42% from 1970 to 2008 (from 163 per 100,000 person-years in 1970–1979 to 94 per 100,000 person-years in 2000–2008) (Feigin et al., 2009). The estimated age-adjusted incidence rate in 2010 was 138.9 per 100,000 person-years in the high-income countries and 182.6 per 100,000 person-years in the low- and middle-income countries (Feigin et al., 2014). Studies from 61 low-income and middle-income countries have shown that between 1990 and 2010, the incidence of haemorrhagic stroke and ischaemic stroke saw a significant increase of 22% (from 5% to 30%) and a non-significant increase of 6% (from -7% to 18%), respectively (Krishnamurthi et al., 2013).

Overall, the prevalence rates of stroke are about 10 per 1000 population, which corresponds to about 50 per 1000 population in those over 65 years of age and it is expected that a third of these will die from stroke (Lindley, 2008). In the high-income countries, the prevalence of stroke (per 100,000 population) was 411.3, 504.0 and 534.8 in 1990, 2005 and 2010, respectively (Feigin et al., 2014). In the low- and middle-income countries, the prevalence of stroke (per 100,000 population) was 278.0, 296.0 and 300.0 in 1990, 2005 and 2010, respectively (Feigin et al., 2014).

Globally, the WHO has estimated that in 2001, death from stroke in the low- and middle-income countries accounted for 85.5% of stroke deaths worldwide, and the disability-adjusted life years (DALYs) lost in those countries was almost seven times that of the high-income countries (Mathers et al., 2006).

GBD 2010 reported that between 1990 and 2010, the mortality rates for ischaemic stroke fell by 14% (from 9% to 19%) and that for haemorrhagic stroke was reduced by 23% (from -18% to 25%) in 61 low-income and middle-income countries (Krishnamurthi et al., 2013).

Stroke is also the primary cause of death in China and other countries in Eastern Asia (Adams, 2007, Asplund, 1996, Kubo et al., 2003, Stegmayr et al., 2000). The Asian populations have higher stroke mortality than the Western populations (Hata and Kiyohara, 2013, Ueshima et al., 2008).

Among the Asian countries, Japan has the lowest stroke mortality, similar to that in the Western countries, with age-adjusted rates of stroke mortality of less than 50 per 100,000 person-years, while in Malaysia, the estimated stroke mortality rate is between 50 and 100 deaths per 100,000 person-years (Hata and Kiyohara, 2013, Ueshima et al., 2008).

1.5.2 Proportional frequency of stroke types

A review of 15 population-based studies (mostly in the developed countries) between 1966 and 2002 showed that the proportional frequency of stroke types was as follows: 67.3% to 80.5% of strokes were ischaemic, 6.5% to 19.6% were PICH, 0.8% to 7.0% were SAH and 2.0% to 14.5% were undefined (undetermined) (Feigin et al., 2003).

A later review based on 56 population-based studies among the high-income countries (2000–2008), the proportion of ischaemic stroke ranged from 73% to 90%, while that for PICH and SAH ranged from 9% to 13% and from 1% to 6%, respectively (Feigin et al., 2009). The review also found that between 2000 and 2008, among the low- to middle-income countries, the proportion of ischaemic stroke, PICH and SAH ranged from 54% to 85%, from 14% to 27% and from 5% to 10%, respectively (Feigin et al., 2009). The ratio of cerebral infarction to cerebral haemorrhage was approximately 2:1 to 3:1 in East Asian countries whereas it is 5:1 to 10:1 in the Western countries (Ueshima et al., 2008).

Among high-income countries over these decades (1970 to 2008), the pooled proportional frequency of stroke subtypes has only slightly changed: ischaemic stroke was 74%

(1980 to 1989), 77% (1990 to 1999) and 82% (2000 to 2008) and the frequencies for PICH was 9%, 13% and 11%, respectively (Feigin et al., 2009). Data for low-income and middle-income countries were only available for 2000 to 2008 and it showed that the proportional frequency for ischaemic stroke was lower in low-income to middle-income countries than in high income countries (67% and 82%, respectively) but the frequencies for PICH and SAH were twice the frequencies in high-income countries for the same decade (Feigin et al., 2009).

In countries where there has been an increase in the proportional frequencies for haemorrhagic stroke, this was largely the result of an increase in the prevalence of risk factors such as raised blood pressure and smoking (Feigin et al., 2009, Tosteson et al., 1990, Mensah, 2008). Other contributing factors include an increase in the prevalence of hypertension and chronic hepatic disease. It is also the case that clinical measures to diagnose and combat stroke, such as the , use of anti-coagulation therapy and the increase use of CT scans which enables earlier detection (by better detection of intracranial bleeding), contribute for the increase in the proportion of haemorrhagic strokes (O'Donnell et al., 2010a, O'Donnell et al., 2010b, Shiber et al., 2010). The situations in India, Pakistan, Bangladesh and Indonesia suggest that with the increase of uncontrolled hypertension, the percentage of haemorrhagic stroke is high (19% to 46%) (Kim, 2014). Unfortunately, many countries especially in the low-income and middle-income countries do not have previous data on stroke subtypes to enable researchers to look for the change in trend of the proportional frequencies for stroke.

1.5.2.1 Proportional frequency for ischaemic stroke (IS)

Generally, the proportion of all strokes that are ischaemic is higher among the high-income countries than among the low- to middle-income countries (Feigin et al., 2009). Among patients with ischaemic stroke, a large proportion of patients in the Western population has

extracranial large artery embolism while small vessel occlusion or intracranial atherosclerosis are more common in the Asian populations (Department of Statistics, 2015).

Ischaemic stroke comprised 51.4% and 62.4% of all stroke cases in Okinawa and China, respectively (Burke and Venketasubramanian, 2006), and in India, 83.9% of first-ever stroke patients were classified as ischaemic stroke (Sridharan et al., 2009). A recent study showed that in the Korean population, the proportion of ischaemic stroke in comparison to haemorrhagic stroke increased from 64.7% to 76.1% over the span of a few years (Hong et al., 2013). In Malaysia, ischaemic stroke comprised 56.3% of all strokes (Jaya et al., 2002).

1.5.2.2 Proportional frequency for haemorrhagic stroke (HS)

PICH and SAH account for 10% to 27% of all stroke worldwide, respectively (Feigin et al., 2009). These estimates however vary with time and geographical area. For example, an earlier review of population studies worldwide showed that the proportional frequency of subarachnoid haemorrhagic stroke was between 1% and 7% of all stroke (Feigin et al., 2003).

The proportion of haemorrhagic stroke in the Asian populations is higher than that in others. For example, a study in Korea showed that in 2000, the proportion of patients with haemorrhagic stroke was 35.3% (Hong et al., 2013). In Okinawa, Japan, and in China, the proportions of patients with SAH were 9.3% and 1.8% (of all stroke), respectively (Burke and Venketasubramanian, 2006). In India, 10.6% of patients with stroke had intracerebral haemorrhage (ICH) and 5.5% had SAH (Sridharan et al., 2009).

The scenario in Malaysia shows that from a hospital-based study in Kelantan, haemorrhagic stroke comprised 32.9% and subarachnoid haemorrhagic stroke comprised 10.8% of all in-hospital stroke cases (Jaya et al., 2002).

1.5.3 Stroke trends

Between 1990 and 2010, the age-standardised incidence of stroke significantly decreased in the high-income countries but increased though non-significantly (by 12%). Between 1990–2010, mortality rates decreased significantly in both the high-income (37%) and low- and middle-income countries (-20%) but the overall global burden of stroke was still great and increasing (Feigin et al., 2014).

Most of the burden of stroke falls on the low- and middle-income countries (Krishnamurthi et al., 2013), with the countries in the South-East Asia (SEA) region having the highest stroke incidence and mortality rates (Krishnamurthi et al., 2014).

In countries such as Japan, where stroke incidence is decreasing, the attributing factors could be: a) improvement in blood pressure management in hypertensive respondents, and b) reduction in the rate of smoking. Unfortunately, the decrease has slowed from 1998 to 2000, possibly due to a steep increase in metabolic risk factors and inadequate blood pressure reduction (Kubo et al., 2003, Hata and Kiyohara, 2013).

With regards to stroke mortality, in stroke referral centres recording decreasing stroke mortality trends (during hospitalisation), the decrease could be due to: a) advancement of healthcare services (Zhang et al., 2007), and b) improvements in acute cardiac and medical treatment (Cheng et al., 2005).

1.6 Demography of Malaysia and Kelantan and stroke as a public health problem

Malaysia is located between 2° and 7° north of the Equator. Peninsular Malaysia is separated from the states of Sabah and Sarawak by the South China Sea. To the north of Peninsular

Malaysia is Thailand while its southern neighbour is Singapore. Sabah and Sarawak are ringed by Indonesia; Sarawak also shares a border with Brunei.

1.6.1 Demography of Malaysia

The demographic information of Malaysia is shown in Table 1-2. The urban population for Malaysia has grown from 26.6% (1960) to 49.8% (1990) and 74.7% (2015), leaving the rural population at near 25% in 2015. The percentage (of total) for population ages 65 and above has increased from 3.6% (1990), 4.0% (2000), 5.0% (2010) and 5.9% (2015) (The World Bank, 2016). The geographical map and the population density of Malaysia are shown in Figure 1-7.

1.6.2 Demography of Kelantan, Malaysia

Malaysia consists of 14 states. The state of Kelantan has a large area of 15,105 km² and in 2012, the population was 1.65 million; 50.4% was male and 49.6% was female. The average annual population growth rate for the three years spanning 2010 and 2012 were 1.3%, 1.8% and 2.0%, respectively. The demography of Kelantan is summarised in Table 1-3.

Table 1-2 Demography of Malaysia for year 2010, 2011, 2012 and 2013.

	2010	2011	2012	2013
Area (km ²)	330,290			
Population (million)				
Total	28.59	29.06	29.51	29.92 ^P
Male	14.73	14.98	15.22	15.43 ^P
Female	13.86	14.08	14.29	14.49 ^P
Life Expectancy (years)				
Male	71.9	72.2 ^P	72.4 ^e	72.6 ^e
Female	76.6	76.8 ^P	77.0 ^e	77.2 ^e
Gross Domestic Product (GDP)				
GDP at current prices (RM million)	797,327	885,339	941,949 ^e	986,733 ^P
GDP at constant 2005 prices (RM million)	676,653	711,760	751,934 ^e	787,611 ^P
GDP Growth (%)	7.4	5.2	5.6 ^e	4.7 ^P

Notes: Between 2011 and 2013, the population estimates were based on the adjusted Population and Housing Census of Malaysia for 2010. ^P Preliminary ^e Estimate
Source: (Department of Statistics, 2015)

1.6.3 Stroke burden in Malaysia

In 2011, diseases of the circulatory system were the fifth principal cause of hospitalisation and the main cause of death (25.6% of all deaths) in all Ministry of Health hospitals in Malaysia (Ministry of Health, 2011).

In 2012, cerebrovascular diseases (ICD-10 Blocks I60–I69) contributed to 28,272 hospital discharges among all Ministry of Health hospitals, which is equivalent to a 97.6 discharge rate per 100,000 population, and comprised 17.8% (28,272/158,788) of all diseases of the circulatory system (Ministry of Health, 2012). In 2012, the total number of deaths from stroke in all Ministry of Health hospitals in Malaysia was 4162, or a 14.4 mortality rate per 100,000 populations. This was 33.8% (4162/12,312) of all deaths from diseases of the circulatory system, second to ischaemic heart diseases (Ministry of Health, 2012).

Malaysia sits among the low-income and middle-income countries specifically in the Southeast Asia region. Based on the estimates made by the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD 2010), most of the burden (98.6% incident strokes, 52.2% prevalent strokes, 70.9% stroke deaths, and 77.7 DALYs lost) exerted by the low-income and middle-income countries (central, east, south, and southeast Asia; eastern Europe; Andean, central, southern, and tropical Latin America and the Caribbean; and North Africa and the Middle East, Oceania, central, east, southern, and west sub-Saharan Africa) (Feigin et al., 2014). The Southeast Asia region had the highest mortality (age-standardized stroke mortality above 136.7 per 100,000 people) in 2010 (Figure 1-8).

We used Pubmed Advanced Search Builder (www.pubmed.com) to review all the studies on stroke ever conducted in Malaysia. We set the title as 'Malaysia' and 'Stroke' in the search builder fields. . The search yielded only fifteen articles with the latest publication published in May 2016 in Neurology Research. There was one study each on stroke registry,

on brief review of stroke in Malaysia, on stroke fatality, on longitudinal study (but small population covered) and on stroke pattern in Kelantan, Malaysia (Aziz et al., 2015, Loo and Gan, 2012, Jaya et al., 2002, Wan-Arfah et al., 2015, Neelamegam et al., 2013). Two studies looked at the costing of stroke and two more on ischaemic stroke (Akhavan Hejazi et al., 2015, Nor Azlin et al., 2012, Hamidon and Raymond, 2003, Aziz et al., 2016). A total of three studies on outcome of treatments (Sudirman et al., 2015, Ali et al., 2015, Aziz et al., 2016) and four studies on psychological and functional aspects after stroke (Sahathevan et al., 2014, Rameezan and Zaliha, 2005, Khoo et al., 2013, Mohd Zulkifly et al., 2016). Almost all were single-centre studies and none was a case-control study. The longitudinal study was conducted in a small area in the Northwest of Malaysia (Neelamegam et al., 2013).

1.6.4 Health care system in Malaysia

In Malaysia, there are two providers of healthcare: the government sector (majority) and the private sector. The government run two types of public hospital, one is run by the Ministry of Health, the other (known as University Hospital) - run by the Ministry of Higher Education. Both types of hospital are subsidized by the government. The Malaysian health care system consists of mainly tax-funded and government-run universal services. Its public sector health services are organized under a civil service structure and are centrally administered by the Ministry of Health. The public sector provides about 82% of inpatient care. People can choose between public and private services depending on their ability to pay (Safurah et al., 2012).

In Malaysia there is no system of family doctors or general practitioners for patients as practiced by NHS in the UK. Patients needing the kind of primary healthcare for which one might see a GP in the UK, typically visit a public hospital (for example to get initial check-up). This decision about where to go is mostly based on patients' preference and logistics. There

are also small private clinics manned by 1 to 3 medical doctors who provide outpatient care and minor surgery (Pagalavan, 2011). But again, unlike the NHS, these private clinics do not have designated patients (meaning they are not the designated GP for particular patients). Patients who come to these clinics pay for the service in full.

For the outpatient services at polyclinics, district hospital and tertiary hospitals, patients need to pay 1.00 RM (Ringgit Malaysia) for the non-specialist treatment and 5.00 RM (Ringgit Malaysia) to 30.00 RM (Ringgit Malaysia) for the specialist treatment. If the patients are admitted, the charge will be free (3rd class), 5.00 RM (Ringgit Malaysia) (2nd class) and 10.00 RM (Ringgit Malaysia) for the 1st class. These charges are for Malaysians only. Other costs for the special services and investigations can be found on the Ministry of Health website (Ministry of Health Malaysia, 2013a, Ministry of Health Malaysia, 2013b). The public but non-Ministry of Health hospitals such as the university hospitals act as teaching hospitals. These type of public hospitals have bigger autonomy than then Ministry of Health hospitals but are still managed according to government regulations.

Patients can also go directly to emergency departments or smaller government hospitals (district hospitals or district health clinics) for treatment and check-ups. At these smaller public clinics/hospitals, doctors can examine patients and refer them to hospitals, such as HUSM or HRZP in Kelantan, if necessary. The place of referral depends on the distance from the referring hospitals/clinics and the availability of expertise at the referred hospitals. People generally have good physical access to health facilities (health facilities was available to 95 percent of the population in 2004) (Merican et al., 2004). And 92% of the urban population live within 3km of a health facility and nearly 69% for the rural population. As a signatory to the 1978 Alma Ata Declaration, Malaysia affirms health as a fundamental human right, in that no one should be denied because of ethnicity, race or religion (Merican et al., 2004, Safurah et al., 2012).

The private sector provides services used by about up to 25% of population (Zan, 2016). Their services concentrate mainly in the cities and all hospital charges are borne by the patients. All patients receiving treatment from the private hospital pay the charges using medical insurances. These medical insurances provide medical plans and the plans are different between different insurance companies and they cover treatment and admission at the private and government public hospitals. Unlike national health insurance – a legally enforced scheme of health insurance that insures a national population against the costs of health care – these private insurances are non-compulsory to the general Malaysian population. In 2014, the Ministry of Health hospitals had 2,465,162 admissions and the non-Ministry of Health hospitals 148,450 admissions, a total of 2,613,612. In the same year, the total number of admissions to the private hospitals was 1,083,201 (Ministry of Health Malaysia, 2015).

There has been no study to our knowledge that compares the quality of services and care between the public and the private services in Malaysia. But it has been shown that privatization increases the drug price in Malaysia (Babar and Izham, 2009). A systematic review did not support the claim that the private sector is usually more efficient, accountable, or medically effective than the public sector in low- and middle-income countries (Basu et al., 2012). Malaysia implements the National Referral System to ensure provision of integrated health care to the population with the key objective of providing greater equity, accessibility and better utilization of resources (Merican et al., 2004). The current healthcare system in Malaysia is considered as progressive and equitable to her population (Yu et al., 2008).

1.6.5 Stroke services in Kelantan, Malaysia

The study took place in the two largest hospitals in Kelantan, Malaysia: a) Hospital Universiti Sains Malaysia (HUSM) and b) Hospital Raja Perempuan Zainab II (HRPZ).

In Kelantan, there are two main tertiary hospitals - HUSM (a public hospital run by the Ministry of Higher Education) and HRPZ (a public hospital run by the Ministry of Health, Malaysia), six kilometres apart at the capital of Kelantan state, Kota Bharu. Both are the two largest hospitals in Kelantan and both are owned by the government (public hospitals). The public has full and equal access to both hospitals and choosing which hospital depends on the preference of the patient. Both hospitals provide advanced medical and surgical services. Each of them has its own niche area for example, HUSM is in the neuromedical and neurosurgical services and HRPZ for endocrine surgical services and respiratory services.

HUSM is in Kubang Kerian town, six kilometres (km) from the capital of Kelantan, Malaysia. Its centroid coordinates are 6.099° latitude and 102.281° longitude. It is the second largest hospital (in Kelantan, Malaysia). And it is also categorised as a public hospital and is managed by the Universiti Sains Malaysia.

HRPZ was previously known as Hospital Kota Bharu. It is the largest public hospital in Kelantan. HRPZ is a Ministry of Health hospital and is in the capital of Kelantan. The centroid of HRPZ is at the coordinates 6.125° latitude and 102.246° longitude.

1.6.5.1 Hospital Universiti Sains Malaysia (HUSM)

Hospital Universiti Sains Malaysia (HUSM) is part of the Health Campus of the Universiti Sains Malaysia. It started with 36 beds in 1983 and now in 2008 it has 747 beds. The missions of HUSM are to provide new patient services and to become a full-fledged medical centre. To achieve the missions, HUSM acts as the referral hospital and also a teaching hospital to the population in the east coast of Malaysia (Kamari, 2009). HUSM has these services:

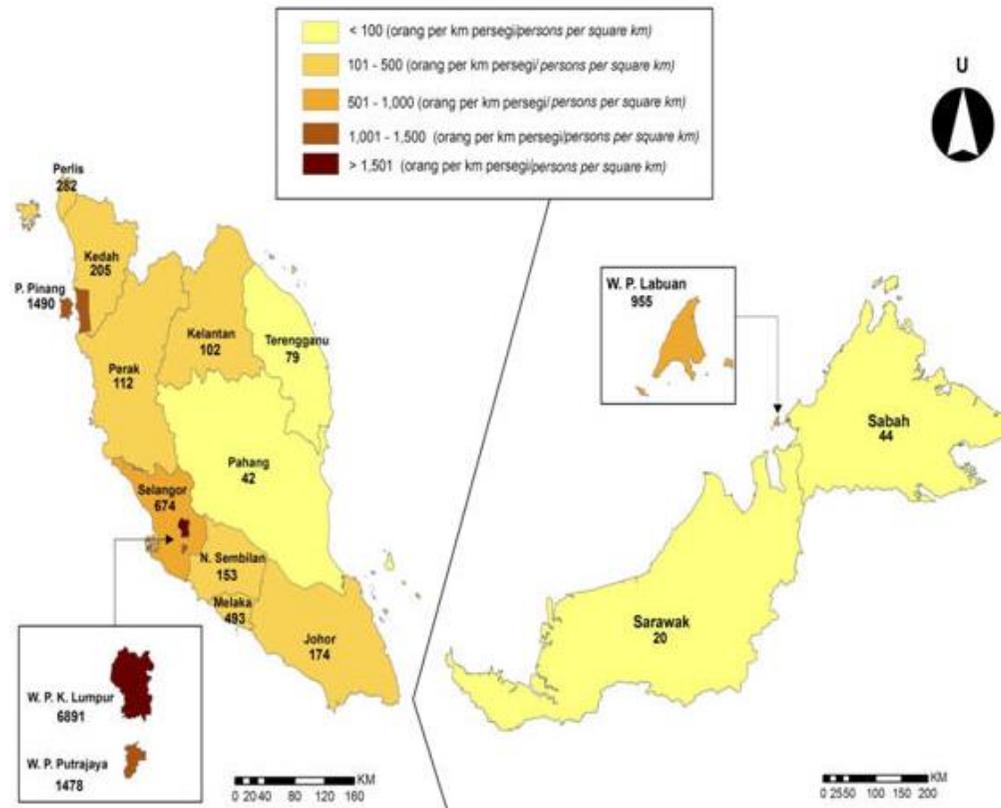


Figure 1-7 Map of Malaysia and the population density by state, Malaysia for year 2010 (Department of Statistics, 2015) .

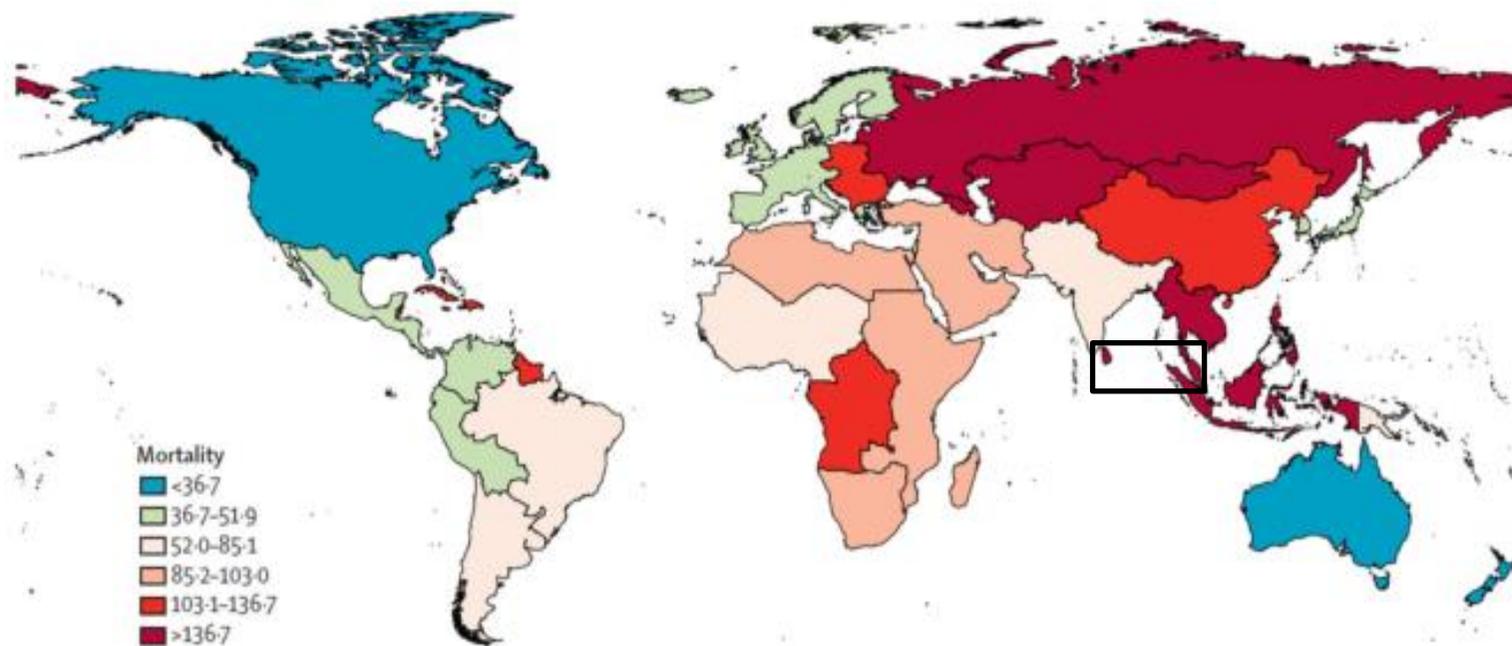


Figure 1-8 Age-standardized stroke mortality per 100,000 people in 2010 (Feigin et al., 2014). Malaysia – comprised of Peninsular Malaysia and East Malaysia in the Borneo – is located inside the box.

Table 1-3 Demography of Kelantan, Malaysia for year 2010, 2011, 2012 and 2013.

	2010	2011	2012	2013
Population (million)				
Total	1.59	1.62	1.65	1.68 ^P
Male	0.8	0.81	0.83	0.85 ^P
Female	0.79	0.81	0.82	0.84 ^P
Average Annual Population Growth Rate (%)	1.3	1.8	2	1.9 ^P
Gross Domestic Product (GDP)				
GDP at constant 2005 prices (RM mil.)	11,991	12,780	13,476	13,963
GDP per capita at current prices (RM)	9,322	10,363	10,568	10,677
GDP Growth (%)	4.9	6.6	5.4	3.6
Employment				
Labour Force ('000)	547.2	554.3	589.2	615.2
Employed ('000)	534.2	543.5	575.6	598.1
Unemployed ('000)	13.1	10.8	13.6	17.1
Labour Force Participation Rates, LFPR (%)	57	56.2	58.2	59.3
Unemployment Rate (%)	2.4	2	2.3	2.8

Notes: Between 2011 and 2013, the population estimates were based on the adjusted Population and Housing Census of Malaysia for 2010. ^P Preliminary ^e Estimate
Source: (Department of Statistic, 2015)

1. Outpatient service
2. In-patient service
3. Clinical support service
4. Laboratory service
5. Pharmaceutical service

HUSM has two CT scanning machines:

1. GE LightSpeed Plus 4-Slice CT (acquired in 2000)
2. Siemens Somatom Definition AS+ 128-Slice (acquired in May 2009)

The neuroscience service is a part of in-patient services and was initiated in 1984 and was formally run by the Neuroscience Unit in 2001. In 2006, the neuroscience service did a major revamp to provide a much better service combining experts from the neurological, neurosurgical, neurophysiological, radiosurgical and neuroradiological teams. It is one of the most active neuroscience unit in the field of traumatic brain injury, haemorrhagic stroke and neurooncology in Southeast Asia (Kamari, 2009).

1.6.5.2 Hospital Raja Perempuan Zainab (HRPZ) II

HRPZ is the oldest hospital in Kelantan. It began operating in the 1920s. It has 16 clinical departments, 9 clinical support services and 8 non-clinical support services. HRPZ provides, among many others, outpatient and in-patient services.

The Department of Medicine is one of the clinical departments in HRPZ. This department is the tertiary referral centre serving the population of Kelantan and North Terengganu (the neighbouring state). It provides general medicine and medical speciality services such as the endocrinology, gastroenterology, cardiology and neurology speciality services. The neurology speciality service is run by neurologists and medical officers. They provide in-patient and out-patient care for patients with neurology problems such as stroke. The department has a dedicated unit the Neurophysiology Unit. The Department of Medicine is the largest department in HRPZ with it has 238 total number of beds.

1.7 Risk factors for stroke

Risk factors, also known as risk markers, for stroke are the characteristics of an individual or population that are associated with an increased risk of stroke as compared with those without these characteristics (Lindley, 2008, Arboix, 2015, Jia et al., 2011, Adams, 2007). Understanding the risk factors for stroke is important because if these risk factors can be controlled in the general population, it could lead to reductions in the risk for stroke incidence and mortality (Hata and Kiyohara, 2013, O'Donnell et al., 2010a, O'Donnell et al., 2010b).

Traditionally, the risk factors can be divided into: a) non-modifiable risk factors, and b) potentially modifiable risk factors (Sacco et al., 1997, Jia et al., 2011).

1.7.1 Non-modifiable risk factors

The non-modifiable risk factors are factors that are not changeable and they include factors or variables such as age, sex, family history of stroke, geography and ethnicity.

Advancing age is the main non-modifiable risk factor for stroke (Ovbiagele and Nguyen-Huynh, 2011). Age is probably the single most important forecaster of a high risk for stroke (Adams, 2007). Age rapidly increases the likelihood of stroke after a certain age, such as 55 years (Rosamond et al., 2007, Rosamond et al., 2008). Age was reported to double the risk for stroke in those aged 65 years and above (Ovbiagele and Nguyen-Huynh, 2011). The risk for stroke is seven times higher among persons older than 70 years than in younger persons (Brott et al., 1997, Brott et al., 1986).

Although haemorrhagic and ischaemic stroke affect people from all ethnic groups in all parts of the world (Sudlow and Warlow, 1996, Sudlow and Warlow, 1997, Pongvarin, 1998), the rates vary between ethnicities. The variation could result from different diets and lifestyles, smoking prevalence, accessibility to health care and public health resources and different reporting quality (Adams, 2007, Asplund, 1996, Truelsen et al., 2006).

A positive family history of stroke is also a risk factor for stroke. This indicates that stroke is perhaps an inherited disease (Adams, 2007). For example, intracranial aneurysms and vascular malformations—components of multi-system genetic diseases—occur in family clusters, and stroke patients with a positive family history of stroke have approximately three to five times increased risk of aneurysmal rupture (Adams, 2007). First-degree relatives also have almost twice the risk for stroke (Ovbiagele and Nguyen-Huynh, 2011).

In some studies, sex is a risk factor for stroke but the findings are inconclusive in all studies. Some studies have found that women, in particular those with atrial fibrillation (AF)

(Wagstaff et al., 2014) and an older age (after 70 or 75 years old) (Bhattacharjee et al., 2011, Wagstaff et al., 2014), have a higher risk of stroke compared to men. The higher risk of stroke among women could be the result of pregnancy, preeclampsia or use of oral contraceptive pills before menopause (Bushnell, 2008) or could be due to increased cardiovascular risk after menopause (Lisabeth and Bushnell, 2012). Others have found that men have 20–30% higher risk of stroke compared to women (Ovbiagele and Nguyen-Huynh, 2011).

1.7.2 Modifiable risk factors

Studies have found that hypertension (Hata and Kiyohara, 2013, Ovbiagele and Nguyen-Huynh, 2011, O'Donnell et al., 2010b), diabetes mellitus (Ovbiagele and Nguyen-Huynh, 2011, Hata and Kiyohara, 2013, O'Donnell et al., 2010b), hypercholesterolaemia (Hata and Kiyohara, 2013, Ovbiagele and Nguyen-Huynh, 2011), smoking (Hata and Kiyohara, 2013, Ovbiagele and Nguyen-Huynh, 2011, O'Donnell et al., 2010b), alcohol intake (Ovbiagele and Nguyen-Huynh, 2011, O'Donnell et al., 2010b) and socioeconomic status (Ovbiagele and Nguyen-Huynh, 2011) are among the important modifiable risk factors for stroke.

Recently, there have been interests in the other modifiable risk factors for stroke such as social and economic status, weather, season, day of the week as well as biochemical and environmental factors and geography (Adams, 2007, O'Donnell et al., 2010b, Pedigo et al., 2011, Yang et al., 2015, Yadav, 2015, Storhaug et al., 2013, Howard et al., 2013, Maheswaran et al., 2005). Populations from a lower socioeconomic class have a higher risk of stroke compared to the more affluent and educated populations (Cox et al., 2006, Kumar et al., 2015). In addition, the incidence of stroke and in particular the incidence of intracranial haemorrhage correlates with cold weather. It also occurs more frequently on Mondays than other days of the week (Adams, 2007). Other new risk factors that have been identified and investigated include inflammation, infection, renal disease, dietary habit (Ovbiagele and Nguyen-Huynh,

2011), ratio of apolipoprotein B (ApoB) to ApoA (O'Donnell et al., 2010b) and homocysteine (Sacco et al., 1998).

Hypertension is the most important modifiable risk factor for all stroke subtypes and is a more potent risk factor for intra-cerebral haemorrhage (ICH) than for ischaemic stroke (IS) (O'Donnell et al., 2010b). A large multi-centre case-control study showed that five risk factors account for more than 80% of global risk for all stroke: a) hypertension, b) current smoking, c) abdominal obesity, d) diet and e) physical activity and with the addition of apolipoprotein, they account for up to 90% of the risk for stroke (O'Donnell et al., 2010b).

Among studies looking at risk factors for stroke such as cholesterol and body mass index (BMI), the findings have been inconclusive. For example, the independent positive relationship between total cholesterol and stroke mortality was reportedly absent in middle age (after allowing the effect of systolic blood pressure) and in those with systolic blood pressure below 145 mmHg (Prospective Studies et al., 2007). The finding from previous large studies on the relationship between body mass index (BMI) and stroke risk is not consistent (Song et al., 2004, Kurth et al., 2005). A large study shows that higher mortality for stroke (about 40%) was associated with each five kg/metre² in the upper BMI range (25 to 50 kg/m²) with no evidence of a positive association between BMI and stroke in the lower BMI range (Prospective Studies et al., 2009). The positive relationship between BMI and stroke risk can be largely attributed by the effect of BMI on blood pressure (Prospective Studies et al., 2009).

1.7.3 Risk factors for ischaemic stroke

Ischaemic stroke is positively associated with hypertension and homocysteine but is inversely associated with apolipoprotein and high-density lipoprotein (HDL) cholesterol (O'Donnell et al., 2010b, Gorgui et al., 2014, Sacco et al., 1998). Other risk factors include low socioeconomic status (Kumar et al., 2015) and cystatin C (Yang et al., 2015). In young Caucasian adults

(younger than 50 years old), the most common risk factors are dyslipidaemia, smoking, hypertension and patent foramen ovale (a heart valve defect) (Renna et al., 2014).

In INTERSTROKE study, however, it was reported that the increased concentration of total cholesterol was not associated with risk of ischaemic stroke but the ratio of non-HDL to HDL cholesterol was (O'Donnell et al., 2010b). The non-association between total cholesterol and ischaemic stroke is consistent with a finding from a large prospective observational studies (Prospective Studies et al., 2007).

1.7.4 Risk factors for haemorrhagic stroke

The risk factors for haemorrhagic stroke include: a) hypertension b) anticoagulant (blood thinning), c) cerebral aneurysms (enlargement of the blood vessels), d) substance abuse, e) family history of stroke, f) smoking, g) diabetes, h) high cholesterol, i) obesity and j) sedentary lifestyle (Torpy et al., 2010).

For intra-cerebral haemorrhage (ICH), the investigators in the United States of America (USA) reported that in their populations, the risk factors were: a) race (non-whites are at higher risk for stroke), b) interaction between race and age, c) systolic blood pressure (but not hypertension status), d) male sex and warfarin use, e) cholesterol levels and f) triglyceride levels (Howard et al., 2013, Sturgeon et al., 2007). In European populations, the risk factors for ICH were: a) hypertension, b) diabetes, c) triglyceride levels, d) short stature, e) history of psychiatric morbidity and f) smoking status (Zia et al., 2006). In the Asian male population, a pooled analysis showed that the risk factor for haemorrhagic stroke was systolic blood pressure, and to a lesser extent, smoking (Hyun et al., 2013). Other studies have found that total cholesterol levels, non-HDL cholesterol levels and ratio of non-HDL to HDL were associated with reduced risk (O'Donnell et al., 2010b, Tziomalos et al., 2009).

A systematic review of 14 longitudinal and 23 case–control studies showed that the most important and consistent risk factors for SAH include: a) smoking status, b) hypertension and c) excessive alcohol intake (Feigin et al., 2005).

1.8 Outcomes of stroke

Depending on when the patient presents and on the underlying causes, the outcome of stroke may be stable, progressive or completely resolved. In stroke outcome studies, the investigators mainly deal with: a) stroke death and b) the functional health status after stroke. Generally, PICH or very large ischaemic strokes contribute to an early mortality rate of 10% within the first week. A further 10% die within a month, and after about a year, a third of people with stroke will have died (Lindley, 2008).

Reliable estimates of the population mortality are one of the problems in many countries especially those in the low-income and middle income countries. The problems of estimating population mortality rates by cause in Malaysia and other countries in the Asian region include incomplete vital registration, expertise of certifying specialists and variation in mortality data. Inadequate case-ascertainment, incomplete health screening, inaccurate diagnosis, lack of investigative equipment and technology and low level of awareness may contribute to inaccurate information about stroke mortality in developing countries (Venketasubramanian, 1998, Thammaroj et al., 2005, Kim, 2014). Providing specialists is important in the less developed countries (Kim, 2014). The stroke workload for Malaysia was only 9.1%, much lower than Indonesia, Singapore and Thailand which was most likely due to the small number of neurologist; hence, many stroke patients may be attended by non-neurologists (Venketasubramanian, 1998).

1.8.1 Stroke mortality and stroke case-fatality

Stroke mortality is calculated by dividing the number of all new strokes by the population size in a specified time interval (Feigin et al., 2003, Bonita, 1992). Stroke case-fatality measures the proportion of people who die because of stroke within a specified time period (Bonita, 1992) or the proportion of fatal strokes in all first strokes within a specified time period (Feigin et al., 2003, Saposnik et al., 2008).

The mortality rate from stroke (M), stroke event rate (C), and case-fatality are related by $M=C \text{ times } F$ but in most populations, changes in stroke mortality were principally attributable to changes in case-fatality rather than changes in event rates (Sarti et al., 2003).

The burden of ischaemic and haemorrhagic stroke has increased significantly between 1990 and 2010. And the absolute number of incident haemorrhagic stroke was twice lower than that of ischaemic stroke, the overall global burden of haemorrhagic stroke (deaths and DALYs) was higher (Krishnamurthi et al., 2013). Stroke is the second leading cause of death worldwide; contributing to 10% of all deaths and 4% of all DALYs worldwide. If the current trend continues, there will be almost 12 million stroke deaths, 70 million stroke survivors and more than 200 million DALYs lost globally by 2030 (Feigin et al., 2014).

The Global Burden of Disease (GBD) study estimated that in 2001, ischaemic heart disease (IHD) and cerebrovascular disease (stroke) were the leading causes of death for low- and middle-income countries and for high income countries, responsible for 12 million deaths globally, or almost one-quarter of the global total (Mathers et al., 2006). In the South-East Asia (SEA) countries alone, stroke had caused 6.6% of total deaths in the region (World Health Organization, 2014). Figure 1-10 shows the geographic distribution of relative mortality from stroke and ischaemic heart diseases worldwide.

Worldwide estimates for the year 2002 (based on data from 192 WHO member states) showed that the age- and sex-adjusted mortality rates for cerebrovascular disease vary 10-fold between countries, from 24.5 per 100,000 population to 251 in 100,000 population (Johnston et al., 2009). Stroke mortality rates appear to be higher than ischaemic heart disease mortality in much of Africa and Asia and lower in North America, Western and Northern Europe and Australia (Kim and Johnston, 2011).

The Global Burden of Disease Study in 2010 reported that the highest overall, male and female mortality rate in SEA countries was 90.1, 105.0 and 77.5 per 100,000 people, respectively (Krishnamurthi et al., 2014).

Country-specific studies show that in India, 28-day stroke mortality in a community study was 27.2% (24.5% for urban populations and 37.1% for rural populations; 72.1% of these deaths occurred within 10 days of stroke) (Sridharan et al., 2009). In-hospital mortality in Singapore was 3.4%, 2.5% and 9.1% among Chinese, Malays and Indians, respectively (Sharma et al., 2012). In Malaysia, the estimated annual age-standardised mortality rates from stroke (per 100,000 population) was 79 for males and 82 for females based on WHO estimates, and higher—103 and 97 for males and females, respectively—based on more reliable adjusted local data (Hoy et al., 2013).

Two studies in Malaysia looking primarily at PICH reported that overall hospital-based fatality (2002–2003) was 43.9% (Sia et al., 2007) and 32.5% (2007–2009) (Yousuf et al., 2012). The stroke fatalities in PICH that occurred during the first 24 hours, first 2 days and first week of admission were 32.7%, 38.5% and 84.6%, respectively (Yousuf et al., 2012).

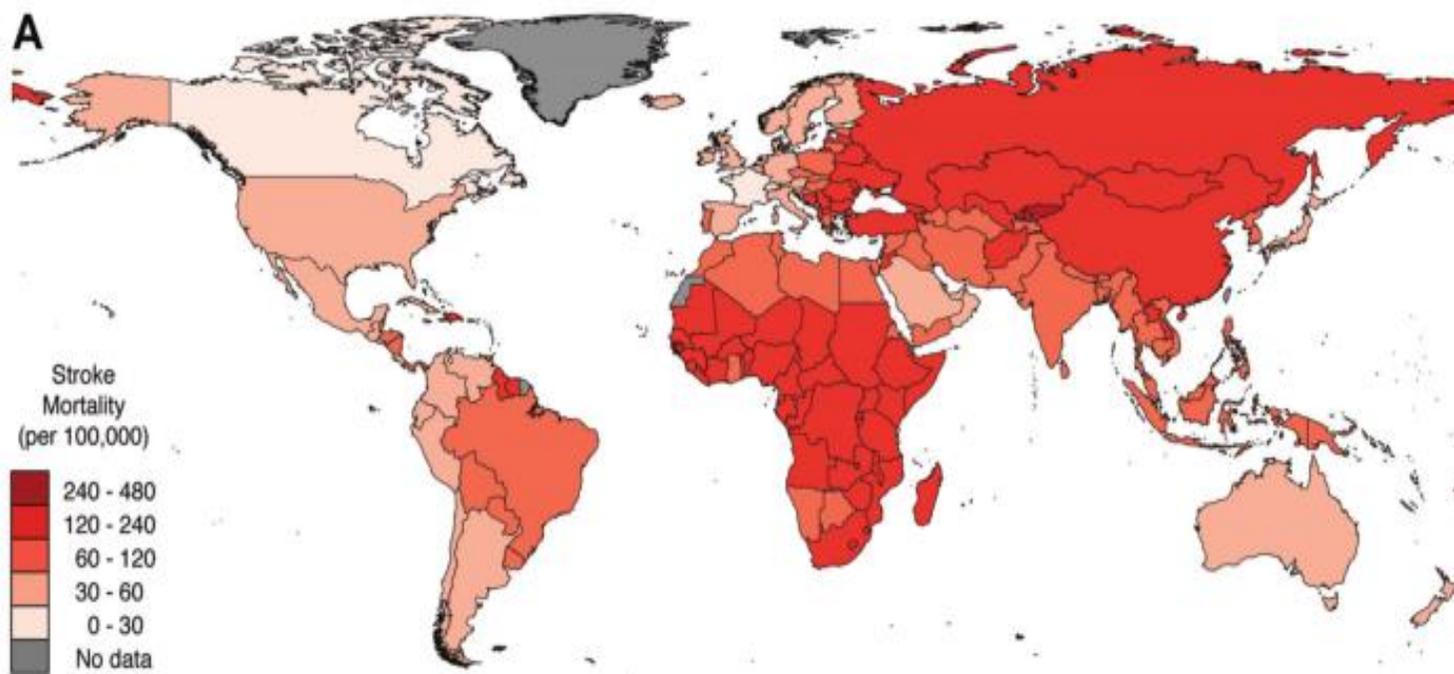


Figure 1-9 Worldwide geographic distribution of relative mortality from stroke (Kim and Johnston, 2011).

In Kelantan, a hospital-based study reported that the overall stroke mortality was 37.3% (January 1997 to December 1998), with a majority of patients (91.5%) dying during the first month, 62.7% of patients dying while in the ward (stroke fatality) and the rest dying at home (Jaya et al., 2002). A study in Kuala Lumpur, Malaysia, based on 2000–2001 data showed that the in-hospital case–fatality rate within 30 days after admission for ischaemic stroke was 11.7%; that for haemorrhagic stroke was 27.3% (Basri and Azman Ali, 2003).

1.8.2 Predictors for stroke mortality and stroke case-fatality

The risk for death in stroke patients is determined by multiple factors such as sex, age (Appelros et al., 2003, Olsen et al., 2011, Liu and McCullough, 2012, Nedeltchev et al., 2010), socioeconomic status (Eriksson et al., 2013, Cox et al., 2006, Lindmark et al., 2013), education (Lindmark et al., 2013), co-morbidities such as diabetes (O'Donnell et al., 2010a, Delbari et al., 2011, Fernandes et al., 2012), hypertension (Delbari et al., 2011), diastolic blood pressure on admission (Sharma et al., 2012), diet (Scarborough et al., 2011), stroke severity (Saposnik et al., 2008, Andersen et al., 2005, Appelros et al., 2003, Basri and Azman Ali, 2003, Sharma et al., 2012), drug treatment and availability and experience of stroke teams and physicians (Saposnik et al., 2008, Kita et al., 2009), time of admission (weekday versus weekend) (Ogbu et al., 2011), anaemia (Li et al., 2013), location of the infarction (Basri and Azman Ali, 2003), quality of primary care (Lee et al., 2011) and stroke subtype (Shigematsu et al., 2013).

In the Asian populations, the AASAP study (a multi-centre study based in 36 hospitals all over Asia) reported that the risk factors for early death in ischaemic stroke were age, diabetes mellitus, smoking, AF, ischaemic heart disease and previous anti-platelet treatment, and the risk factors for death in ICH were age and hypertension, which mirror the risk factors in Europe and North America (Wong, 1999).

In Malaysia, the predictors for ischemic stroke mortality were infarction of the mid-cerebral artery (MCA) area, AF, diabetes mellitus, Barthel Index and Glasgow Coma Scale (GCS) score (Basri and Azman Ali, 2003). The predictors of intracerebral mortality in Malaysian stroke patients were GCS score, ICH score, volume of haemorrhage, presence of intraventricular extension, presence of midline shift on CT images and posterior fossa bleeding among others (Yousuf et al., 2012, Sia et al., 2007).

1.8.3 Predictors of functional health status after stroke

Studies have shown the functional status, as well as other outcomes, after stroke are associated with age (Sohrabji et al., 2013, van Almenkerk et al., 2013, Khan et al., 2012, Kong and Lee, 2014), stroke severity, size of the stroke (van Almenkerk et al., 2013) and stroke type (Di Carlo et al., 2006, Paci et al., 2011, McNaughton et al., 2001).

One the best-known measures of functional status after stroke is the Barthel Index. It is the most-cited measure of disability in stroke rehabilitation trials in the published literature (Sangha et al., 2005). It was developed initially to act as a simple index of independence post-stroke and a useful scoring assessment in rehabilitation (Quinn et al., 2011). It has excellent reliability (Duffy et al., 2013), and the most popular version of the Barthel Index is the 10-item Barthel Index (scores from 0 to 100 with 5-point increments) (Shah et al., 1992, Tibaek and Dehlendorff, 2011, Balu, 2009). The maximal score is 100 (total independence) and the lowest score is 0 (totally dependent bedridden state) (Tibaek and Dehlendorff, 2011, Uyttenboogaart et al., 2007) .

1.9 Modelling binary outcome: Logistic regression model

1.9.1 Logistic regression model

Linear logistic regression is the commonest statistical method with which to model data with a binary outcome variable, (Hosmer et al., 2013, Hosmer et al., 2011). The other less used method is binary probit model; both – the linear logistic and probit models – are jointly known as the binary regression model (Long and Freese, 2006).

Logistic or logit models are popular in health and medical studies because: a) the common linear regression is not appropriate for binary data as the fitted values $x^T \beta$ may be less than zero or greater than one (Katz, 2011, Dobson, 2002), b) the logistic regression is mathematically flexible, c) the logistic model parameters provide the basis for clinically meaningful estimates of effect (Faraway, 2006, Hosmer et al., 2013) and lastly, d) they are computationally easier to model (Dobson, 2002).

Logistic regression produces linear predictor, which is the weighted combination of the independent, predictor or prognostic variable (Nashef et al., 1999, Roques et al., 1999, Knaus et al., 1981). The logistic specific form is:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

The quantity $\pi(x) = E(Y|x)$ represents the conditional mean of Y given x when the logistic distribution is used. A transformation of $\pi(x)$ is known as the logit transformation (or the linear logistic model) and it is a special form of the general logistic regression model (Dobson, 2002, Hosmer et al., 2013). This transformation is defined, in terms of $\pi(x)$, as:

:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] \quad (2)$$

$$= \beta_0 + \beta_1$$

The logit is the natural logarithm of the odds of the outcome. It can take on any value from minus to plus infinity. This means that in a regression analysis when the outcome variable is dichotomous, the model for the conditional mean (the probability) of the regression equation extends from zero to one (Katz, 2011, Hosmer et al., 2013).

1.9.2 Model fitting for a logistic regression

If Y is coded 0 or 1 the expression for $\pi(x)$ given in equation (1) provides the conditional probability that Y is equal to 1 given x. This is denoted as $\pi(x)$. It follows that the quantity $1 - \pi(x)$ gives the conditional probability that Y is equal to zero given x, $Pr(Y = 0/x)$.

The general method for estimating parameters in the logistic model is via the maximum likelihood. The contributions to the likelihood function for the pair (x_i, y_i) of data when Y is coded as zero (0) or one (1), is through this expression:

$$\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (3)$$

As the observations are assumed to be independent, the likelihood function is defined as:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

However, it is easier mathematically to work with the log of the equation using the expression known as *log-likelihood*:

$$L(\beta) = \ln[l(\beta)]$$

$$= \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (5)$$

To find the value of β that maximises $L(\beta)$, we differentiate $L(\beta)$ with respect to β_0 and β_1 by using the likelihood equations:

$$\sum [y_i - \pi(x_i)] = 0 \quad (6)$$

$$\sum x_i [y_i - \pi(x_i)] = 0 \quad (7)$$

In equation (6) and equation (7), it is understood that the summation is over i varying from 1 to n .

1.9.3 Model checking for a logistic regression model

In a logistic model, model checking includes the assessment of: a) summary measures of goodness of fit (a single number that summarises the overall model) and b) regression diagnostics (numbers that summarise the fit of individual subjects in data forming the overall model) (Hosmer et al., 2013, Dobson, 2002).

There are three general summary measures of goodness of fit in logistic regression: a) the Hosmer–Lemeshow statistics, b) the Pearson chi-square test and c) the receiver operating characteristic (ROC) (Hosmer et al., 2013).

The Hosmer-Lemeshow statistics measure the overall model fitness and is the commonest measure of fit in logistic regression (Dobson, 2002). It compares the probabilities of the outcome occurring (predicted by the logistic model) against the observed data. Similarly, another overall model fitness test, i.e. the Pearson chi-square test, also compares the difference between the expected probability of the outcome occurring (predicted by the logistic model) and the observed outcome in the data (Long and Freese, 2006).

The third measure of the overall goodness of fit is the area under the ROC curve, also known as the c-index (Royston and Altman, 2010, Hanley and McNeil, 1982). The ROC curve quantifies the ability of a logistic model to discriminate between patients having or not having the outcome of interest (Royston and Altman, 2010). Although there has not been any consensus for values for a model to classify the ROC curve based on its performance, it is generally accepted that for a logistic model, an ROC curve greater than 0.7 indicates good discrimination for the fitted model (Brown et al., 2007, Burd et al., 2006).

As many logistic data contain individual rather than grouped observations, logistic model checking uses covariate patterns derived from the data. Covariate patterns are observations containing the same values of all the explanatory variables (Dobson, 2002).

The residuals (the difference between the predicted outcome calculated from the model and the observed outcome) are calculated based on these grouped observations in the model (Long and Freese, 2006).

1.10 Modelling time-to-event outcome: Survival analysis and the Cox proportional hazard regression

In some epidemiological studies, researchers observe subjects until a certain time using the cohort study design. In a cohort study, there are at least two end points for each subject, where: a) the subject develops an event during the period of follow-up; this is also known as failure, or b) the subject does not develop an event at the end of the follow-up. Under these circumstances, the subject is said to be censored. In a cohort study, all participants either have the same follow-up time, or each participant has his or her own follow-up time (including different starting and end points). In the latter circumstance, survival or duration analysis is the method of choice for data analysis.

Survival analysis generally comprises three methods: a) non-parametric, b) semi-parametric and, c) parametric. Non-parametric analysis includes the Kaplan-Meier method useful to describe the survival probability. Semi-parametric analysis such as the Cox proportional hazard regression is the most commonly used method for analysing survival data in an epidemiological study. The parametric analysis include exponential, Weibull and log-normal survival models (Hosmer et al., 2011, Kleinbaum and Klein, 2012, Katz, 2011).

The Kaplan-Meier method and the Cox model are widely used in clinical studies with time-to-event data. In such studies, all subjects who do not develop the outcome of interest are grouped together, and they are known as the censored observations. Both the Kaplan-Meier and the Cox model require these assumptions about censoring: a) Independent censoring, b) Random censoring, and c) Non-informative censoring (Kleinbaum and Klein, 2012). The reasons for subjects being censored can be used to help properly assess the nature of the missing data. For example, if the subjects were censored because of a lack of response or side effects, then this will give rise to – in statistical terms – informative missing data. Outcome-related censored subjects are informative and should not be disregarded in an analytical study without careful thought. Sections 1.10.4, 1.10.5 and 1.10.6 will provide further details about the censoring mechanism, the implications of informative censoring and the methods of handling informative censoring. In Appendix H of this thesis, two recent imputation methods, a) the Gamma imputation method and b) the Risk Score imputation method, are described in greater length.

1.10.1 Descriptive method for survival analysis

The initial analysis for survival data includes describing the survival function of the subjects. The survival function is defined as ‘the fraction of individuals who does not have the event at any given time’ (Glantz, 2005).

The default estimator of survival function in most statistical software is the Kaplan–Meier estimator or product limit estimator. This estimator incorporates information from all observations, both uncensored and censored (Hosmer et al., 2011). Assume we have a sample of n independent observations denoted (t_i, c_i) , $i = 1, 2, \dots, n$ of the underlying survival time variable T and the censoring indicator variable C . Assume that, among the n observations, there are $m \leq n$ recorded times of failure and $n - m$ censored values. Let the number at risk of dying at $t_{(j)}$ be denoted n_i and the observed number of deaths be denoted d_i . The Kaplan–Meier estimator of the survival function at time t is defined as:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} \quad (8)$$

With the convention that $\hat{S}(t) = 1$ if $t < t_{(j)}$. d_i is denoted as the number of deaths observed at the first failure time and n_i the number of subjects alive at any specified time point (number at risk of dying or simply the number at risk).

1.10.2 The Cox proportional hazard regression

Semi-parametric survival analysis is adequate for comparing survival between different groups (levels) of a categorical covariate, but a fully parametric hazard function, which requires the specification of a baseline hazard function, yields more information from the data (Kleinbaum and Klein, 2012, Hosmer et al., 2011).

Semi-parametric Cox proportional hazard regression allows the incorporation of subjects with differing lengths of follow-up and censoring, either the result of loss to follow-up or of death due to other causes (Katz, 2011). One advantage of using Cox proportional hazard regression for statistical non-specialists is that it is more intuitive to interpret hazard

ratios than regression parameters estimated as effects on a time scale in parametric models or accelerated failure time models (Hosmer et al., 2011).

The form of the Cox regression model (hazard function) is specified as:

$$h(t, x, \beta) = h_0(t)r(x, \beta) \quad (9)$$

where $h_0(t)$ is the hazard function that changes as a function of survival time, while $r(x, \beta)$ characterises how the hazard function changes as a function of subject covariates. The ratio of the hazard functions for 2 subjects with a covariate denoted as x_1 and x_0 is:

$$HR(t, x_1, x_0) = \frac{h(t, x_1, \beta)}{h(t, x_0, \beta)} \quad (10)$$

hence, the model can be written as (Cox, 1986):

$$\begin{aligned} HR(t, x_1, x_0) &= \frac{h_0(t)r(x_1, \beta)}{h_0(t)r(x_0, \beta)} \quad (11) \\ &= \frac{r(x_1, \beta)}{r(x_0, \beta)} \end{aligned}$$

With this parameterisation (equation (11)), the hazard function becomes:

$$h(t, x, \beta) = h_0(t)e^{x\beta} \quad (12)$$

and the hazard ratio is:

$$HR(t, x_1, x_0) = e^{\beta(x_1 - x_0)} \quad (13)$$

1.10.3 Model fitting for the Cox proportional hazard regression

It is not possible to use the log-likelihood function in survival analysis because the error component must be specified; alternatively, an expression—the partial likelihood function—is proposed (Cox, 1986).

This partial likelihood function depends only on the parameter of interest and is given by the following expression (Hosmer et al., 2011) on page 75:

$$l_p(\beta) = \prod_{i=1}^n \left[\frac{e^{x_i\beta}}{\sum_{j \in R(t)} e^{x_j\beta}} \right]^{c_i} \quad (14)$$

where the assumption is: no tied times and it is often modified to exclude terms when $c_i = 0$, yielding:

$$l_p(\beta) = \prod_{i=1}^m \frac{e^{x_i\beta}}{\sum_{j \in R(t)} e^{x_j\beta}} \quad (15)$$

Where the product is over m distinct ordered survival times and $x_{(i)}$ denotes the value of the covariate for the subject with ordered survival time $t_{(i)}$. The log partial likelihood is:

$$L_p(\beta) = \sum_{i=1}^m \left\{ x_{(i)}\beta - \ln \left[\sum_{j \in R(t)} e^{x_j\beta} \right] \right\} \quad (16)$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of the equation with respect to β (beta), setting the derivatives equal to 0 and solving for the unknown parameter (Hosmer et al., 2011).

1.10.4 Censoring mechanism and informative censoring

Survival analysis becomes complicated because the failure times are unobserved for a proportion of individuals and the data only provide the last time that the subjects were under observation, known as censoring time. Also, there is always a possibility that the hazard of failure in those subjects at risk at a given time and who have not yet failed or been censored is different from the hazard at that time in those with the same values of all relevant covariates, but who have been censored. This contradicts the standard assumption that if an observation is censored at time C , the contribution to the likelihood is just the probability that lifetime T exceeds C . The fact that the censoring has occurred when it did has not altered the distribution of T , hence the censoring mechanism is irrelevant for inference about the distribution of T (Siannis et al., 2005).

In survival analysis, the censoring mechanism can be classified as (Shih, 2002):

- 1) Non-informative censoring (ignorable missing)
 - a. Missing completely at random (MCAR): the censoring mechanism is independent of both observed and missing data
 - b. Missing at random (MAR): the censoring mechanism is conditionally independent of missing data given observed data.
- 2) Informative censoring (non-ignorable missing)
 - a. Missing not at random (MNAR): the censoring mechanism is conditionally dependent on missing data given observed data

The critical distinction is between 1) and 2) because under 1), likelihood-based inferences correctly address properties of the uncensored process, which is often what is wanted. Standard survival analysis methods assume that censoring is non-informative.

One of the most important assumptions of censoring is that censored patients are considered to have survival prospects similar to the participants who continued to be followed (Ranganathan and Pramesh, 2012, Bland and Altman, 1998). In practice, censoring is almost invariably assumed to be non-informative or ignorable.

1.10.5 Implication of informative censoring: An example in stroke setting

Here, we consider an example using a study of stroke mortality to illustrate the relationship of censoring and the survival time. A survival analysis that uses the Kaplan–Meier method and Cox model assumes that censoring is non-informative. In this setting, ICU patients discharged alive from the hospital are assumed to be representative of all other individuals who have survived to this time of discharge but who are still in hospital. In this case, the distribution of the censoring time is unrelated to the distribution of the survival time.

This censoring assumption is 'non-informative' about the mortality pattern of the population. This is perhaps true if the censoring process operates at random (which is usually the case when mortality is assessed at a point in calendar time), provided that this time point is selected before the study is initiated (Resche-Rigon et al., 2006). However, this assumption is false if, for example, censoring is a result of a deterioration or an amelioration in the patient's health status. This is probably the case in neurology wards where patients are discharged in relatively good health, and thus potentially censored informatively. This can happen when the patient needs no more stroke care (due to amelioration or deterioration of their vital conditions). Patients are therefore discharged alive (censored) because they have a lower or higher risk of hospital death (due to stroke) than the average. These patients are therefore not the same patient population as those who stayed within the hospital. The censoring is therefore 'informative', and carries information about the remaining survival time (Resche-Rigon et al., 2006).

1.10.6 Methods to handle informative censoring

Several methods have been described to deal with the problem of informative censoring. These include imputation techniques for missing data, often together with sensitivity analyses to mimic best and worst-case scenarios, and considering the censoring event itself as an important end-point (Shih, 2002, Ranganathan and Pramesh, 2012).

By conceptualizing the censored failure times as missing data, the problem of non-independent censoring becomes a missing data problem. Using multiple imputation in such cases provides a natural context for also accommodating missing covariates (Jackson et al., 2014). It was developed to: a) remove bias in standard survival analysis when it is thought that the censoring may be informative and b) improve efficiency by imputing events times for individuals who were censored (Burkoff et al., 2016b, Hsu and Taylor, 2009).

A procedure has been developed to quantify the sensitivity of the conclusion from fitted Cox proportional hazards model when the independent censoring is in doubt, rather than focusing on the reasons why the assumption of independent censoring may be false (Jackson et al., 2014). This procedure uses intuition that censoring associates with either a harmful effect or protective effect. By modelling such association between censoring and failure, investigators gain an informed view of the model's assumptions and the plausible range of sensitivity parameter or parameters to investigate. In this approach, the censored observations provide unobserved failure times and multiple imputation is used to impute censored failure times (Burkoff et al., 2016a). Estimation proceeds via multiple imputation where censored failure times are imputed by the users. Users can define that censoring may be associated with either a harmful effect or protective effect. The basic idea is to assume a step-change in the hazard at the censoring time. The size of this step-change cannot be estimated from the available data, but instead is used as a sensitivity parameter. The censored observations provide missing (unobserved) failure times and the user then imputes multiple

censored failure times. The user can also set different parameter or parameters that quantify the departure from independent censoring. From there, a sensitivity analysis can be performed to explore the implications of a range of possible values of sensitivity parameters. Once the imputed datasets have been created, the user can estimate the model parameters of interest by fitting an analysis model to each of the imputed datasets. At the end, the resulting parameter estimates are combined using Rubin's rules (Jackson et al., 2014). For further details, see Appendix H.

Another procedure that utilises multiple imputation is known as the risk score imputation. It uses non-parametric multiple imputation. The risk score imputation procedure creates multiple imputed event times for those subjects whose event times were censored (Burkoff et al., 2016b). In this procedure, auxiliary variables were used to compare survival distributions. Covariates which are known or believed to be related both to the hazard of failure and the hazard of censoring should be included in the imputation process. To increase the efficiency of the imputation, the user can include covariates which are only related to the hazard of failure. This procedure assumes that each subject provides other information that is informative about the health condition of the subjects and can be used for prediction of both event and censoring times. In risk score imputation, to improve robustness, two proportional hazard (PH) models are used: a) a PH model for the event times and b) a PH model for the censoring times. Each risk score is a linear combination of auxiliary variables. It is centred and scaled by subtracting the mean and dividing by the standard deviation of the risk scores. Based on the imputing risk set, a non-parametric multiple imputation, Kaplan-Meier imputation, imputes a future event or censoring time for each censored observation (Hsu and Taylor, 2009). In risk-score imputation, the event time is drawn from a Kaplan-Meier estimator of the distribution of event times calculated from the observations in the imputing risk sets. (Hsu and Taylor, 2009). Once the data sets have been imputed, the user can perform standard time-to-event statistical analyses on the augmented datasets and the results are then combined to

produce parameter estimates and hypothesis tests. In this case, the correctness of the analysis relies on the correctness of the imputation model. For further details, see Appendix H.

1.10.7 Residuals and proportionality assumption for the Cox proportional hazard regression

An examination of model adequacy provides a measure of its validity and examining the residuals does this. The Cox proportional hazard model produces residuals that are not obvious. Residuals (the difference between the predicted and observed outcome) generated from the Cox model include: a) Schoenfeld residuals, b) scaled Schoenfeld residuals and c) Martingale residuals or Cox–Snell residuals (Kleinbaum and Klein, 2012, Hosmer et al., 2011, Cleves et al., 2010).

The proportional hazard assumption is vital in the interpretation of the estimated Cox regression parameters. It has a log-hazard function of the form (15)(Hosmer et al., 2011):

$$\ln[h(t, x, \beta)] = \ln[h_0(t) + x'\beta] \quad (17)$$

where the left part is the log of the baseline hazard function, and on the right, the linear predictor.

Methods to assess proportionality include: a) a plot of the log-hazard over time and b) a plot of scaled Schoenfeld residuals over time. A plot that fits the proportionality assumption will have the residuals lying ‘randomly’ scattered about the zero (Cleves et al., 2010, Hosmer et al., 2011). A plot of the log-hazard over time would produce two continuous curves: one for $x = 0, \ln[h_0(t)]$, and the other for $x = 1, \ln[h_0(t)] + \beta$ (Cleves et al., 2010).

1.11 Modelling longitudinal data analysis: Linear mixed models

Data with measurements taken repeatedly from a subjects is known as a wave are called longitudinal data or panel data (Diggle et al., 2002, Rabe-Hesketh and Skrondal, 2012). In this type of data, the analysis of choice must reflect the presence of correlation between successive measurements (Dobson, 2002, Diggle et al., 2002, Rabe-Hesketh and Skrondal, 2012).

There are two approaches to modelling longitudinal data: a) drop the usual assumption of independence between outcomes Y_i and model the correlation explicitly, or b) consider the hierarchical structure of the study design, where the outcomes are assumed to be independent but with the presence of correlation between a subject's measurements (Dobson, 2002). Problems in longitudinal data analysis can be grouped into: a) where regression of the dependent variable (Y) on independent variable (\mathbf{x}) is the scientific focus, and b) where correlation is the main interest (Diggle et al., 2002).

Different statistical models in longitudinal data include: a) random-effect models, b) fixed-effect models, c) dynamic models, d) marginal models (Rabe-Hesketh and Skrondal, 2012, Diggle et al., 2002), e) transition models and f) analysis of variance (Diggle et al., 2002).

1.11.1 Naïve analysis for the correlated data

Naïve analysis, sometimes called pooled analysis, utilises all observations from the subjects (assuming independence between subjects) with:

$$E(y_{ijk}) = \alpha_i + \beta t_k + e_{ij} \quad (18)$$

where Y_{ijk} is the score at time t_k ($k = 1, \dots, n$) for patient j ($j = 1, \dots, n$)

One way of fitting a model with naïve analysis is by using a generalised estimating equation (GEE), assuming independence between observations for the same subject, which can estimate the population-averaged coefficients (Diggle et al., 2002).

1.11.2 Multi-level models for longitudinal data

The multi-level approach is an alternative to analysing longitudinal data. Repeated observations on the same units are also clustered data (Rabe-Hesketh and Skrondal, 2012). For example, suppose a model is specified as:

$$Y_{jk} = \mu + a_j + e_{jk} \quad (19)$$

where a_j is the effect of time j . It is independent, identically distributed random variables with $a_j \sim N(0, \sigma_a^2)$. Similarly, the terms are independent, identically distributed random variables with $e_{jk} \sim N(0, \sigma_e^2)$. In this case,

$$E(Y_{jk}) = \mu \quad (20)$$

$$\begin{aligned} \text{var}(Y_{jk}) &= E[(Y_{jk} - \mu)^2] = E[(a_j - e_{jk})^2] \\ &= \sigma_a^2 + \sigma_e^2 \end{aligned} \quad (21)$$

$$\text{cov}(Y_{jk}; Y_{jm}) = E[(a_j - e_{jk})(a_j - e_{jm})] = \sigma_a^2 \quad (22)$$

for the times in the same subjects, and

$$\text{cov}(Y_{jk}; Y_{lm}) = E[(a_j - e_{jk})(a_l - e_{lm})] = 0 \quad (23)$$

for different patients (Dobson, 2002).

In a linear model where longitudinal data with Y_{jk} is the measurement at time t_k on subject j , it can be written as:

$$Y_{jk} = \beta_0 + a_j + (\beta_1 + b_j)t_k + e_{jk} \quad (24)$$

The parameter μ is a fixed effect and the term a_j is a random effect (an example of mixed model with both fixed and random effects). The parameters of interest are μ , σ_α^2 and σ_e^2 . β_0 and β_1 are the intercept and slope parameters for the population, a_j and b_j are the differences from these parameters specific to subject j , t_k denotes the time of the k_{th} measurement and e_{jk} is the random error term (Dobson, 2002).

In general, mixed models for normal response can be written in the form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (25)$$

where $\boldsymbol{\beta}$ is the fixed effects, and \mathbf{u} and \mathbf{e} are the random effects. The matrices \mathbf{X} and \mathbf{Z} are the design matrices. Both \mathbf{u} and \mathbf{e} are assumed to be normally distributed.

1.11.3 Model checking for the linear mixed models

We invoke statistical assumptions whenever we fit a statistical model. In a mixed model, the assumptions involve the structural and stochastic parts. The assumptions include checking the functional form, normality and homoscedasticity. A fundamental assumption of the mixed model is that the level-2 residuals or \mathbf{U}_i (a sample of independent unobservable variables from random effects distribution) are independent of the explanatory variable (Diggle et al., 2002).

Checking the functional form involves looking at the plot of ‘outcome’ against ‘predictor’; for normality checking, visualising residual distribution is adequate; and to check for homoscedasticity assumption, we plot the residuals against the predictor (Singer and Willet, 2003).

1.12 Issues in model building

1.12.1 Effect of the third variable

In epidemiology, model building serves to: a) produce a prediction model and b) adjust for confounding variables (Neter, 1996). When prediction is the aim, model fit and mean square prediction error are the main criteria used to evaluate model adequacy. Adjustment for confounding variables is important in clinical epidemiology to minimise bias. To achieve this, any 'third' predictors thought to be influential are identified and included in the model (Jewell, 2004). The adjusted model provides insight into the relationship between the predictors and the outcome through the model structure.

1.12.1.1 Confounders

In epidemiological analysis, there are three types of effect exerted by the 'third' variable: a) confounding, b) mediation and c) suppression. Confounding does not necessarily imply a causal relationship among the variables as opposed to mediation. The presence of this third variable affects the relationship between a risk factor and an outcome (Kleinbaum and Klein, 2012). Due to this 'confounding' effect, the third variable is more often known as a confounder (Katz, 2011, Szklo and Nieto, 2007).

In confounding, the association between variables has these essential features: a) the confounding variable is causally associated with the outcome, b) the confounding variable is causally or non-causally associated with the exposure and c) the confounding variable is not an intermediate variable in the causal pathway between exposure and outcome (Szklo and Nieto, 2007, Jewell, 2004).

Several assessments identify a particular variable as an important confounder: a) the application of automatic variable selection, b) the comparison of adjusted and unadjusted

effect estimates and c) the combination of statistical association from the data with some background knowledge about the causal network (Hernan et al., 2002). A variable has a considerable effect on the model if the coefficient of at least one of the explanatory variables changes by more than 10% (Hernan et al., 2002) or 20% by the inclusion or removal of a suspected third variable (Hosmer et al., 2011, Hosmer et al., 2013). Conversely, relying only on statistical significance to identify confounding is not recommended but if unavoidable, use a larger p-value than the traditional 0.05 during model building (Szklo and Nieto, 2007).

1.12.1.2 Interaction

In interaction, at least two explanatory variables (predictors) interact with each other and influence the parameter estimation during model building. The product of the two 'interacting' variables is known as the two-way interaction term (Katz, 2011, Hosmer et al., 2011, Hosmer et al., 2013). In interaction, the effect of a covariate in the model may differ depending on the level of the other covariate in the same model (Kleinbaum and Klein, 2012).

Interaction can be described as instances in which potential intervention on secondary exposure is in view. It is often used interchangeably with effect modification (Hosmer et al., 2011, Hosmer et al., 2013, Katz, 2011). In contrast to interaction, effect modification is merely conditioning of a secondary variable (VanderWeele, 2009, Knol and VanderWeele, 2012).

Unlike a suspected confounder variable, the interaction term is tested for its statistical significance before keeping or removing it from the model using the likelihood ratio (LR) test. A variable in a model that has a p-value of less than 0.05 from a chi-square test suggests the effect of a significant interaction effect on the model (Hosmer et al., 2011, Hosmer et al., 2013).

1.12.2 Modelling the functional form of continuous covariates

In developing a regression model, $x = (x_1, \dots, x_k)$ denotes the vector of predictor variables under consideration and $g(x) = (\beta_{1 \times 1} + \dots + \beta_{k \times k})$ is their linear function.

For a logistic model (binary outcome), the relevant model is $\text{logit Pr}(Y = 1|x) = \beta_0 + g(x)$. In a Cox proportional hazards model (time-to-event outcome), the effect of predictors is modelled through the hazard function $h(t|x) = h_0(t) \exp[g(x)]$ (Sauerbrei et al., 2007).

For generalised linear regression models, if the covariates or predictors are continuous, they must fulfil the assumption of linearity in the logit (Hosmer et al., 2013) or log hazard (Hosmer et al., 2011).

1.12.2.1 The functional forms of continuous covariates

The shape of the relation between the numerical or continuous predictor variables and the outcome in the model should be correctly specified.

One of the many ways to represent the effect of continuous predictor variables on the outcome is to categorise the continuous predictor. It is grouped into about four or five groups at 'suitable' cut-off points. The products of this approach—the corresponding dummy variables—are used as a predictor in the model. Unfortunately, the use of cut-off points is problematic because: a) the resulting step function is a poor approximation of the true relationship, b) cut-off points almost always fit the data much less well than a suitable continuous function, c) it is hazardous because effect estimates are biased, d) p-values are much too small and unlikely to be reproducible (Royston et al., 2006, Sauerbrei et al., 2007), e) the number of cut-off points and where they are placed is arbitrary and f) results are not necessarily robust (Royston et al., 2006, Wong et al., 2011).

The use of splines is the next alternative. To use splines, a continuous predictor variable can be fitted by generalised additive models (Hastie and Tibshirani, 1990, Sauerbrei et al., 2007, Wong et al., 2011). The limitations when using spline-based estimation include: a) susceptibility to erroneous fitting of local noise (Sauerbrei et al., 2007) and b) instability, with very ‘wiggly’ fitted function, render an analysis less useful and difficult to reproduce (Rosenberg et al., 2003).

Other options such as quadratic polynomials, or less frequently, cubic polynomials, have been used in the past. The problem with these two methods is that the range of curve shapes afforded by polynomials is limited (Royston and Altman, 1994, Sauerbrei et al., 2007).

1.12.2.2 Fractional polynomials (FP) and other methods

The functional relationship between numerical predictors and the outcome can be non-linear (J-shaped or U-shaped), which more accurately represents the risk than a straight-line linear relationship does (Wong et al., 2011). In addition, modelling numerical predictors correctly to represent non-linearity in a multivariable model will improve model fitting (Hosmer et al., 2011, Hosmer et al., 2013, Wong et al., 2011).

Fractional polynomial (FP) function and various types of spline are two of the most commonly used methods to represent the non-linear relationship between numerical predictors and the outcome variable (Sauerbrei et al., 2007). The spline method has not gained wide acceptance (Sauerbrei et al., 2007, Wong et al., 2011) because spline modelling, though extremely flexible, generates fitted curves with uninterpretable ‘wiggles’ hence is problematic to many readers (Sauerbrei et al., 2007, Royston and Sauerbrei, 2005).

In FP, the transformation is applied to the numerical covariates. The power for the transformation function of x^p , with p coming from a set, $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. In the case of $\beta_0 + \beta_1 x^p$ with p in S , the model is known as FP model of degree 1, or FP1. FP1

can be extended to degree 2 or FP2 (generalisations of quadratics) (Royston and Altman, 1994). In multivariable modelling, use of degree 3 or FP3 or higher order functions is rarely sensible, as instability is greatly increased (Royston and Sauerbrei, 2005).

1.12.2.3 Software to model the functional form of the continuous covariates

In Stata (StataCorp., 2010), the multivariable FP (MFP) function can produce various functional forms of continuous (numerical) covariates to best represent the relationship between a predictor variable and an outcome variable. The advantage of using MFP is that in both univariable and multivariable model building, MFP will fit data using a simple method with interpretable and transportable results (Royston and Sauerbrei, 2005, Sauerbrei et al., 2007).

In MFP, a 'linear' model for a numerical predictor x assumes a risk score or linear predictor of the form. Similarly, a non-linear model denotes non-linearity in x in the risk score. For example, in Cox regression models of the form, $h(t|x) = h_0(t)\exp \beta(x)$ and $h(t|x) = h_0(t) \exp(\beta\sqrt{x})$ respectively represent the linear and non-linear form in x (Sauerbrei et al., 2007).

MFP uses the closed test procedure (Royston et al., 2006, Royston and Sauerbrei, 2013, Sauerbrei et al., 2007, Wong et al., 2011). During the procedure, these algorithms take place:

1. Determine the best-fitting second-degree polynomial by choosing power transformations from the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, where 0 denotes the log-transformation.
2. The best-fitting FP2 is then compared against the null model using a deviance difference test with 4 degrees of freedom to determine whether the continuous covariate should be included in the model.

3. If the first test is statistically significant, a second deviance difference test with 3 degrees of freedom is applied to compare the best-fitting FP2 against the linear model.
4. If the second test is significant, a final deviance difference test with 2 degrees of freedom is used to compare the best-fitting FP2 with the best-fitting FP1.
5. If the final test is significant, FP2 is included; otherwise, FP1 is chosen.

1.12.3 Variable selection

Subject matter knowledge should guide model building whenever possible (Sauerbrei et al., 2007). Throughout the process of model building, variables that are extraneous, redundant, have a lot of missing data, or intervene between the risk factor and the outcome should be excluded (Katz, 2011). A predictor that is not significant statistically can remain in the model if it is important in practice and if it confounds the effect of the other predictors. On the other hand, predictors that are systematically biased can be excluded (Sauerbrei et al., 2007).

In model building, predictor variables are added into the model or are removed from it. This addition and removal of a variable changes the regression parameters, and the newly developed model and the previous one must be compared to indicate the magnitude of change due to its addition or removal; all is done using logic, theory, prior research, hypothesis testing and model fit comparison (Singer and Willet, 2003). The degree of change can be checked using the LR test or the Wald test. Between the two, the LR test is preferred; unfortunately, there has been no statistical theory that provides clear evidence to support this (Long and Freese, 2006).

1.12.4 Goals of model building

In epidemiology, regression methods have become an integral component for modelling the relationship of risk factors and the disease or outcome of interest. Generalised linear models for example can describe the relationship between a response variable and one or more explanatory variables (Hosmer et al., 2013, Hosmer et al., 2011)—multifactorial conditions or risk factors (Katz, 2011).

One of the goals of disease modelling is finding the best-fitting and most parsimonious, clinically interpretable model which describes the relationship between the response variable and explanatory variable (Hosmer et al., 2013, Hosmer et al., 2011). When the aim of model building is to explore important predictors or independent variables, it is recommended to approach model building by: a) data collection and preparation, b) reduction of explanatory or predictor variables, c) model refinement and selection and d) model validation (Neter, 1996). This approach requires a combination of substantive theory, research questions and statistical evidence (Singer and Willet, 2003).

Methods to reduce the number of predictors include: a) manual selection and b) automatic selection. Automatic selection such as stepwise procedures perform better than literature-based assessment (Sauerbrei et al., 2007) unless the number of predictors or independent variables is large. With a large number of predictors, a parsimonious model is preferred. A parsimony model contains a subset of 'important' predictors whose regression coefficients differ from zero. For this task, sequential strategies include forward selection, stepwise selection or backward elimination procedures or all-subset selection with different optimisation criteria. In practice, the criteria for assessing the model after these strategies are Mallows' C_p , the Akaike information criteria (AIC) or the Bayesian information criteria (BIC) (Burnham and Anderson, 2002, Sauerbrei et al., 2007).

1.13 Thesis overview

This thesis focuses on the use of current and established epidemiological and statistical methods to develop a model of the risk factors for stroke, to explore the prognostic factors for stroke fatality, to compare the prognostic effect of the two main stroke subtypes: a) cerebral infarction and b) haemorrhagic stroke and lastly to model the longitudinal functional health status of stroke patients over 3 months. The information provided by this project would be useful for improving the overall understanding of the natural history of stroke, particularly in stroke prevention and control, the management of acute stroke in the hospital and finally, in post-stroke assessment, specifically in Kelantan and generally in Malaysia, where data on stroke are very scarce.

First, I investigated the risk factors for stroke using a case–control study design. I included individual- and area-level risk factors (Paper I). Using binary logistic regression and in-hospital case and control patients, I developed an epidemiological risk factor model to predict the risk for stroke. The independent risk factors for stroke in Kelantan are: a) age, b) sex, c) race, d) population density and e) average number of household members. The model was complicated by the interaction of the variables age and sex and by the non-linear relationship between age and the odds for stroke.

Next, using similar hospital-based stroke data, I extracted data from a cohort of stroke patients admitted to HUSM. Using the time-to-event data, I developed a prognostic model based on Cox proportional hazard regression to predict in-hospital stroke fatality. The independent prognostic factors of in-hospital stroke fatality were age and the GCS score—a well-known scale that quantifies the severity of a coma (Paper II).

Based on the information from Paper II, the focus in the subsequent study was to compare the clinical presentation (on admission), fatalities and prognostic effect of stroke

fatality between two main stroke subtypes: a) haemorrhagic stroke and b) cerebral infarction) (Paper III).

Lastly, to understand the impact of stroke on functional status after acute stroke (post-stroke), I recruited 98 stroke patients admitted to HUSM and HRPZ. I collected their baseline information and scored each patient using the Barthel Index—a well-known functional status questionnaire in stroke—at discharge then at 1 and 3 months after discharge (Paper IV).

1.13.1 Draft Paper 1

Individual and area-level risk factors for stroke in Kelantan, Malaysia: A hospital-based case–control study by Kamarul Imran Musa, Peter J Diggle, and Thomas J Keegan is a study in which we investigated the risk factors for stroke using in-hospital patient data from two major hospitals in Kelantan: a) HUSM and b) HRPZ. Our general objective was to explore the relationship of the individual- and area-level variables (risk factors) with the outcome (stroke or no-stroke) using a case–control study design. We defined the cases (stroke) and the controls (no-stroke) using the ICD-10. We requested the hospital data containing the individual-level data for the cases and controls. The Department of Statistics, Malaysia, provided the area-level data. We performed binary logistic regression to model the risk factors and during the modelling, we used MFP to correctly specify the non-linear relationship between the continuous variables and the outcome. We plotted a variogram to assess the presence of significant spatial correlation (whether the geographical locations of stroke patients are related with the risk for stroke). In this paper, we analysed 3118 patients (1369 cases and 1749 controls) and found that the variables age, population density, average number of household members and race and a 2-way interaction term between transformed age and sex were significant in the final model. The final model also passed the goodness of fit test. There was

no strong evidence to suggest the presence of spatial correlation, hence justifying the use of traditional binary logistic regression. We concluded that a model with age, race, population density and an interaction term between age and sex predicts stroke but that locations do not.

Dr Thomas Keegan and I designed the study. I requested data from the hospitals and the Department of Statistics and cleaned and analysed the data. Professor Peter Diggle critically reviewed the analysis and improved the results. I drafted the manuscript and Dr Thomas Keegan performed the review and final editing of the overall manuscript.

1.13.2 Draft Paper 2

Glasgow Coma Scale and age as independent prognostic factors for in-hospital stroke fatality by Kamarul Imran Musa, Peter J Diggle, and Thomas J Keegan is a study with the following objectives: a) to compare the on-admission variables between stroke patients who were alive and dead at discharge, and b) to identify the prognostic factors for stroke fatality. In this study, we recruited 226 consecutive stroke patients admitted to HUSM. Stroke patients were eligible if they had been discharged with the ICD-10 Chapter IX Blocks I60–I69 (cerebrovascular diseases). We used an independent *t*-test and the Pearson chi-square test to compare the on-admission variables between stroke patients who were alive or dead on discharge. We performed Cox proportional hazard regression analysis to identify the prognostic factors for stroke fatality. In this study, 57.1% (129/226) of stroke patients were female and 42.9% (97/226) were male. On admission, the overall mean age and GCS score were 60.8 years and 12.4, respectively. Overall, 65.4% (148/226) and 32.7% (74/226) had high blood pressure and diabetes mellitus, respectively. Two models were considered: a) age and GCS score, and b) age and stroke subtype. Model (a) was considered the best prognostic model for stroke fatality because it is more practical in clinical settings.

Dr Thomas Keegan and I designed the study. I requested the patient files and extracted the relevant data from each patient file. I entered, cleaned, analysed and interpreted the data. Professor Peter Diggle critically reviewed the analysis and improved the results. I drafted the manuscript and Dr Thomas Keegan performed the review and final editing of the overall manuscript.

1.13.3 Draft Paper 3

Comparing cerebral infarction and haemorrhagic stroke in Asia: case–fatality and prognostic effect for in-hospital survival by Kamarul Imran Musa, Juhara Haron, Peter J Diggle, and Thomas J Keegan is a study in which our aims were to: a) describe the on-admission variables for patients with cerebral infarction and haemorrhagic stroke, and b) compare stroke fatalities (3 days, 7 days and 14 days) and the prognostic effect for case–fatality between cerebral infarction and haemorrhagic stroke. We collected data from two groups of consecutive stroke patients admitted to HUSM: a) cerebral infarction (n = 150) and b) haemorrhagic stroke (n = 142). We extracted patient variables from the hospital medical records and patient clinical folders and reviewed all CT scan images and reports. The overall stroke fatality and fatalities at 3 days, 7 days and 14 days were described. The survival probability and prognostic effect were estimated using Kaplan–Meier survival estimates and Cox proportional hazard regression, respectively. We found that patients in the haemorrhagic stroke group were younger, had poorer GCS scores and had a higher mean diastolic blood pressure than those in the cerebral infarction group. The median survival time for patients with cerebral infarction was 28 days; that for patients with haemorrhagic stroke was 14 days. Haemorrhagic stroke patients had more than twice the risk for stroke fatality than those in the cerebral infarction group. We concluded that haemorrhagic stroke patients had less favourable clinical presentation on admission, higher 3-, 7- and 14-day fatalities and more than twice the risk for stroke fatality than those with cerebral infarction while in hospital.

Dr Thomas Keegan and I designed the study. I requested the patient files and extracted the relevant data from each patient file. Dr Juhara Haron verified the case note diagnosis with CT scan images archived at HUSM or the CT reports of each patient. I entered, cleaned, analysed, interpreted the data and drafted the manuscript. Professor Peter Diggle critically reviewed the analysis and improved the results and Dr Juhara Haron reviewed the methods. Dr Thomas Keegan performed the review and final editing of the overall manuscript.

1.13.4 Draft Paper 4

Barthel Index scores over three months are related to age and stroke subtype in Asian stroke patients by Kamarul Imran Musa, Peter J Diggle, Thomas J Keegan is a study with the following objectives: a) describing the change in the Barthel Index at baseline (discharge) to 1 month and then 3 months post-baseline, and b) estimating the relationship of age, sex and stroke subtype with the change in Barthel Index scores over the 3-month period. We recruited 98 stroke patients from HRPZ and HUSM when they were admitted for stroke. I scored the Barthel Index for all patients at baseline (discharge) and then at 1 and 3 months after discharge. We used the random intercept model to account for the individual subject random effect. In this study, the mean age of the patients was 60.7 years, 65.3% (64/98) were female and 73.7% (70/95) had ischaemic stroke. The unadjusted Barthel Index for all patients increased from baseline (mean score = 35.1) to 1 month (mean score = 64.4) and 3 months (mean score = 78.0). Between discharge and 3 months, 13 patients had died and the distribution of scores between patients who died and who were alive during the study period were statistically different (p-value from Kolmogorov–Smirnov test = 0.048). Using a linear mixed effect (*lme*) model, the significant predictors for the change in Barthel Index over the 3-month period were measurement occasions, age and stroke type. Over the 3-month period, the adjusted mean score—from *lme*—for the Barthel Index increased from 35.1 to 68.7. We reported that with increasing age, the Barthel Index score decreased and that haemorrhagic stroke patients had

lower Barthel Index scores than ischaemic stroke patients. We concluded that, overall, the Barthel Index scores increased from discharge to 3 months and that measurement occasion, age, and stroke type, but not sex, were related with the change in Barthel Index score.

Dr Thomas Keegan, Professor Peter Diggle and I designed the study. I interviewed all patients at baseline (discharge) and at 1 and 3 months after the baseline interview. I entered and cleaned the data and analysed the results. Professor Peter Diggle critically reviewed the analysis, helped in the interpretation and improved the results. I drafted the manuscript and Dr Thomas Keegan performed the review and final editing of the overall manuscript.

Chapter 2 Individual and Area-level Risk Factors for Stroke in Kelantan, Malaysia: A Hospital-based Case-control Study

Kamarul Imran Musa ^{1,2}, Peter J Diggle ¹, Thomas J Keegan ¹

¹ CHICAS, Lancaster Medical School, Lancaster University, Bailrigg, Lancaster, United Kingdom LA14YW

² Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian Kelantan, Malaysia

Corresponding author:

Kamarul Imran Musa

Lancaster Medical School,

Lancaster

United Kingdom LA14YW

Email: k.musa@lancaster.ac.uk

Telephone: +44 1524594539

2.1 Abstract

Aims To identify the risk factors for stroke using the individual-level and the area-level variables among in-hospital patients in Kelantan, Malaysia.

Methods A case-control study using in-hospital patients was undertaken. The individual-level data were retrieved from the hospital records. The area-level data were derived from the 2010 Malaysia nationwide census. The cases and controls were defined based on the International Classification of Diseases 10th Revision (ICD-10). Logistic regression was used to model the risk factors and multivariable fractional polynomials (MFP) to capture non-linear associations of numerical variables with stroke risk. The sample variogram and envelopes were simulated after fitting the model and plotted to explore the residual spatial correlation at the level of sub-sub districts (n=288).

Results A total of 3,118 patients (1,369 cases and 1,749 controls) were analysed. At univariable logistic regression analysis, variables age, sex, race, marital status and population density were significantly associated with stroke admission. After fitting multivariable logistic regression, variables age, population density, average household and race and two-way interaction between age and sex were significantly related with stroke. The area under the Receiver Operating Characteristic (ROC) curve was 0.733 and the final model passed Hosmer-Lemeshow and Pearson goodness of fit tests. The sample variogram showed no evidence of residual spatial autocorrelation.

Conclusion Age, sex, race, population density and average household are independent risk factors for stroke in Kelantan, Malaysia. A quantitative description of the odds for stroke is complicated by the non-linearity and by the interaction of age and sex.

Keywords: Stroke, Risk factors, Logistic regression, Fractional polynomials

2.2 Introduction

Stroke is defined as a clinical syndrome of permanent brain dysfunction having an improving or worsening temporal profile commonly due to a brain infarction, brain ischaemia or brain haemorrhage (Tegos et al., 2000). It is 'a clinical syndrome characterized by rapidly developing clinical symptoms and/or signs of focal, and at times global, loss of cerebral function, with symptoms lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin' (Hatano, 1976). The World Health Organization (WHO) reported that in 2008 stroke was the second commonest cause of death worldwide. (Lopez et al., 2006, Sarti et al., 2000).

Knowing the risk factors for stroke is important for its control and prevention. Customarily, the risk factors for stroke are categorized into modifiable risk factors, such as biochemical profiles or smoking status, and the non-modifiable risk factors such as sex and age (Adams, 2007, Lindley, 2008). Alternatively, risk factors can also be categorized into traditional (such as smoking status or blood pressure) and novel risk factors (such as geographical location) (Romero et al., 2008).

As research in stroke progresses, new domains of social and environmental risk factors have been studied in addition to the big three risk factors for stroke - hypertension, diabetes mellitus and cigarette smoking (Adams, 2007). These new domains include; environmental risk factors such as geographical locations (Havulinna et al., 2008, Pedigo et al., 2011) and ethnic background (Carson et al., 2012, Pullicino et al., 2009). Geographical locations and improved computation enable spatial analysis and these has proven useful to track system progress of stroke management and to act as a tool in stroke surveillance (Gropen et al., 2009). In the United States of America, studies such as the REasons for Geographic And Racial Differences

in Stroke study (REGARDS) allowed for the creation of a national cohort to address geographic and ethnic differences in stroke (Howard et al., 2006).

The majority of stroke studies has taken place in high-income countries (O'Donnell et al., 2010a). We used Pubmed Advanced Search Builder (www.pubmed.com) and set the titles as 'Malaysia' and 'Stroke'. The search yielded fifteen articles, with the latest publication published in May 2016, in Neurology Research. Only one was a longitudinal study and it was about a small population in the Northwest Malaysia. None was a case-control study. Eleven articles came from studies based on single-centre study (Mohd Zulkifly et al., 2016, Wan-Arfah et al., 2015, Sudirman et al., 2015, Ali et al., 2015, Akhavan Hejazi et al., 2015, Sahathevan et al., 2014, Khoo et al., 2013, Nor Azlin et al., 2012, Rameezan and Zaliha, 2005, Hamidon and Raymond, 2003, Jaya et al., 2002b), three articles were from multi-centre studies (Neelamegam et al., 2013, Aziz et al., 2015, Aziz et al., 2016) and one article is a review article (Loo and Gan, 2012).

Motivated by the need to improve the understanding of risk factors for stroke in Malaysia, and with the availability of the latest census tract from 2010 nationwide census, we have carried out a hospital-based case-control study. To our knowledge, the study is the first of its kind in Malaysia to investigate the incidence of stroke, traditional risk factors using census tract data. The current study would be a valuable resource for healthcare providers and policy makers in the control and prevention programme for stroke in Malaysia specifically and in Asia generally. We examined the relationship between two groups of risk factors variables – the traditional individual-level variables and the census area-level variables – and risk of hospital admission for stroke in Malaysia.

2.3 Methods

2.3.1 Study design and setting

We performed a hospital-based case-control study in the state of Kelantan, in the north-east of Malaysia. Kelantan, is bounded by Thailand in the north, the South China Sea in the northeast, the state of Terengganu in the east, the state of Pahang in the south and the state of Perak in the west. It covers 15,099 km square, and in 2011 had a population of 1.6 million.

Our study was set in the two public hospitals in Kelantan that offer a neurology service. These were a Ministry of Health hospital, Hospital Raja Perempuan Zainab II (HRPZ) and a public university teaching hospital, Hospital Universiti Sains Malaysia (HUSM).

In this study, the individual-level data were obtained from the hospital records and the area-level data were provided by the Department of Statistics, Malaysia (in shapefile format). There are 288 administrative districts sub-sub districts (known as 'mukim kecil') in Kelantan, Malaysia, and we received area-level data for each. The area-level data consisted of data on: surface area (square km²), average household size and population size for each sub-sub district.

2.3.2 Selection of cases and controls

Cases and controls were selected from in-patients at both study hospitals.

A patient was eligible to be recruited as a case for the study if they met all the following criteria; at least 18 years old, a Kelantan resident, a Malaysian citizen and was treated in either Hospital Universiti Sains Malaysia (HUSM) or Hospital Raja Perempuan Zainab II (HRPZ) as in-hospital patients within the study period, June 2010 - June 2011.

Cases were patients who had been discharged with the diagnosis of cerebrovascular disease (ICD-10) Blocks I60 to I69 (Cerebrovascular diseases) in the Chapter IX (Diseases of the Circulatory Systems). The discharge diagnosis had to have been made by either the neuro-medical or the neuro-surgical team at HUSM and HRPZ, Kelantan.

Controls were selected from patients who had been discharged from either one of the study hospitals with the ICD-10 diagnoses of other than cerebrovascular disease Blocks I60-I69. This means we selected patients as controls even though they have circulatory problems such as ischaemic heart diseases (ICD-10 I10-I15), hypertensive diseases (ICD-10 I10-I15), pulmonary heart diseases and diseases of pulmonary circulation (ICD-10 I16-I28) and other circulatory problems that share many of the risk factors with cerebral infarct (Table 2-1). In other case-control study, these type of controls were excluded (O'Donnell et al., 2010a). Our intention is that our study will be able to identify risk factors that are more related to stroke than other cardiovascular diseases. We excluded a patient as a control if the discharge diagnoses was either ICD-10 Chapter XV Pregnancy, Childbirth and the Puerperium (O00-O99) or ICD-10 Chapter XVI Certain Conditions Originating in the Perinatal Period (P00-P96) (Table 2-1).

Case-control studies in which the cases are drawn from a tertiary referral centre (such as ours) can be prone to selection bias as the cases are a selection of all possible stroke cases (in particular they are the cases that have survived to get to hospital). We used hospital controls which raises the possibility of bias being introduced, in that the controls may differ in some way from the population from which the cases were drawn and this itself may be related to exposure. Limitations in resources prevented us from selecting controls from the community, as has been the practice in other studies (O'Donnell et al., 2010a, O'Donnell et al., 2010b, Kwon et al., 2016).

Table 2-1 Diagnostic categories for controls

Diagnostic categories for controls	Conditions
Included ICD-10 categories: All ICD-10 excepted those in the excluded list	
Excluded ICD-10 categories: ICD-10 Chapter IX I60-I69 ICD-10 Chapter XV	Cerebrovascular diseases Pregnancy Childbirth Puerperium
ICD-10 Chapter XVI	Certain Conditions Originating in the Perinatal Period (P00-P96)

2.3.3 Statistical analysis

All data analyses were carried out using Stata version 11.2 (StataCorp., 2010) except for the variogram analysis for which we used R version 3.11. (R Development Core Team, 2013). Summary descriptions were means and standard deviations for numerical variables, frequencies and percentages for categorical variables.

We performed univariable logistic regression followed by multivariable logistic regression using as the outcome variable admission to hospital for stroke, coded as 1 for 'yes' and 0 for 'no'.

We used purposeful manual variable selection whereby independent variables were added one at a time (Hosmer et al., 2013). Briefly, during the selection process, we performed these steps:

- 1) we began our analysis with careful univariable analysis of each independent variable. Through the use of these analyses, we identified, as candidate variables for the first multivariable model, any variable whose univariable test has a p -value less than 0.25 or 0.20, or whose has clinical importance.

- 2) we fitted the multivariable model containing all the covariates identified for inclusion in Step 1. We assessed the importance of each covariate using the p -value of its Wald test or the likelihood ratio test.
- 3) We compared the reduced model against the full model. In particular, we were concerned with any variables whose coefficients has changed markedly
- 4) We checked the assumption that the logit had increased/decreased linearly as a function of the covariate.
- 5) We checked for interaction among the variables in the model.
- 6) We assessed the model fitness.

At each step of adding or removing a variable in the model, the likelihood ratio test was performed and the degree of change in the beta coefficients of the variables was checked for the presence of significant confounding effects.

We checked all the numerical variables to determine whether they should enter our model linearly or transformed using the Multivariable Fractional Polynomials (MFP) function in Stata (Royston and Sauerbrei, 2004, Sauerbrei et al., 2007, Hosmer et al., 2013). Based on MFP, variable age was transformed to two terms of $age1=age/10$ and $age2= (age/10)^2$ (quadratic function).

The interaction between transformed age and sex was significant (p -value <0.05 on two degrees of freedom) and included in the model accordingly. Goodness of fit was assessed using the Hosmer-Lemeshow test, C, Pearson chi-square test (Hosmer et al., 2013) and area under the Receiver Operating Characteristics (ROC) curve (Zelmanovitz et al., 1997).

We used the sample variogram of the residuals to check for spatial autocorrelation between stroke risk and location. A significant spatial autocorrelation would indicate a significant relationship between stroke and location after allowing for the effects of all

independent variables included in the model. Standard statistical regression models often assume independence of the observation and these models are not appropriate for analysing spatially dependent data (Auchincloss et al., 2012). The availability of geographic coordinates in our data enabled the assumption of independence of the observation to be checked.

To perform the variogram analysis, we needed the geographical coordinates of each patient in the study. Unfortunately, neither hospital recorded the address coordinates for patients. Hence, we obtained coordinates (geocoded) addresses by matching each patient's address with the database of villages provided by the Malaysian Centre for Geospatial Data Infrastructure (MaCGDI) at <http://1malaysiamap.mygeoportal.gov.my/> website. This database contains the names of villages in Malaysia with the corresponding geographical coordinates. These geographical coordinates are the latitude and longitude coordinates pointed at the centre of each of the village. Alternatively, we utilized the open-source mapping service providers: a) Wikimapia (<http://wikimapia.org/>) and, b) Google map (<https://www.google.com.my>) to locate the villages that are not available in the MacGDI database. Using ArcGIS, we determined the geographical coordinates of these villages. These village geographical coordinate (the geographical centroid) acted as the coordinates for patients address in this study. To prevent 'stacking' of multiple geographical coordinates of the villages, we used R software to slightly 'jitter' the coordinates.

We calculated the spatial residuals as the difference between the observed number of strokes in a sub-sub district and the expected number of stroke in a sub-sub district (n=286) after fitting the logistic model. We measured the distances between each pair of centroids of the sub-sub districts and calculated the sample variogram up to a maximum distance of 50,000 metres (50km).

In the plot of the sample variogram, the rising curve indicates positive spatial correlation. Conversely, a plot that is confined within the envelopes of variograms after

random permutation of the residuals indicates absence of spatial autocorrelation (Diggle and Ribeiro, 2007).

On completion of model building and model checking, the final estimated logit model for our stroke data is given in the following equation:

$$\begin{aligned}\hat{g}(x) = & -6.575 + 1.501(age1) - 0.086(age2) + 0.130(pop) - 1.332(sex) \\ & + 0.596(race) + 0.179(averhh) + 0.582(age1)(sex) \\ & - 0.058(age2)(sex)\end{aligned}$$

where: $age1$ is $(age/10)$, $age2$ is $(age/10)^2$, pop is population density - every increase in 1,000 population per 1 km square, sex is male vs female (reference), $race$ is Malay vs non-Malay (reference), and $averhh$ is average number of household, 5 or more household vs less than 5 household (reference).

2.3.4 Ethical approval

We received the ethical approvals from the Medical and Research Ethics Committee (MREC), Ministry of Health, Malaysia (NMRR-12-471-12139), the Human Research Ethics Committee USM (HREC), Universiti Sains Malaysia (JEPERM[242.4.(1.4)]) and Lancaster University Ethics Committee.

2.4 Results

We analysed data on 3,118 patients from two hospitals, Hospital Universiti Sains Malaysia (HUSM) and Hospital Raja Perempuan Zainab 11 (HRPZ), Kelantan, Malaysia. Of these, 1,369 were cases and 1,749 were controls. Sixty-four percent of cases and controls came from HRPZ and the rest from HUSM.

Table 2-2 shows the individual and area-level variables (characteristics) of the cases and controls. There were more males (54.8%) than females (45.2%). A majority of patients (75.9%) were married at the time of admission and 94.2% of all patients were Malays. The mean age for all patients was 55.2 years old (SD = 17.9) and the cases were older than the controls (63.1 years vs 49.0 years).

In Table 2-3 we show the final logistic model containing the independent variables and their adjusted odds ratios for stroke. These individual-level variables; age1 and age2 (both transformed from age), race (Malay vs Non-Malay), sex (Male vs Female) and the interaction term for age1 and age2 with sex; and population density and the average household size from area-level variables were significant the final model.

Table 2-2 also shows the area-level variables obtained from the Malaysia 2010 census. The mean population density was 1328.2 people per km square and was higher in cases (1427 people per km square) than in controls (1252 people per km square). Of all, 29.8% of the census-tracts had the average household size of 5 or more.

In Table 2-2, we show the results from the univariable or simple logistic regression. The independent variables were; age, transformed variable age, age1 and age2, sex, race, marital status, population density, average household size and were analysed one at the time. The outcome variable was admission to hospital for stroke. Significant relationships with the outcome of stroke were seen with all independent variables except race and average household size. In the last column of

Table 2-3, we show values for the log likelihood of each of the independent variable. The lowest value for the log likelihood the model with the best fit, so the model with 2 terms (age1 and age2) transformed from age has a better fit (log likelihood = -1843.72) than the model using the untransformed age (linear form) (log likelihood = -1879.28).

In Table 2-3 we show the final logistic model containing the independent variables and their adjusted odds ratios for stroke. These individual-level variables; age1 and age2 (both transformed from age), race (Malay vs Non-Malay), sex (Male vs Female) and the interaction term for age1 and age2 with sex; and population density and the average household size from area-level variables were significant the final model.

Table 2-2 Individual and area-level variables of cases and controls. The outcome was coded as a binary variable either as a case (has stroke) or a control (does not have stroke).

Variables		Cases, n=1,369		Controls, n=1,749		All, n=3,118	
		n	%	n	%	n	%
Sex	Male	723	52.8	987	56.4	1,710	54.8
	Female	646	47.2	762	43.6	1,408	45.2
Marital Status	Married	1,148	83.9	1,225	70.0	2,377	75.9
	Others	221	16.1	524	30.0	98	3.1
Race	Malay	1,300	95.0	1,637	93.6	2,937	94.2
	Other	69	5.0	112	6.4	181	5.8
Household size	≥5	430	31.4	499	28.5	929	29.8
	< 5	939	68.6	1,250	71.5	2,189	70.2
		Mean	SD	Mean	SD	Mean	SD
Age	Years	63	12.9	49	18.8	55	17.9
Population density	Pop/ km ²	1,427	1405.4	1,252	1355.5	1,329	1,380.2

Note: Age: Mean and SD, Pop density: Population density. Pop density and average household are area-level variables

A sensitivity analysis, in which the observed status of the patients was compared to the predicted status using the final model, showed that the final performs acceptably. The results show that the overall rate of correct classification: sensitivity, specific, positive predictive value and negative predictive value were 66.4%, 71.5%, 62.4%, 59.8% and 73.7%, respectively.

Two goodness-of-fit tests, the Pearson and the Hosmer-Lemeshow goodness of fit tests show that our model has good fit. The former test shows no significant difference between the observed and fitted value (chi-square p-value=0.215) and the latter, which compares the observed and estimated frequency in different group of probabilities, was also not significant (Hosmer-Lemeshow test= 6.519, df=8, chi-square p-value=0.591).

Table 2-3 Simple logistic regression results showing the estimated regression coefficients, respective 95% confidence intervals, p-values and the log-likelihood values for each variable.

Independent variable	n	Beta ^a	Lower 95% CI ^b	Upper 95% CI	p-value ^c	LL ^d
Constant only model	3118	-0.24	-0.32	-0.17	<0.001	- 2138.02
Untransformed age	3118	0.05	0.05	0.06	<0.001	- 1879.28
Transformed age	3118					
Age1		1.80	1.47	2.12	<0.001	-
Age2		-0.12	-0.14	-0.09	<0.001	1843.72
Sex	3118	-0.15	-0.29	0.00	0.044	- 2135.99
Race	3118	0.25	-0.06	0.56	0.107	- 2136.70
Marital status	3118	0.80	0.62	0.98	<0.001	- 2096.54
Population density (1,000 per km ²)	3118	0.09	0.04	0.14	<0.001	- 2131.87
Average household size	3118	0.14	-0.02	0.29	0.08	- 2136.50

^{a,c} are the unadjusted parameters and p-values. ^b Confidence interval. ^d are the log-likelihood values. Age1 and Age2 are the transformed variables from age generated by Multivariable Fractional Polynomials (MFP) in Stata. Age1=Age/10, Age2=(Age/2)²

Table 2-4 The final model developed from multiple logistic regression analysis (n=3118) showing the estimated odds ratios, respective 95% confidence intervals and p-values.

Independent variables in the model	OR ^a	Lower 95% CI ^b	Upper 95% CI ^b	p-value
age1	4.49	2.76	7.29	<0.001
age2	0.92	0.88	0.96	<0.001
Malay v. non Malay	1.81	1.30	2.54	0.001
Male v. female	0.26	0.04	1.71	0.162
age1*male	1.79	0.93	3.46	0.084
age2*male	0.94	0.89	1.00	0.044
Population density (1000 per km ²)	1.14	1.07	1.21	<0.001
≥ 5 household size v. < 5	1.20	1.01	1.42	0.043

^a Odds ratio. ^b 95% confidence interval for odds ratio

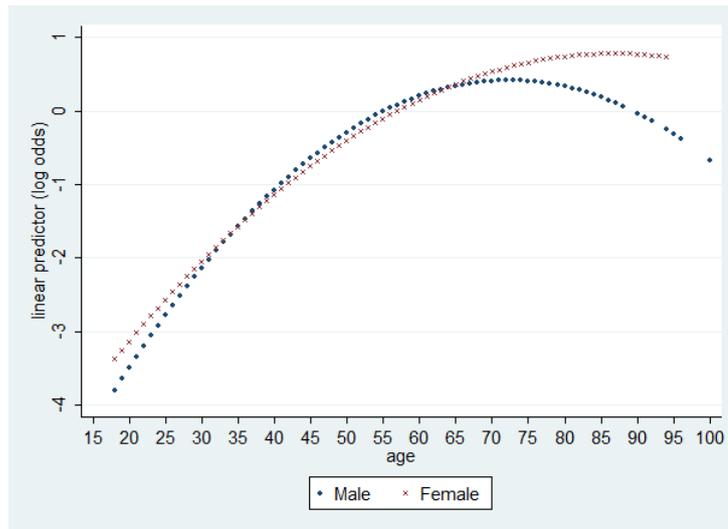


Figure 2-1 The estimated log-odds from the final model (n=3118) are plotted against the variable age. The estimated age effects for males (circles) and females (crosses) shows that the log odds for stroke in men are lower than women at ages up to 35 years, slightly higher between ages 40 years and 60 years, and increasingly lower starting at the age of 65 years.

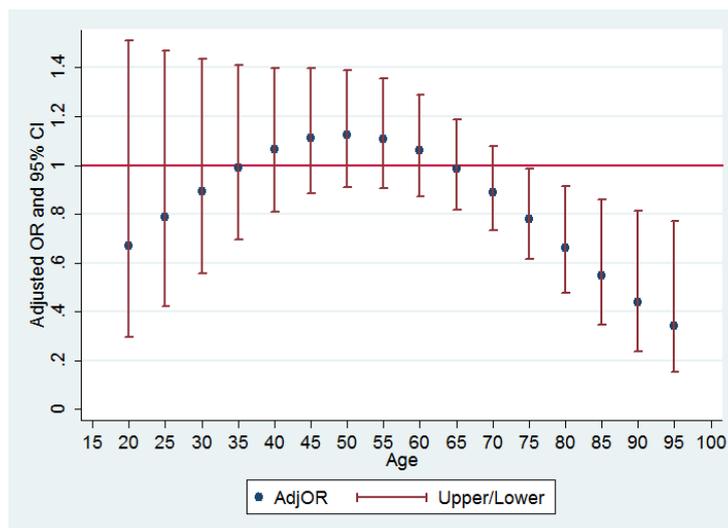


Figure 2-2 The adjusted odds ratios (OR) from the final model (n=3118) are plotted against the variable age. Though the log odds show different age effect for men and women in throughout the age as shown in Figure 2-1, but based the 95% confidence intervals (CI) for the odds ratio, they only reached statistical significance at the age of 75 years and beyond (as 95% CIs are below 1).

Figure 2-1 and Figure 2-2 show the relationship between age (independent variable) and log odds and odds ratios of having stroke by sex, respectively. In Figure 2-1 the curves for the log odds of stroke in men and women intersect around 40 years for men and 65 years for

women. Before the age of 65, the odds of stroke in both men and women increase with age. From the age of 65, the odd of stroke in women increases slightly whereas the odd of stroke in men decreases.

In Figure 2-2, we plotted the odds ratios (OR) from the final model (n=3118) against age (between the age of 20 years to 95 years old with 5-year intervals). In the analysis, women were grouped in the reference group. The plot shows that between the age of 20 years old and 35 years old, the odds of having stroke in men were lower than the odds of stroke in women (odds ratio < 1) albeit non-significant (95% confidence intervals cross 1). However, as all patients aged, the difference in odds decreased. Between 35 years to 65 years of age, the odds for having stroke in men surpassed that of in women (odds ratio > 1) but the 95% confidence intervals still cross 1. The biggest difference in the odds between sexes was around the age 50 years old. From then on, the difference in the odds progressively decreased and around 75 years, the odds for having stroke in men are now being surpassed by the odds of having stroke in women. All the 95% confidence intervals between the age of 75 years to 95 years are below one (1).

In Figure 2-3, the sample variogram to check for spatial autocorrelation taken from the centroids of the sub-sub districts (n=286) is shown. The flatness of the ordinates indicates absence of strong relationship between the spatial residuals and locations of stroke patients measured by the distance a case lives far from one another. This justifies the use of conventional logistic regression – excluding locations (coordinates) of patients – in our model building rather than a more sophisticated geostatistical modelling.

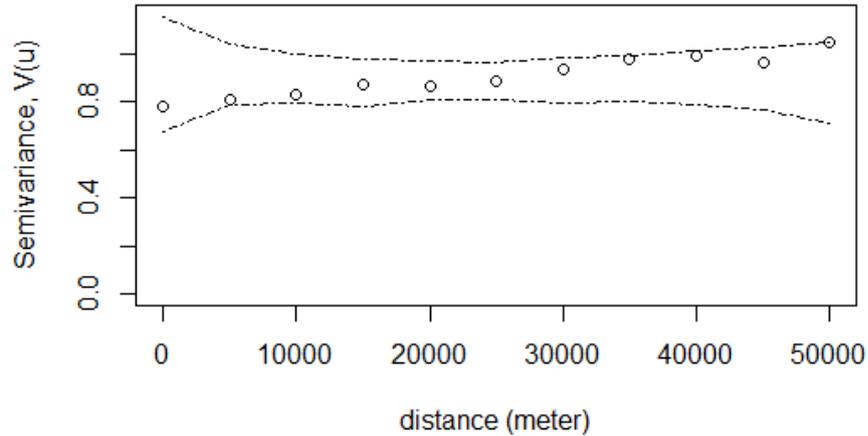


Figure 2-3 Sample variogram and envelopes generated by simulating the semivariance against the distance (in metre). The variances come from the final fitted multiple logistic model. The variogram was based on observed residuals generated by the logistic model and the ordinates were taken from the centroid coordinates of the sub-sub districts ('mukim kecil'), n=286.

2.5 Discussion

In this case-control study, we explored the individual-level and area-level risk factors for stroke using data from in-hospital patients in the state of Kelantan, Malaysia. We found that the important individual-level risk factors for stroke were age, race and sex, and the area-level risk factors for stroke were population density and average household size. There was also no evidence that a patient's location and stroke risk were correlated, hence, use of a 'traditional' logistic regression was adequate.

It is generally accepted that as people age, the risk of chronic diseases such as stroke, increases (Schaller, 2007). It has been shown that the incidence of stroke doubles every decade in both females and males after the age of 55 years (Rojas et al., 2007). This study confirms such a relationship in stroke. But, it is also important to consider if age should be one of the important risk factors in the model rather than as a nuisance variable (Sanderman et al., 2006, Joffe, 2003). In this study, we assessed the role of age to try to understand its role in the 'aging process' in our populations, hence the requirement to model age as one of the covariates in our model.

We transformed the variable age into two new variables because the odds for stroke with age was not linear. We used the fractional polynomials method for the transformation. The non-linearity in the risk for cardiovascular diseases in numerical variable such as age is consistent with other epidemiological studies (Drefahl et al., 2012, Driver et al., 2008, Modig et al., 2013a). The implications for our analysis of non-linearity are two-fold: a) firstly, the need to transform the variable age to better fit the statistical model, and b) the relationship between age and risk stroke is not a straight line so is best presented in a graph.

In addition to nonlinearity, our model was also complicated by the interaction between age and sex. We showed the odds for stroke in male and female patients following two different trajectories. In males, odds of stroke increased in a curvilinear fashion with age and plateaued around age 65. And after age 75, the odds for stroke in men decreased. In females, however, the odds of stroke increased and plateaued only at the age 80 years and with no subsequent decrease. This result suggests that women were protected from stroke compared to men until certain age and this is consistent with studies in the US (women were protected until the age of 80 years old) (Sealy-Jefferson et al., 2012) including results from the Framingham Heart Study (Petrea et al., 2009). Unfortunately, the reason for the protective effect is still not known (Sealy-Jefferson et al., 2012) but social and medical factors could be the contributing factors (Petrea et al., 2009). We acknowledge that in our study, it is possible that the age result may have been driven by selection bias. This bias could have resulted from differences in the referral pattern by the sex of controls, such that at younger ages male controls may have been younger than male cases because the nature of the conditions for which they were in hospital.

In general, there is no consensus in the literature over whether either men or women have a higher risk for cardiovascular disease. It has been postulated that as a result of various biological, environmental and social factors, men have a higher risk for stroke than do women,

possibly in the range from 30-40% (Roger et al., 2012). The National Health and Nutrition Examination Surveys (NHANES) from 1999-2004, has reported that women were more than twice as likely to have had a stroke than men and earlier; from the age of 45 to 54 years (Towfighi et al., 2011, Towfighi et al., 2007). Some age-matched studies show that women had higher stroke incidence (Appelros et al., 2009). Our study suggests the odds for stroke were higher in men than in women between the age range of 35 and 65. Outside this range, our data suggests the odds for stroke were higher in women. Our model shows that after the age of 65, the risk for stroke in women levelled off, though still at a rate consistently higher than in men.

The age when the risk for stroke is higher in females differs between studies. For example, we reported that women started to have higher risk for stroke than in males at the age of 65. In Argentina, the risk for stroke in females increased higher than males at the age of at least 80 years old (Rojas et al., 2007).

We also found that ethnicity was significantly associated with risk of admission for stroke. Malay patients had higher odds for stroke admission compared to non-Malay patients. Ethnicity has been shown to play a role in the development of cardiovascular disease, for example, a study in Singapore has shown different disease rates between members of Malay, Chinese and Indian ethnic groups (Lee et al., 2001). More generally, Asian ethnic groups have higher risk for stroke compared to white Caucasian men (Eastern Stroke Coronary Heart Disease Collaborative Research Group., 1998) and in America, African Americans have a higher risk of ischaemic stroke compared to other ethnic groups (Ohira et al., 2006). The difference in risk could be related to different health status and lifestyles between ethnicities, for example in Singapore, obesity rates in Malays were higher than in other ethnic groups (Hong et al., 2004).

On possible reason, why people of Malay ethnicity may be at higher risk of overweight and cardiovascular disease is in adhering to different diet. It is known that Malay people consume less daily fruits than the people of Chinese or South Asian ethnicity, another risk factor for cardiovascular disease (Yen and Tan, 2012). Further studies looking at ethnic disparities will help in the planning of practical preventive strategies (Hong et al., 2004).

In this study, the population density and average size of household are significantly related to the risk of hospital admission for stroke. An increase in population density was associated with an increase in odds of stroke – a 14% increase in odds for every additional 1,000 populations per km². Areas in which the average household contained more than 5 people had a 20% higher odd of stroke compared to smaller households. Our results perhaps suggest that these population density and average household size are proxies for the degree of urbanization, crowdedness, air pollution and the greater pressure of life – all of which contribute to increased stroke prevalence (Lin et al., 2007). In developing countries, urbanization has been shown to improve socio-economic but at the same time, increases the prevalence of cardiovascular risk factors (Nakibuuka et al., 2015).

We avoided categorizing numerical variables to avoid losing statistical information and power (Royston et al., 2006, Royston and Altman, 2010) and instead, used fractional polynomial method to model them. Fractional polynomial is a robust method which can improve the model fit (Wong et al., 2011). One of the advantages of using STATA software is that it has the most functions for performing fractional polynomials in comparison to other software (Sauerbrei et al., 2007, Royston and Sauerbrei, 2013, Royston and Sauerbrei, 2004). Using a parametric Multivariable Fractional Polynomials (MFP) function in STATA, we show that the odds for stroke were not linear in both male and female at different age. MFP also allowed us to show graphically the interaction between the variables age and sex.

Studies have recently reported that location of patients is a risk factor for stroke (Hunt et al., 2014, Lee et al., 2014). The relationship between a location and an event, a phenomenon known as spatial correlation, can be explored using a variogram (Diggle and Ribeiro, 2007) or alternatively, can be treated as a covariate in a more complex generalized linear geostatistical model (Stanton and Diggle, 2013). In our study, the absence of spatial autocorrelation – geographical relationship between stroke patients – means that logistic regression was adequate. In future studies, we would recommend using individual-level coordinates in stroke spatial studies rather than the coordinates of the centroid of an area. This will increase the power of study and minimize the risk of ecological bias.

The strength of our study includes the use of multivariable fractional polynomials (MFP). MFP allowed us to: a) assess the non-linear relation between age and odds for stroke and b) developed an improved logistic model. This finding is consistent with other cardiovascular studies (Drefahl et al., 2012, Modig et al., 2013a, Modig et al., 2013b). We did not categorize age, as doing so would reduce statistical information and power (Wong et al., 2011, Sauerbrei et al., 2007, Royston and Sauerbrei, 2013). Next, we used the nationwide census data, which is conducted every 10 years in Malaysia. Thirdly, we used semi-variogram to justify the use of 'conventional' logistic regression.

The use of in-hospital patients limits the generalizability of the findings. Our findings are on the population of stroke patients who reach hospital. It is likely that there are many patients with less severe stroke who may not attend the emergency departments and many stroke victims will die before reaching hospital. The introduction of a stroke register in Malaysia would improve population level studies of the outcome and risk factors for stroke in the country. Secondly, the use of the disease classification (ICD-10) without a secondary clinical review from the neurologist may introduce selection bias. The recommended case ascertainment would be the verification of the ICD-10 diagnosis by a neurosurgeon or

neurologist, using the CT scan or MRI imaging taken during hospital admission, and without this inaccuracies in diagnosis and coding may occur. Thirdly, two biases could present in our study. The selection bias which was mentioned earlier occurred especially in the retrospective case-control study such as ours because by design it requires both the cases and the controls are representative of the same population (Geneletti et al., 2009). The use of only hospital-based controls would introduce hospitalization bias, also known as Berkson's bias (Geneletti et al., 2009).

In this study, we were aware of the limitation caused by the hospitalized controls and understood that selecting controls in case-control studies tends to be more problematic. The controls would satisfy two requirements; a) within the constraints of any matching criteria, their exposure to risk factors and confounders should be representative of that in the population "at risk" of becoming cases, and b) also, the exposures of controls should be measurable with similar accuracy to those of the cases (Coggon et al., 2003, Bandera et al., 2013).

Efforts have been made to reduce bias that may result from control selection, however it is possible that our control selection methods may introduce bias. Hospital control may be drawn from a population that differs from the cases in terms of their likelihood of exposure. Both our cases and controls were drawn from hospitals. People admitted to hospital for a reason other than stroke may still be likely to be exposed to inherent risk factors for stroke, such as age, or modifiable ones such as high blood pressure. While we were restricted to using hospital controls we did take care to use controls from a variety of wards, and with a variety of diagnoses because exposure measurement using patients a range of control diagnosis rather than single group is a better strategy (Coggon et al., 2003). The effect of the bias may vary: if cases have a higher frequency of the risk factor of interest than do the controls, then a more positive association may emerge. On the other hand, if the prevalence of risk

factors is higher among the controls, a protective association may occur (Grimes and Schulz, 2005). It is possible that the probability of CVD-risk factors may be higher in hospital controls compared similar community controls, which could bias risk estimates towards the null. But a study in China comparing the hospital (out-patients) controls and population controls have found that the hospital out-patient controls performed only little different from population controls for most exposures (Li et al., 2011).

Unfortunately, we do not have exposure variables such as age, body mass index and blood pressure from our data (due to our data came from limited hospital electronic record) that we can compare our controls with the Malaysian general population. Perhaps, because we used hospital (in-patients) controls (whom in our opinion has more co-morbidities), this could have attenuated the relationship between the exposure and the outcome as shown in elsewhere (Neupane et al., 2010). Had we used the population controls, we would have seen a more positive relationship between the risk factors and the outcome (stroke). In the future, when it is possible, we propose taking two controls (from both the hospital and the population) to better assess the relationship between the risk factors and the outcome (Li et al., 2011).

Despite these limitations, in this region, where the studies on stroke are still scanty, and our study has provided additional knowledge on the risk factors for stroke.

2.6 Conclusion

In conclusion, the variables age, race and sex are important predicative individual-level factors for hospital admission with stroke. The interaction between age and sex and the non-linearity relationship between age and the odds for stroke complicates the quantitative description of

the odds for stroke in the final model. The population density and average size of household were the area-level risk factors for stroke. Based on exploratory use of variogram, the spatial correlation between locations and stroke were deemed not important, hence indicative that where people lived in relation to one-another was not a risk factor for admission with stroke.

2.7 Acknowledgements

The authors would like to acknowledge the support from the Director General of Health, Ministry of Health, Malaysia and from the Directors of the Hospital Raja Perempuan Zainab II and Hospital Universiti Sains Malaysia, Kota Bharu Kelantan.

2.8 Funding

KIM is funded by the SLAI and ASHES PhD programme through the Ministry of Education, Malaysia and Universiti Sains Malaysia, Malaysia. The project is funded by the Short Term Grant, Universiti Sains Malaysia (304/PPSP/61312028).

Chapter 3 Glasgow Coma Scale and Age as Independent Prognostic Factors for In-hospital Stroke Fatality

Kamarul Imran Musa¹, Peter J Diggle², Thomas J Keegan²

¹ Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kbg Kerian, Kelantan, Malaysia

² CHICAS, Lancaster Medical School, Lancaster University, LA1 4YM Lancaster, UK

Corresponding author:

Kamarul Imran Musa

Lancaster Medical School,

Lancaster

United Kingdom LA14YW

Email: k.musa@lancaster.ac.uk

Telephone: +44 1524594539

3.1 Abstract

Aims To compare the on-admission variables for stroke patients who were alive or dead at discharge and to identify the prognostic factors for in-hospital stroke fatality.

Methods A total 226 consecutive in-hospital stroke patients admitted to Hospital Universiti Sains Malaysia were recruited between January 2011 and June 2012. All patients matched the International Classification of Disease 10 Chapter IX Blocks I60–I69 (cerebrovascular diseases) codes. The on-admission variables were compared using the independent *t*-test and Pearson chi-square test. Cox proportional hazard regression was used to identify the prognostic factors for stroke fatality.

Results On admission, the variables Glasgow Coma Scale (GCS), haemoglobin level, total white cell count, sex, type of referral and abnormal lipid profile were significantly different between patients who were alive or dead at discharge. Based on univariable Cox proportional hazard regression, only age and GCS were significantly related with stroke fatality. Multivariable Cox proportional hazard regression showed that 2 models predicted in-hospital stroke fatality: model 1 involving age and GCS, and model 2 involving age and stroke subtype. Between the two, model (a) was practical because it is easier and quicker to use in clinical settings. There was no significant 2-way interaction term in the models.

Conclusion Age and GCS are two independent prognostic factors for best predicting in-hospital stroke fatality. More recent clinical assessment methods, however, could be added to improve acute stroke assessment predictive of stroke fatality.

Keywords: Glasgow Coma Scale, Age, Stroke, Case-fatality

3.2 Introduction

The World Health Organisation (WHO) defines stroke as ‘rapidly developed signs or focal (or global) disturbances of cerebral function lasting longer than 24 hours (unless interrupted by death), with no apparent non-vascular cause’ (Hatano, 1976, Feigin et al., 2009, Thorvaldsen et al., 1995).

Stroke case fatality rates vary markedly (28-day fatality ranges from 15% to 57%) between populations (Thorvaldsen et al., 1995), but are higher in most of the Asian countries than in any other region in the world (Kim and Johnston, 2011). In the low- to middle-income countries, the mean total stroke case fatality rate was 35.2% (1980–1989), falling to 23.0% during 1990–1999 and then increasing to 26.6% during 2000–2008 (Feigin et al., 2009).

The predictors for stroke fatality include demographic variables such as sex (Reeves et al., 2008, Appelros et al., 2003, Appelros et al., 2009, Olsen and Andersen, 2010), age (Andersen et al., 2005, Olsen et al., 2011, Fernandes et al., 2012), socioeconomic status (Cox et al., 2006, Eriksson et al., 2013, Lindmark et al., 2013) and education (Lindmark et al., 2013). Clinical variables include co-morbidities such as diabetes (Delbari et al., 2011, O'Donnell et al., 2010b, Fernandes et al., 2012, Basri and Azman Ali, 2003), hypertension (O'Donnell et al., 2010b, Delbari et al., 2011), diet (Scarborough et al., 2011, O'Donnell et al., 2010b) and stroke severity (Saposnik et al., 2008, Andersen et al., 2005, Appelros et al., 2003, Basri and Azman Ali, 2003, Yousuf et al., 2012). Other variables include drug treatment and the availability and experience of stroke teams and physicians (Saposnik et al., 2008), time of admission (weekday versus weekend) (Ogbu et al.), anaemia, location of the infarction (Basri and Azman Ali, 2003) or bleeding (Yousuf et al., 2012) and stroke subtype (van Gijn and Rinkel, 2001, Shigematsu et al., 2013).

To date, there has been no consensus on the most important prognostic factors for stroke fatality, although understanding of the factors is important (Shigematsu et al., 2013). To the best of our knowledge, there have been only 4 published papers indexed by PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) on stroke fatality in Malaysia (Jaya et al., 2002, Basri and Azman Ali, 2003, Yousuf et al., 2012, Sia et al., 2007). All of them were hospital-based studies and two had been published more than 10 years ago. One paper reported that in 2002, the overall in-hospital stroke case fatality rate was 37.3% and that it was more common among male stroke patients (Jaya et al., 2002). The other discussed the stroke case–fatality rate for acute ischaemic stroke in Hospital Kuala Lumpur (situated in Kuala Lumpur, the capital of Malaysia) (Basri and Azman Ali, 2003, OECD/World Health Organization, 2014) but did not explore the prognostic factors for stroke fatality. The remaining two papers described the fatalities among intracerebral haemorrhage (ICH) patients in the East Coast of Malaysia and Kuala Lumpur, respectively, and concluded that the predictors of fatality among Malaysian ICH patients include stroke type, location of haemorrhage and bleeding volume but not sex or age (Yousuf et al., 2012, Sia et al., 2007).

In view of the scarcity of published papers on stroke fatality in Malaysia and with Malaysia being a country in the Asian continent—a region with a serious burden of stroke, high stroke mortality, prevalence and poorly controlled risk factors (Kim, 2014)—more studies on stroke fatality are necessary. This study will add to the body of knowledge on the prognostic factors for stroke fatality and will be useful for improving hospital performance and quality of care for stroke patients (Donabedian, 1988, Saposnik et al., 2008).

In this study, we first aimed to compare the on-admission variables between stroke patients who were alive or dead at discharge using data recorded on admission. Next, we developed multivariable models to identify variables important as prognostic factors for in-hospital stroke fatality.

3.3 Methods

3.3.1 Study design and data source

We obtained data from consecutive stroke patients admitted to Hospital Universiti Sains Malaysia (HUSM), in Kelantan, Malaysia, between 1 January 2011 and 30 June 2012. HUSM is one of the largest tertiary hospitals in Malaysia, serving as one of the three referral centres for stroke in the East Coast of Malaysia. HUSM has a dedicated team of neurologists, neurosurgeons and neuroradiologists who work with all neurology cases. The team also services a neuro-intensive care unit and two neurology wards.

At HUSM, the medical record unit stores databases containing admission and discharge information, including patient basic demographic profiles, date of admission and discharge, survival status on discharge and final diagnosis. HUSM records the discharge diagnosis of each patient in its own electronic hospital registration system (administrative diagnosis) and uses the International Classification of Diseases 10th Revision (ICD-10) for this purpose. The final diagnosis is made by the medical specialists at discharge, then a trained coder inserts the appropriate codes into the registration system.

3.3.2 Case definition

We extracted data from eligible records based on the patient discharge diagnosis (cold pursuit method) (Thorvaldsen et al., 1995). The diagnosis had to match ICD-10 Chapter IX Blocks I60–I69 (cerebrovascular diseases) criteria.

The haemorrhagic stroke patients matched the ICD-10 Block I60, I61 or I62 criteria or any combination thereof. The non-haemorrhagic stroke patients were expected to match the

ICD-10 Block I63 (cerebral infarction) or other type of stroke (Blocks I64–I69) criteria. All the patients in our study had CT scan done for the stroke diagnosis and stroke subtypes.

At HUSM, all patients with suspected stroke will undergo CT scan examination at the emergency department, neuromedical unit or neurosurgical unit as early as possible. Based on the CT scan images and clinical presentations, the treating physicians will confirm the stroke diagnosis and proceed with further management. The radiologist will review the CT scan images usually on the same day, to provide further detailed diagnosis including the stroke subtypes.

The patients had to meet the following eligibility criteria for inclusion in the study: a) stroke as the primary reason for admission, and b) of Malaysian nationality. We excluded patients with: a) shock and hypertensive encephalopathy, as these conditions are not caused directly by cerebrovascular events, and b) discharge diagnosis of transient ischaemic attack (TIA) because the symptoms of TIA resolve within 24 hours (Shigematsu et al., 2013, Devries et al., 2013, Thorvaldsen et al., 1995).

3.3.3 Variables

Through the electronic hospital registration system, the records office administrator at HUSM provided the following demographic and administrative variables in electronic format: a) sex, b) age, c) race, d) marital status, e) date of admission, f) date of discharge, g) discharge diagnosis and h) survival outcome on discharge (dead or alive).

The neurology team gave the final diagnosis of stroke (discharge diagnosis) and the coder at the records office coded it based on the ICD-10 criteria. This type of diagnosis is known as administrative diagnosis.

KIM extracted abstracted the biophysical, medical history and biochemistry data from the case notes. The biophysical data contained the variables a) systolic blood pressure (SBP, mmHg), b) diastolic blood pressure (DBP, mmHg), c) Glasgow Coma Scale (GCS) score and d) capillary blood sugar (mmol/l). The medical history data provided the variables a) history of high blood pressure (yes or no), b) diabetes mellitus (yes or no) and c) abnormal lipid profile (yes or no). The biochemistry data contained the variables a) total white cell (TWC) count, b) haemoglobin (Hb) level (mg/dl), c) platelet count, d) sodium level (mmol/l), e) potassium level (mmol/l) and f) urea level (mmol/l). These variables, especially the biophysical and biochemistry data, are taken as standard diagnostic tests for all patients suspected with acute stroke. Their predictive role in stroke should be assessed to indicate their usability in early stroke care in HUSM setting.

The GCS is used in the acute setting to measure the level of consciousness and is predictive of stroke outcome (Chen et al., 2011, Stroke Unit Trialists, 2007). The GCS score consisted of the best eye, motor and verbal responses, and ranged from a minimum of 3 (worst) to a maximum of 15 (best).

The outcome variable was the time (in days) until an event (death due to stroke during admission). Other outcomes were considered censored observations.

3.3.4 Statistical analysis

We used EpiData Entry (Lauritsen, 2000) for data entry and Stata version 11.2 (StataCorp., 2010) for data cleaning and analyses.

We described the variables on admission using mean (SD) and frequency (%) where appropriate based on the overall patients' data and then based on survival status at discharge (alive or dead). For comparisons, we used the independent *t*-test and Pearson chi-square test.

In assessing survival, the outcome variable was time-to-stroke fatality (event = death) after admission to HUSM. The time was calculated in days (the difference of days between the date of admission and the date of discharge). The event was defined as either failure (death due to stroke, coded as 1) or censored (alive at discharge or death from causes other than stroke, coded as 0).

We used Cox proportional hazard regression—a widely used semi-parametric survival analysis method in medicine—to explore the important prognostic factors for in-hospital stroke case fatality (Hosmer et al., 2013, Hosmer et al., 2011). During model building, we performed the following: a) univariable Cox proportional hazard regression, b) manual selection of variables, c) checking of the functional form of numerical variables, d) checking of interaction between prognostic factors and lastly, e) checking of the assumptions for the hazard proportionality of the chosen model.

Based on univariable analysis (crude), blood pressure, blood count, blood urea, serum electrolytes, capillary blood sugar and marital status were selected for multivariable selection. In multivariable analysis, each candidate variable was added to the model individually. At each step of adding or removing a variable, we performed the likelihood ratio (LR) test, retaining variables with a significance level of less than 2-tailed 0.05. Simultaneously, we examined any change in the coefficient, as a change of 20% or more indicates important confounding effects (Hosmer et al., 2013).

We used fractional polynomials to estimate the relationship between numerical covariates (age and population density) and the outcome (case or control). We did not categorize the numerical covariates because to do so would have reduced the power of analysis and provide less information on the relationship between the covariates and the outcome (Royston and Sauerbrei, 2004, Royston and Sauerbrei, 2005). Fractional polynomials also improve model fit and provide more realistic non-linear relationship if the model (Wong

et al., 2011, Royston and Sauerbrei, 2013, Royston and Sauerbrei, 2005, Royston and Sauerbrei, 2004, Royston and Altman, 1994). We checked the functional form of two numerical variables: a) age and b) GCS score, using fractional polynomial (FP) analysis (Wong et al., 2011, Hosmer et al., 2011, Hosmer et al., 2013, Royston and Altman, 1994, Sauerbrei et al., 2007, Royston et al., 2006). FP determines whether the variables age and GCS score should enter the model in their linear or transformed form. In Stata, the mfp function executes FP analysis. Three models were analysed: 1) the null model, 2) the untransformed model (age and GCS score in linear form) and 3) transformed model (age and GCS score transformed by mfp). In our model, both GCS score and age were best presented in their untransformed (linear) form.

We generated a 2-way interaction term between age and GCS score but the product term was not statistically significant ($p = 0.294$), hence it was excluded from the model.

We tested the proportionality assumption—the estimated hazard does not depend on time—for Cox hazard regression using Schoenfeld residuals. Using these residuals, we performed two tests: a) the ‘global test’ (the overall model test), and b) the ‘detail test’ (test for each numerical covariate in the model). Our chosen model passed both tests.

When performing survival analysis for our data, we assumed our data fulfilled these three assumptions about censoring (for examples to those censored because they were discharged well) : a) independent censoring, b) random censoring and, c) non-informative censoring (Resche-Rigon et al., 2006, Kleinbaum and Klein, 2012). The common analyses of survival data using the Kaplan-Meier method and the Cox regression model will provide bias results when these assumptions are violated. Specifically, in the case of informative censoring, censored observations provide important relationships between censoring and the outcome of interest (the remaining survival time). When informative censoring is suspected, the imputation method for missing observations and sensitivity analysis to estimate the models in

various scenarios can be performed. Further details of imputation and sensitivity analysis are provided in Appendix H. The presence of shared dependencies between the covariates and the outcomes is one of the ways to support the presence of non-informative censoring. The methods to assess it and the results obtained from the assessments are shown in Appendix H.

3.3.5 Ethical approval

We received ethical approval from the Medical and Research Ethics Committee (MREC), Ministry of Health, Malaysia (NMRR-12-471-12139), the USM Human Research Ethics Committee (HREC), Universiti Sains Malaysia (JEPERM [242.4.(1.4)]) and the Lancaster University Ethics Committee.

3.4 Results

In this study, a total 226 consecutive in-hospital stroke cases at HUSM were analysed. Table 3-1 shows that the mean age of the whole cohort was 60.8 years (SD = 14.0). Stroke patients who died when in hospital were older (mean age = 62.2 years, SD = 15.0) compared to those who survived (mean age = 60.4 years, SD = 13.6) but the difference was not statistically significant at $p = 0.05$. The mean GCS score for the whole cohort was 12.4 but patients who were alive at discharge had a significantly higher mean GCS score than those who died (mean GCS score = 13.7 vs. 8.4; $p < 0.001$). A higher GCS score indicates a higher level of consciousness. The difference in the mean SBP and DBP between patients who were alive and dead at discharge was small (SBP, 163.1 mmHg vs. 163.3 mmHg, $p = 0.971$; DBP, 91.6 vs. 92.8, $p = 0.705$).

Table 3-1 also shows that patients who were alive at discharge had a significantly higher mean Hg level than those who were dead at discharge (13.3 g/dl vs. 12.4 g/dl, $p = 0.010$), but there was no difference in mean platelet levels between the two groups ($238.5 \times$

$10^9/l$ vs. $229.5 \times 10^9/l$, $p = 0.474$). The mean TWC count was higher in patients who died during admission ($12.5 \times 10^9/l$ vs. $9.8 \times 10^9/l$, $p < 0.001$). The mean serum potassium, serum sodium, blood urea and capillary blood sugar levels were not significantly different between patients who were dead and alive at discharge (potassium = 3.9 mmol/l vs. 3.8 mmol/l, $p = 0.251$; sodium = 137.5 mmol/l vs. 138.6 mmol/l, $p = 0.178$; urea = 7.6 mmol/l vs. 8.7 mmol/l, $p = 0.201$; capillary blood sugar = 9.6 mmol/l vs. 10.5 mmol/l, $p = 0.274$). The mean length of stay (LOS, in days) between patients who were alive or dead at discharge were not statistically different ($p=0.210$).

Table 3-2 shows that 93.3% (211/226) of patients in our study cohort were Malay, which is consistent with the racial demographics of the state of Kelantan. A nationwide population survey in 2010 showed that 92.7% of the Kelantan population was Malay. Married patients on admission represented 90.7% (205/226) of the cohort, and 57.0% (129/226) of patients were female. Of the total subjects, 38.5% (87/226) had been referred from either tertiary hospitals or district hospitals. Based on self-reporting, 32.7% (74/226) of the cohort had a history of diabetes, 65.5% (148/226) had a history of high blood pressure and 11.9% (27/226) had an abnormal lipid profile.

Of all the categorical variables, types of referral were the most significant variable (p -value < 0.001) associated with the survival status at discharge, followed by abnormal lipid profile (p -value = 0.010) and sex (p -value = 0.014). More referred patients, female patients and patients with normal lipid profile were dead on discharge.

Table 3-1 On-admission numerical variables for all patients (regardless of survival status at discharge) and for patients based on survival status at discharge, n=226

Variable	All patients Mean(SD) n=226	Alive at discharge, Mean(SD) n=173	Dead at discharge, Mean(SD) n=53	p-value ^a
Age (years)	60.8(14.0)	60.4(13.6)	62.2(15.0)	0.412
Glasgow Coma Scale	12.4(3.8)	13.7(2.7)	8.4(4.0)	<0.001
Systolic BP (mmHg)	163.2(34.4)	163.1(32.0)	163.3(41.7)	0.971
Diastolic BP (mmHg)	91.9(20.0)	91.6(18.2)	92.8(25.2)	0.705
Heart rate (per min)	82.7(20.9)	81.9(20.3)	85.4(22.6)	0.293
Haemoglobin (g/dl)	13.1(2.2)	13.3(2.1)	12.4(2.4)	0.010
Platelet count (per mm ³)	236.3(74.6)	238.5(72.7)	229.5(80.5)	0.474
Total white cell count (per microlitre)	10.5(3.9)	9.8(3.5)	12.5(4.2)	<0.001
Sodium (mmol/l)	137.8(5.2)	137.5(3.6)	138.6(8.5)	0.178
Potassium (mmol/l)	3.9(0.6)	3.9(0.5)	3.8(0.7)	0.251
Urea (mmol/l)	7.9(5.5)	7.6(4.8)	8.7(7.2)	0.201
Capillary Blood Sugar (mmol/l)	9.8(4.9)	9.6(5.1)	10.5(3.9)	0.278
LOS ^b	6.5(7.8)	6.1(7.5)	7.8(8.7)	0.210

^a Independent t-test ^b Length of stay (days)

For the risk of dying in hospital from stroke, age and GCS score were the only significant prognostic factors in simple Cox regression analysis, as shown in Table 3-3. The unadjusted risk for in-hospital stroke fatality decreased by 17% with a 1-point increase in the GCS score. A 1-year increase in age increased the unadjusted risk for in-hospital fatality by 2.4% (95% confidence interval [CI]: 1.002, 1.047); a 10-year increase in age increased the crude risk for stroke fatality by 26.7%.

When we examined the relationship between categorical variables and stroke fatality using Cox proportional hazard regression, we found that stroke subtype (haemorrhagic/non-haemorrhagic stroke) was the only significant variable in the simple Cox proportional hazard regression. Others, such as sex, types of referral and abnormal lipid profile that had been shown to be statistically significant (based on Pearson chi-square analyses) in Table 3-2 were

no longer significant. The likely reason for these results to differ is because the Pearson chi-square examines the independence of distribution, i.e. how well the observed distribution of data fits with the distribution expected were the variables to be independent, while the Cox proportional hazard regression method models the incidence or hazard rate per population at-risk per unit time.

Table 3-5 shows the two models that predict in-hospital stroke fatality at HUSM: model 1 (n=225) contains age and GCS score and model 2 (n=226) contains age and stroke subtype. We propose that the best model is model 1 (age and GCS score) based on subjective assessment. We felt that model 1 is more practical in daily clinical practice. Even though model 1 seems to have a better fit (log-likelihood = -205.60, LR chi-square = 40.1, degree of freedom [dof] = 2) than model 2, the quantitative comparison could not be done especially using the likelihood ratio test because of different sample sizes (however, the sample sizes are only different by 1; n=225 vs n=226). Being practical means that the GCS score can be assessed very quickly even by non-clinicians at almost no extra cost. With model 2, computed tomography (CT) scan images to assist clinicians are needed. It is also a poorer fit than model 1 (log-likelihood = -223.33, LR = 12.9, dof = 2). The need for CT scans translates into extra cost, extra time and more training for clinicians. When the model has stroke subtypes, GCS and age together as the covariates, the adjusted hazard ratios become 1.52 (p-value = 0.181), 0.83 (p-value < 0.001) and 1.03 (p-value = 0.006), respectively. Based on the level of significance at 10% for the Wald statistic, the covariate stroke subtype can be dropped from the model. In addition to that, the confounding effect between stroke subtypes and GCS is very likely, hence only one of them should be the covariate at one time.

Based on the 'global test' and the 'detail test' in Stata, our model does not violate the proportional hazard assumption in Cox regression. The p-value for the global test was 0.855 and for the specific test, the p-value for GCS score and age was 0.986 and 0.576, respectively.

Based on the non-significant p-values, we could not reject the null hypothesis that both the hazard for age and GCS score were proportional (slopes of residuals against time were zero).

3.5 Discussion

In this study of 226 consecutive stroke patients at HUSM, sex, history of abnormal lipid profile, referral type, GCS, Hg level and platelet count were significantly different between patients who were alive and dead at discharge. Cox proportional hazard regression showed that age and GCS score are independent prognostic predictors for in-hospital stroke fatality.

Identifying early prognostic factors for stroke fatality is useful for classifying stroke patients based on possible outcome. This will then help clinicians anticipate the risk for fatality among patients with acute stroke. In our study, we have shown that the GCS score and age are independent prognostic predictors for stroke fatality at discharge. Every 1-unit increase in the GCS score corresponded with a decrease in fatality risk of about 18% (hazard ratio [HR] = 0.82).

The reduction could be as small as 12% or as large as 23% (95% CI of adjusted HR from 0.77 to 0.88) when adjusted for age. The GCS is an easy and quick physical assessment of neurological status and can be performed quickly in the emergency setting by clinicians or paramedics. Regarding age, our best model (model 1) showed that with a 1-year increase in age, the risk for in-hospital stroke fatality increased by 3% (95% CI from 1.01 to 1.05) when adjusted for GCS.

Table 3-2 On-admission categorical variables for all patients (regardless of survival status at discharge) and for patients based on survival status at discharge, n=226

Variables	Categories	All patients, n(%)	Alive, n(%)	Dead, n(%)	p-value ^a
Race	Malay	211(93.4)	163(94.2)	48(90.6)	0.350
	Non-Malay	15(6.6)	10(5.8)	5(9.4)	
Sex	Male	97(42.9)	82(47.4)	15(28.3)	0.014
	Female	129(57.1)	91(52.6)	38(71.7)	
Marital status	Married	205(90.7)	158(91.3)	47(88.7)	0.561
	Not-married	21(9.3)	15(8.7)	6(11.3)	
Types of referral	Hospital	87(38.5)	55(31.8)	32(60.4)	<0.001
	GP/Home	139(61.5)	118(68.2)	21(39.6)	
Diabetes mellitus	Has diabetes	74(32.7)	60(34.7)	14(26.4)	0.262
	No diabetes	152(67.3)	113(65.3)	39(73.6)	
High blood pressure	Has high BP	148(65.5)	113(65.3)	35(66)	0.924
	No high BP	78(34.5)	60(34.7)	18(34)	
Abnormal lipid profile	Abnormal	27(12.0)	26(15.0)	1(1.9)	0.010
	Normal lipid	199(88.0)	147(85.0)	52(98.1)	

^a Pearson chi-square p-value

Table 3-3 Numerical variables, crude hazard ratios (HR) with 95% confidence intervals (CI) and the respective p-values. The outcome variable was a time-to-event variable; where the event was status at discharge (dead or alive) and the time was the numbers of days of hospitalization.

Variable	n	Crude HR ^a	Lower 95% CI ^b	Upper 95% CI ^b	p-value
Age (per year)	226	1.02	1.00	1.05	0.034
Glasgow Coma Scale (GCS)	225	0.83	0.78	0.88	<0.001
Systolic BP (mmHg)	225	1.00	1.00	1.01	0.720
Diastolic BP (mmHg)	225	1.00	0.98	1.01	0.678
Haemoglobin (g/dl)	220	0.96	0.85	1.08	0.504
Platelet count (per mm ³)	220	1.00	1.00	1.01	0.466
White cell count (per microlitre)	220	1.06	0.99	1.12	0.088
Sodium (mmol/l)	225	1.02	0.98	1.06	0.373
Potassium (mmol/l)	225	0.93	0.57	1.52	0.774
Blood Urea (mmol/l)	225	1.03	0.99	1.07	0.189
Capillary Blood Sugar (mmol/l)	209	1.03	0.97	1.09	0.329

^a Hazard ratios (HR) were obtained from the simple Cox proportional hazard (PH) regression

^b confidence interval (CI)

Table 3-4 Categorical variables, crude hazard ratios (HR) with 95% confidence intervals (CI) and the respective p-values. The outcome variable was a time-to-event variable; where the event was status at discharge (dead or alive) and the time was the numbers of days of hospitalization

Variables	Category	n	Crude HR ^a	Lower 95% CI ^b	Upper 95% CI ^b	p-value
Sex	Female	129	1			
	Male	97	0.66	0.36	1.21	0.180
Race	Non-Malay	15	1			
	Malay	211	0.48	0.19	1.22	0.122
Marital	Not married	21	1			
	Married	205	1.12	0.48	2.64	0.795
Diabetes	No diabetes	152	1			
	Have diabetes	74	0.64	0.35	1.18	0.152
High BP	Normal BP	78	1			
	Have high BP	148	0.98	0.55	1.74	0.936
Abnormal lipid	Normal lipid	199	1			
	Abnormal lipid	27	0.25	0.03	1.80	0.168
Types of referral	GP ^c /Home	139	1			
	Hospital	87	0.61	0.35	1.07	0.085
Stroke Subtype	Non-	149	1			
	Haemorrhagic	77	2.17	1.18	3.97	0.012

^a hazard ratios obtained from simple Cox proportional hazard (PH) regression

^b confidence interval (CI) ^c General Practitioner

Table 3-5 Two prognostic models for in-hospital stroke fatality. The best model (Model 1, n=225) contains two independent predictors: age (years) and Glasgow Coma Scale (range of score is from 3 to 15). The alternative model (Model 2, n=226) has age (years) and stroke subtypes (haemorrhagic stroke vs non-haemorrhagic stroke) as the independent predictors. The parameters shown are the adjusted hazard ratios (HR), standard error (SE), lower and upper 95 confidence intervals (CI) for adjusted hazard ratios (HR) and the respective p-values

	Adj HR ^a	SE ^b	Lower 95% CI ^c	Upper 95% CI ^c	p-value
Best model (Model 1)					
n=225					
Glasgow Coma Scale (GCS)	0.82	0.03	0.77	0.88	<0.001
Age (years)	1.03	0.01	1.01	1.05	0.011
Alternative model (Model 2)					
n=226					
Haemorrhagic	2.40	0.75	1.30	4.44	0.005
Non-haemorrhagic	1.00				
Age (years)	1.03	0.01	1.01	1.05	0.014

^a hazard ratios obtained from multivariable Cox PH regression ^b standard error for hazard ratio

^c confidence interval

Our finding replicates the findings of at least two previous studies—both involved patients with ischaemic stroke—which had shown only age and stroke severity as the prognostic factors for stroke case fatality rates in their models. The first study was in Denmark (n = 26,818), and reported that age and the Scandinavian Stroke Scale (SSS) score predicted stroke fatality within the first week post-stroke (Andersen et al., 2011). The second study (n = 479), done in Switzerland, showed that advanced age and the National Institute of Health Stroke Study (NIHSS) were both independent predictors of early stroke mortality (Nedeltchev et al., 2010). Other studies have found that in addition to age and stroke severity, the important prognostic factors of stroke fatality include diabetes, sex and history of previous stroke (Andersen et al., 2005, Saposnik et al., 2008, Koton et al., 2009).

A review has reported that the relationship between sex and stroke fatality varies between studies (Reeves et al., 2008). In our results, sex was not a predictor of stroke fatality, consistent with the findings from other studies (Wahab et al., 2008, Appelros et al., 2003, Nedeltchev et al., 2010). Studies with positive relation between sex and stroke fatality results have shown that, in general, women had either: a) a greater risk than men for stroke fatality (Devries et al., 2013, Niewada et al., 2005) or b) a lower risk for stroke fatality than men (Sheikh and Bullock, 2007, Koton et al., 2009).

Our findings also showed that co-morbidities and physical profiles were not significantly related to stroke fatality. However, others have demonstrated that blood pressure status was a significant predictor of fatality in all major types of stroke (Johnston et al., 2009, Okumura et al., 2005). Besides blood pressure, diabetes mellitus (Koton et al., 2009, Wong, 1999, Okumura et al., 2005), abnormal heart rate such as atrial fibrillation (Appelros et al., 2003, Andersen et al., 2005), temperature and glucose (Koton et al., 2009), body mass index (Johnston et al., 2009) and anaemia (Hao et al., 2013) also predict fatality in certain types of stroke.

The GCS is widely used to assess stroke severity in the acute setting. However, newer measurement tools such as the SSS and the NIHSS can improve assessment of stroke severity if used alongside the GCS. The SSS (Andersen et al., 2005), level of consciousness (Andersen et al., 2005, Saposnik et al., 2008) and NIHSS (Teasdale and Jennett, 1974) are predictive of stroke outcome. Our study suggests that the roles of these new tools for assessing stroke severity should be investigated in Malaysia, as both of these tools are not standard clinical assessment tools in the emergency setting here.

Age has been found to be a significant predictor of morbidity and mortality (Devries et al., 2013, Fernandes et al., 2012, Saposnik et al., 2008, Wong, 1999). In our study, we found that the risk for fatality with age increased linearly. Studies looking specifically at the relationship between age and stroke fatality risk have found that the relationship can be non-linear (Olsen et al., 2011).

There are several limitations to this study. First, we used stroke patients from only one centre (HUSM). Furthermore, HUSM is an advanced university hospital providing one of the best stroke care in Malaysia, and the population who attend this hospital may not be a random subset of the whole population of those at risk of stroke. Because of this, it limits the generalisability of our results to all patients treated at other hospitals in Malaysia. Second, because of the unavailability of a stroke registry, data were abstracted from patient electronic medical registration information, which might create bias due to poor standardisation in data entry and quality control. Thirdly, many important stroke fatality predictors were not routinely collected at HUSM thus limiting the number of covariates in our model. We also used in-hospital patients who have different characteristics from out-of-hospital stroke patients (Shigematsu et al., 2013).

In the survival analysis, we assumed the censoring processes to be non-informative. This is based on the shared dependency between the covariates (for example, stroke subtypes and

Glasgow Coma Scale) and the outcome as shown by the Cox regression and the reversed Cox regression in Appendix H. This evidence has allowed the use of Kaplan-Meier and Cox proportional hazard models in this study. However, in the case when this evidence is regarded as inadequate, and thus informative censoring poses a limitation to the numerical estimation by the Kaplan-Meier and the Cox regression models in this study, multiple imputation methods and sensitivity analysis can be performed. Multiple imputation methods, such as the Gamma imputation and the Risk Score imputation methods, impute observation to the missing observations, and sensitivity analysis allows – for examples – the Cox estimation to be done in different scenarios (see Appendix H). Next, in this study, data consisted of 226 patients with only 53 of them deceased at discharge. There were also ‘thin cells’ (cells with small frequencies), for example, for the variables of race, marital status and abnormal lipids (Hosmer et al., 2011). These, too, would reduce the power of our Cox proportional models. The strengths of the study include the use of CT scan images or radiological reports for case confirmation. A CT scan is the diagnostic test of choice for stroke imaging (Falcone et al., 2013a) and improve the validity of the study because: a) CT scan images help provide accurate stroke subtyping, (Kalantri and Kalantri, 2010, Runchey and McGee, 2010), and b) CT scan images provide more objective assessment of severity than physical examination alone (Thorvaldsen et al., 1995, Andersen et al., 2011, O'Donnell et al., 2010a, Keir et al., 2002).

3.6 Conclusion

Sex, type of referral, history of abnormal lipid profile, GCS, Hg level and TWC count are significantly different between in-hospital stroke patients who were alive and dead at discharge. A Cox proportional hazard model for predicting in-hospital stroke fatality contained two independent prognostic factors: age and GCS score. Both variables are easy and quick to perform in an emergency setting. However, clinicians in Malaysia should consider newer

clinical assessments for stroke severity such as the SSS or NIHSS to complement current stroke assessment on admission.

3.7 Acknowledgements

We thank Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia for funding the study under the short-term research grant (304/PPSP/61312028). We appreciate the contribution of Dr Juhara Haron in reviewing the CT brain scan images and the assistance of Ms Kartini Daud, during data entry.

Chapter 4 Comparing Cerebral Infarction and Haemorrhagic Stroke in Asia: Case–Fatality and Prognostic Effect for In-hospital Survival

Kamarul Imran Musa¹, Juhara Haron², Peter J Diggle³, Thomas J Keegan³

¹ Department of Community Medicine, School of Medical Science, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

² Department of Radiology, School of Medical Science, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

³ Lancaster Medical School, Lancaster University, LA1 4YD, UK

Corresponding author:

Kamarul Imran Musa

Lancaster Medical School,

Lancaster

United Kingdom LA14YW

Email: k.musa@lancaster.ac.uk

Telephone: +44 1524594539

4.1 Abstract

Aims To describe the on-admission baseline demographic and clinical variables for patients with cerebral infarction or haemorrhagic stroke and to compare the stroke case–fatalities and the prognostic effect between them in an Asian population.

Methods Data from two groups of consecutive stroke patients: a) cerebral infarction (n = 150) and b) haemorrhagic stroke (n = 142) were collected from hospital medical records and patient clinical folders. The overall stroke fatality and fatalities at 3 days, 7 days and 14 days were described. The survival probability was estimated using the Kaplan–Meier estimate. Cox proportional hazard regression was used to estimate the crude and adjusted hazard ratio for stroke fatality between haemorrhagic stroke and cerebral infarction. The assumption of hazard proportionality was checked.

Results On admission, patients in the haemorrhagic stroke group were younger, had poorer Glasgow Coma Scale scores and higher mean diastolic blood pressure than those in the cerebral infarction group. The overall stroke case fatality for cerebral infarction was 12.0%; that for haemorrhagic stroke was 46.5%. In the haemorrhagic stroke group, patients with subarachnoid haemorrhage had the highest fatality. The median survival time for patients in the cerebral infarction group was 28 days; that in the haemorrhagic stroke group was 14 days. In the crude and adjusted models (adjusted for sex then age, respectively), patients in the haemorrhagic stroke group had more than twice the risk for stroke fatality than those in the cerebral infarction group.

Conclusion Patients in the haemorrhagic stroke group had less favourable presentation on admission, higher overall stroke fatality and higher 3-day, 7-day and 14-day fatality than those with cerebral infarction. They also had much higher risk for stroke fatality than the patients

with cerebral infarction. This indicates a need for improved critical care to minimise intracranial bleeding.

Keywords: Cerebral infarction, Haemorrhagic stroke, Stroke case fatality

4.2 Introduction

Stroke is the second leading cause of death worldwide (Feigin, 2007). It is a heterogeneous disease commonly classified into three major subtypes: a) ischaemic stroke (cerebral infarction), b) haemorrhagic stroke and c) subarachnoid haemorrhage (SAH) (Amarenco et al., 2009).

Ischaemic stroke constitutes about 70% of all stroke but the overall burden of haemorrhagic stroke is significantly higher (Krishnamurthi et al., 2014) because of its severity (Testai and Aiyagari, 2008). In 2000–2008, the frequency of haemorrhagic stroke in the low- to middle-income countries was twice that in the high-income countries (Feigin et al., 2009) and was also greater in the Asian populations (Wong, 1999, Xu et al., 2010, Basri and Azman Ali, 2003, Kaul et al., 2009, O'Donnell et al., 2010b, Durai Pandian et al., 2007).

Much of stroke research carried out to date has tended to focus on cerebral infarction, probably because it is more common than haemorrhagic stroke (Fernandes et al., 2012, Basri and Azman Ali, 2003, Feigin et al., 2009). Stroke research also largely comes from the developed countries despite the greater burden of stroke being in the low- and middle-income countries (Johnston et al., 2009), many of which are in Asia. Statistics on stroke also show large variations between countries (Ingall et al., 2000). Given these shortcomings, more studies are necessary to better understand the differences between stroke subtypes.

To the best of our knowledge, there have been only 4 published papers on stroke fatality in Malaysia. All were hospital-based studies, and two are more than 10 years old. One

paper reported that in 2002, the in-hospital stroke case fatality rate was 37.3% (Jaya et al., 2002). A 2003 paper reported that the 30-day in-hospital stroke case–fatality rate was 11.7% for ischaemic stroke and 27.3% for haemorrhagic stroke (Basri and Azman Ali, 2003). The other two papers were on primary intracerebral haemorrhage (ICH) studies (Sia et al., 2007, Yousuf et al., 2012). In Kuala Lumpur, Malaysia, the fatality rate at discharge was 43.9% (Sia et al., 2007), while a recent study in Kuantan, Malaysia, showed that among ICH patients, the 24-hour fatality rate was 32.7% and that the predictors of fatality included stroke subtype, location of haemorrhage and bleeding volume but not sex or age (Yousuf et al., 2012).

We aimed to address these shortcomings in data by studying stroke admissions to a large hospital in Malaysia, where the patients are of Asian origin. Our interest was in the proportion of strokes that were haemorrhagic and relative survival by stroke type. Access to computed tomography (CT) scans for each patient admitted to our study hospital allowed us to use accurate and reliable diagnoses to categorise patients by stroke subtype.

This study had two primary aims: a) to describe the on-admission baseline demographic and clinical variables for cerebral infarction and haemorrhagic stroke and b) to compare the prognostic effect of haemorrhagic stroke against cerebral infarction for stroke fatality.

4.3 Methods

4.3.1 Study venue

The study took place at Hospital Universiti Sains Malaysia (HUSM), Kelantan, Malaysia. HUSM is one of the two largest tertiary hospitals—hospitals with specialists—in Kelantan. At HUSM, a neurology team manages all stroke patients. Those admitted to the ward are treated in the neurology unit and attended to by a group of specialists consisting of neurologists, neurosurgeons and neuroradiologists.

4.3.2 Case definition

Stroke cases were identified from HUSM records. HUSM uses the International Classification of Diseases 10th Revision (ICD-10) to code diagnosis at discharge; this coded diagnosis is termed 'administrative diagnosis'.

Diagnoses eligible for inclusion in our study were ICD-10 Chapter IX Block I60 (subarachnoid haemorrhage), I61 (intracerebral haemorrhage), I62 (other non-traumatic intracranial haemorrhage such as extradural or subdural haemorrhage) and I63 (cerebral infarction). Collectively, Blocks I60, I61 and I62 are known as haemorrhagic stroke. Our cases were all cases of stroke with ICD-10 Blocks I60, I61, I62 and I63 who had been admitted and diagnosed at HUSM between December 2008 and June 2012 (ICD-10 Block I60, I61, I62) or between January 2011 and June 2012 (ICD-10 Block I63).

Data on 150 consecutively admitted stroke patients in the cerebral infarction group and 150 consecutive patients in the haemorrhagic stroke group were extracted from hospital records. Eight patients in the haemorrhagic stroke group were excluded because three had a stroke secondary to trauma and five were misdiagnosed as having had a stroke, resulting in a final study sample of 150 patients with cerebral infarction and 142 with haemorrhagic stroke.

The stroke diagnosis extracted from each medical record was verified by a radiologist (JH), who cross-checked the written diagnosis in the medical records with the CT images stored by the HUSM picture archiving and communication system (PACS) server. A CT scan is the investigation of choice to subtype a stroke based on imaging (Falcone et al., 2013a) to differentiate stroke subtypes (Falcone et al., 2013a, Kalantri and Kalantri, 2010, O'Donnell et al., 2010b, Warlow, 1998, Runchey and McGee, 2010). In the event where CT scan images were not available in the PACS server (e.g. a CT scan was done by the referring centre), we verified

the stroke diagnosis based on the radiological diagnosis recorded in the referral letter provided by the patient.

4.3.3 Variables

For this study, the HUSM medical records office provided data on: a) age, b) sex, c) date of admission, d) date of discharge and e) survival status at discharge for each patient. In addition, KIM reviewed each patient's hard copy medical record and extracted the following additional data: a) type of referral centre, b) Glasgow Coma Scale (GCS) score, c) systolic blood pressure (SBP, mmHg), d) diastolic blood (DBP, mmHg) and e) stroke diagnosis.

At HUSM, stroke severity is assessed using the GCS. The GCS is a widely-used measure of a patient's state of consciousness. It uses a scale between 3 (the worst) and 15 (the best) points. The GCS score is a total of scores from its 3 components: a) best eye response (1–4 points), b) best verbal response (1–5 points) and c) best motor response (1–6 points) (Singh et al., 2000).

4.3.4 Statistical analysis

All data were entered into EpiData Entry software (Lauritsen, 2000) and analysed using Stata 11.2 SE (StataCorp., 2010).

We calculated the stroke case fatality by dividing the number of deaths from all causes after admission by the number of admissions for stroke, and multiplying by 100 (Roberts and Goldacre, 2003, Saposnik et al., 2008). Because the calculation was performed for only deaths during hospitalization of acute stroke in our study, our stroke fatality is best coined as 'in-hospital stroke fatality' (Roberts and Goldacre, 2003). We calculated the overall stroke case fatality and case fatality at 3 days, 7 days and 14 days. Stroke fatality studies differs in; 1) their chosen time frame – for example within 28 days, 21 days to one month, 7 days, 30 days, and

1 year of stroke admission – and 2) whether the researchers restrict or do not restrict to stroke deaths during hospital admission (Thorvaldsen et al., 1995, Feigin et al., 2009, Saposnik et al., 2008).

In assessing survival, the outcome variable was time-to-stroke fatality (death) after admission to HUSM. The time was calculated in days (date of admission until date of discharge) and the event was defined as either failure (death due to stroke, coded as 1) or censored (alive at discharge or death from causes other than stroke, coded as 0). Kaplan–Meier survival curves were plotted and the survival experience between cerebral infarction and haemorrhagic stroke was tested using the log-rank test (Hosmer et al., 2011). Kaplan-Meier estimate provided the fraction of subjects (stroke patients) living for a certain amount of time (days) after acute stroke and can be formally defined as ‘the probability of surviving in given length of time while considering time in many small intervals’ (Goel et al., 2010).

Cox proportional hazard regression was used to produce hazard ratios (HR) and to compare the prognostic effect between stroke subtypes. To facilitate the interpretation of HR, the cerebral infarction group was chosen as the reference (baseline) category because the risk of death from cerebral infarction is lower than that from haemorrhagic stroke. We compared the prognostic effect of stroke subtypes in 3 models: 1) stroke subtypes alone, 2) stroke subtypes with age and 3) stroke subtypes with sex. Since the aim of analysis is to look at the prognostic effect of stroke subtypes, we felt adding only age and sex covariates, one at a time, was adequate.

The Schoenfeld residuals generated by the Cox proportional hazard regression were used to check for the assumption of hazard proportionality (Hosmer et al., 2011). In Stata, ‘*estat phtest*’ tests the hazard proportionality for the overall model (global test) and for each covariate (detailed tests). All tests were 2-tailed and p-values of less than 0.05 were considered significant.

In the Cox regression analysis, three assumptions about censoring (censored because they were discharged alive) including the independent censoring, the random censoring and the non-informative censoring (Resche-Rigon et al., 2006, Kleinbaum and Klein, 2012) were considered not violated. This can happen when there is a presence of shared dependency between the covariate (such as stroke subtypes in the model) and the outcome (Figure H.1). The results to support the shared dependency in this study are shown in Table H.3 and Table H.4. But the independence between time (T) and censoring ($\delta = 0$) to provide non-bias estimates in survival analysis requires a strong assumption. In the event that informative censoring is suspected, multiple imputation methods can be performed. Two recent methods are: a) the Gamma imputation method and b) the Risk Score imputation method. Both methods will replace missing observations with imputed values based on different approaches. Imputation can be done in different parameters set by users, and this will produce different estimates depending on the parameters. Sensitivity analysis can then be performed on these different estimates (See Appendix H for further details).

4.3.5 Ethical approval

We received ethical approval from the Medical and Research Ethics Committee (MREC), Ministry of Health, Malaysia (NMRR-12-471-12139), the USM Human Research Ethics Committee (HREC), Universiti Sains Malaysia (JPEM [242.4.(1.4)]) and the Lancaster University Research Ethics Committee.

4.4 Results

Data from 150 consecutive patients with cerebral infarction and 142 consecutive patients with haemorrhagic stroke treated at the HUSM neurology ward were extracted from the HUSM records office and clinical folders. In the haemorrhagic stroke group, 21.1% (30/142) had SAH,

59.2% (84/142) had primary ICH and 19.7% (28/142) had other types of intracranial haemorrhagic stroke. All patients were of Asian origin.

Table 4-1 shows the variables age (years), length of stay (days), GCS score, SBP and DBP for the two groups as stratified by sex. The haemorrhagic stroke group was further grouped into a) SAH, b) primary ICH and c) others. Patients in the haemorrhagic stroke group were significantly younger, had poorer GCS scores and had higher DBP (mean age = 57.6 years, mean GCS score = 10.1, mean DBP = 94.3 mmHg) than the patients admitted with cerebral infarction (mean age = 62.5 years, mean GCS score = 13.5, mean DBP = 91.8 mmHg).

The distribution of stroke patients based on referral type (centre) is presented in Table 4-2. A chi-square test showed a significant difference ($p < 0.001$) in the distribution of referral type: a large proportion of patients in the cerebral infarction group (71.3%) had attended as self-referral to HUSM but a large proportion of patients in the haemorrhagic stroke group were referred from other tertiary or district hospitals (32.4%, 26.8% and 33.3%, 27.5%, respectively).

4.4.1 Stroke fatality

Table 4-3 shows the overall stroke case fatality by stroke type, and stroke fatalities at 3 days, 7 days and 14 days. The fatalities were deaths due to all causes while our stroke patients were being treated at the HUSM. The overall case fatality rate for cerebral infarction was 12.0% (18/150). At 3 days, 7 days and 14 days, the fatalities were 4.0% (6/150), 8.7% (13/150) and 10.7% (16/150), respectively. The overall case fatality rate in the haemorrhagic stroke group was 46.5% (66/142), approximately four times higher than that of the cerebral infarction group. At 3 days, 7 days and 14 days, the fatalities were 22.5% (32/142), 33.1% (47/142) and 39.4% (56/142), respectively.

Among the haemorrhagic stroke subgroups, the SAH group had the highest overall case fatality 66.7% (20/30). Case fatalities in SAH patients were 6 times higher at 3 days and almost 5 times higher at 7 days and 14 days post admission than that in patients admitted with cerebral infarction. More than half (56.0%) of patients with ICH died within 14 days of admission. The distribution of stroke fatalities at each of the 4 time points differed significantly between the haemorrhagic stroke and cerebral infarction groups. The distribution of overall fatality differed significantly between the subgroups of haemorrhagic stroke but not the distribution of fatalities at 3 days, 7 days and 14 days.

4.4.2 Prognostic factors between ischaemic and haemorrhagic stroke

Figure 4-1 shows the Kaplan–Meier survival estimates for the cerebral infarction group (n = 150) and haemorrhagic stroke group (n = 142). The median survival time was significantly shorter (log-rank test $p < 0.001$) in the haemorrhagic stroke group (median=14 days) than in the cerebral infarction group (median=28 days). At 10 days after admission, there were only 11 patients out of 150 CI patients who were in the ward while the rest (139 patients) have been discharged alive by the treating physicians. They were considered as censored patients in our analysis. Among the 142 HS patients, there were still 40 patients who were alive and being treated in the ward at 10 days. At day 50, none of the CI patients was in the ward but two (n=2) HS patients were still being treated.

Table 4-4 and Table 4-5 show the survival parameters estimated for the Cox proportional hazard regression model in a model with stroke subtype as the only covariate and for two covariate-adjusted models.

Table 4-4 shows that the crude and adjusted risks (adjusted for sex) for stroke fatality in the haemorrhagic stroke group were both 2.61 times higher than that of the cerebral infarction group (95% confidence interval [CI]: 1.53, 4.44). When the model was age-adjusted,

the risk for stroke fatality for the haemorrhagic stroke group increased to 2.78 (95% CI: 1.62, 4.76).

The risk estimates for stroke case fatality for the haemorrhagic stroke subgroups, i.e. SAH, ICH and others, were compared against cerebral infarction, as shown in Table 4-5. The univariate Cox regression model shows that all haemorrhagic stroke subtypes posed significantly higher risk for stroke fatality than cerebral infarction: the crude HR was 2.88, 2.68 and 2.17 times greater, respectively, than that for cerebral infarction. When adjusted for age then sex in multivariate Cox hazard regression, the risk estimates for stroke fatality of the haemorrhagic stroke subgroups remained higher than that of the cerebral infarction group, with SAH carrying the highest risk for stroke fatality.

4.5 Discussion

Our study describes the on-admission characteristics of stroke patients and compares stroke case-fatalities and prognostic effects between patients with cerebral infarction or haemorrhagic stroke in an Asian population. We found that haemorrhagic stroke patients were younger, more commonly referred from hospital settings than other settings and had lower GCS scores and higher DBP on admission than cerebral infarction patients. Overall, 3-day, 7-day and 14-day stroke case fatalities (in-hospital stroke fatality) were all higher in the haemorrhagic stroke group than in the cerebral infarction group. The median survival time was shorter in the haemorrhagic stroke group (median=14 days) than in the cerebral infarction group (median=28 days). When not adjusted or when adjusted for either age or sex, haemorrhagic stroke patients had more than twice the risk for in-hospital stroke fatality than cerebral infarction patients. Our findings show that among the haemorrhagic stroke subgroups—SAH, ICH and other types of brain haemorrhage—SAH had the highest risk for in-hospital stroke fatality, followed by ICH, then by the other types of brain haemorrhage.

Table 4-1 The means and standard deviations (SD) for age, length of stay (LOS), Glasgow Coma Scale (GCS), systolic blood pressure (SBP in mmHg) and diastolic blood pressure (DBP in mmHg) stratified for sex and stroke subtype

Variable	Levels ^e	CI ^f		HS ^g		SAH ^h		ICH ⁱ		Others ^j		All	
		n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)	n	Mean(SD)
Age	M	69	61.9(12.7)	57	59.5(14.1)	6	46.3(20)	36	56.4(13.8)	15	61.6(15.6)	126	59.5(14.1)
	F	81	63.1(13.9)	85	60.7(13.7)	24	61.9(10.3)	48	56.4(14.6)	13	59.1(11.6)	166	60.7(13.7)
LOS ^a	M	69	4(4)	57	5.8(7.4)	6	12(9)	36	6(4.8)	15	10.7(16.2)	126	5.8(7.4)
	F	81	5.4(6.2)	85	7.9(8.5)	24	10.8(9.1)	48	6.2(6.5)	13	9.1(12.3)	166	6.7(7.5)
GCS ^b	M	69	14(2.6)	57	12.7(3.5)	6	7.7(4.5)	36	11.2(3.7)	15	12.3(3.6)	126	12.7(3.5)
	F	81	13.2(3)	85	9.4(4.4)	24	10(4)	48	9.3(4.5)	13	9.1(4.6)	166	11.3(4.2)
SBP ^c	M	69	159.5(26.8)	57	165.4(30.7)	6	151(24)	36	184.4(30.1)	15	150.8(31.6)	126	165.4(30.7)
	F	81	168.1(36.8)	85	161.5(37.3)	24	156.3(34)	48	167(38)	13	150.5(39.7)	166	164.7(37.1)
DBP ^d	M	69	90.5(17.8)	57	95.6(22)	6	93.7(19.1)	36	109.3(25.6)	15	86.3(16.8)	126	95.6(22)
	F	81	92.9(19.2)	85	93.6(22.1)	24	86.8(14.6)	48	99.7(24)	13	83.8(20.6)	166	93.3(20.7)

^a LOS=length of stay (days) ^b GCS=Glasgow Coma Scale (min=3,max=15), ^c SBP=Systolic blood pressure ^d DBP=Diastolic blood pressure ^e M=male,F=female ^f CI=cerebral infarction ^g HS=haemorrhagic stroke ^h SAH=subarachnoid haemorrhage ⁱ ICH=intracerebral haemorrhage and ^j Others=other intracranial haemorrhage .

Note: HUSM uses administrative diagnosis based on the International Classification of Diseases (ICD) 10. In ICD-10, cerebrovascular diseases were coded in Chapter I60 to I69 where cerebral infarction is coded as I63, subarachnoid haemorrhage is coded as I60, Intra-cerebral haemorrhage is coded as I61. We grouped stroke patients under 'other intra-cranial haemorrhage' for stroke patients with ICD-10 I62 or for stroke patients with CT images/reports show the combination of I60 and/or I61 and/or I62

Table 4-2 The frequencies and percentages of the type of referral for stroke patients to the Emergency Department, Hospital Universiti Sains Malaysia (HUSM) for different stroke subtypes

Referring centre ^a	Cerebral infarction, n(%)	Haemorrhagic stroke, n(%)
From other tertiary hospital	7(4.9)	46(32.4)
From general practice	4(2.8)	2(1.4)
From district hospital	29(20.3)	38(26.8)
Direct from scene (Self-referral)	102(71.3)	52(36.6)
Not documented	7(4.7)	4(2.8)
TOTAL	150(100.0)	142(100.0)

^a Pearson chi-sq(dof)=46.4(4), p-value<0.001

Table 4-3 The overall stroke case fatality and the 3 days, 7 days and 14 days stroke fatalities for patients with cerebral infarction (CI) and haemorrhagic stroke (HS)

Fatalities for the cerebral infarction (CI) and haemorrhagic stroke (HS)					
Stroke subtypes	All patients	Overall Dead (%)	3-day Dead (%)	7-day Dead (%)	14-day Dead (%)
CI	150	18(12.0)	6(4.0)	13(8.7)	16(10.7)
HS	142	66(46.5)	32(22.5)	47(33.1)	56(39.4)
	292	p-val<0.001	p-val<0.001	p-val<0.001	p-val<0.001

Fatalities for the subtypes of the haemorrhagic stroke (HS)					
Stroke subtypes	All HS patients	Overall Dead (%)	3-day Dead (%)	7-day Dead (%)	14-day Dead (%)
SAH	30	20(66.7)	8(26.7)	11(36.7)	13(43.3)
ICH	84	33(39.3)	18(21.4)	37(44.0)	47(56.0)
Others	28	13(46.4)	3(10.7)	10(35.7)	10(35.7)
	142	p-val=0.036	p-val=0.301	p-val=0.644	p-val=0.137

^a CI=cerebral infarction ^b HS=haemorrhagic stroke ^cSAH=subarachnoid haemorrhage ^d ICH=intracerebral haemorrhage ^e Others=other intracranial haemorrhage.

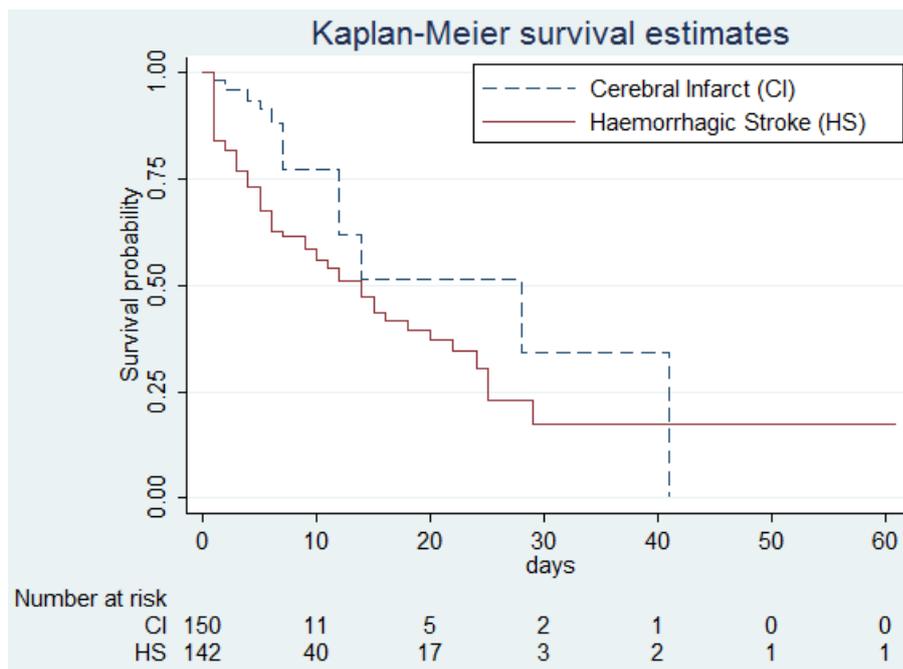


Figure 4-1 Kaplan-Meier survival estimates and curves of overall survival for the cerebral infarction (CI) group and haemorrhagic stroke (HS) group. The number at risk shows the numbers of patients still alive, and therefore at risk for stroke death at different period of time (days). The survival probability indicates the Kaplan-Meier probabilities (conditional probabilities) that a patient survives so many days or longer. The median survival time for CI was 28 days and for HS was 14 days – meaning that a patient with CI has a 50% probability of surviving 28 days or longer and for a patient with HS has a 50% probability of surviving 14 days or longer. The log rank test (results not shown) gave the p-value <0.001

Table 4-4 The crude hazard ratios (HR) and adjusted HR (adjusted for age then sex), standard errors and 95% confidence intervals for HR estimated using the Cox proportional hazard regression model. The haemorrhagic stroke (HS) was compared against the cerebral infarction (CI as the baseline group). In all comparisons, p-values yielded were less than 0.001.

Covariates	n	HR ^a	SE ^b	Lower 95% CI ^c	Upper 95% CI ^c	p-value
stroke subtypes	HS ^d 142	2.61	0.71	1.53	4.44	<0.001
	CI ^e 150	1.00				
stroke subtypes + age	HS 142	2.78	0.76	1.62	4.76	<0.001
	CI 150	1.00				
stroke subtypes + sex	HS 142	2.61	0.71	1.53	4.44	<0.001
	CI 150	1.00				

^a HR=Hazard ratio ^b SE=Standard error for HR ^c CI=confidence interval for HR ^d HS=Haemorrhagic stroke

^e CI=cerebral infarction

Table 4-5 The crude hazard ratios (HR) and adjusted HR (adjusted for age then sex), standard errors and 95% confidence intervals for HR estimated using the Cox proportional hazard regression model. The haemorrhagic stroke (HS) was compared against the cerebral infarction (CI as the baseline group). All p-values were less than 0.050.

Covariates		n	HR ^a	SE ^b	Lower 95% CI ^c	Upper 95% CI ^c	p-value
stroke subtypes	SAH ^d	30	2.88	0.97	1.48	5.59	0.002
	ICH ^e	84	2.67	0.79	1.50	4.77	0.001
	Others ^f	28	2.17	0.82	1.04	4.56	0.040
	CI ^g	15	1.00				
stroke subtypes + age	SAH ^d	30	3.11	1.06	1.59	6.06	0.001
	ICH ^e	84	2.90	0.86	1.61	5.20	<0.001
	Others ^f	28	2.20	0.84	1.05	4.64	0.038
	CI ^g	15	1.00				
stroke subtypes + sex	SAH ^d	30	2.74	0.93	1.41	5.34	0.003
	ICH ^e	84	2.68	0.79	1.50	4.79	0.001
	Others ^f	28	2.28	0.86	1.09	4.79	0.029
	CI ^g	15	1.00				

^a HR=Hazard ratio ^b SE=Standard error for HR ^c CI=confidence interval for HR ^d SAH=subarachnoid hemorrhage ^e ICH=intracerebral haemorrhage ^f Others=intra-cranial haemorrhage (acute non-traumatic subdural haemorrhage or non-traumatic extradural haemorrhage) or patients with CT imaging showing the combinations of SAH, ICH or CI ^g CI=cerebral infarction

We also reported differences in on-admission variables between patients with haemorrhagic stroke and patients with cerebral infarction. Patients with haemorrhagic stroke were younger on admission than those with cerebral infarction were. This reinforces evidence from previous research showing that the frequency of haemorrhagic stroke is greater in young adults than in the elderly (Subha et al., 2015). In this study, patients with haemorrhagic stroke were also more commonly referred from other hospitals rather than from other settings. This was also seen in earlier studies. A likely explanation is that it is a more severe condition (Testai and Aiyagari, 2008, Andersen et al., 2009, Aronowski and Zhao, 2011) and requires more advanced care at a better-equipped centre (Giraldo, 2015c, Giraldo, 2015a).

In Kelantan, Malaysia, patients with haemorrhagic stroke are referred to HUSM from other hospitals because it is a centre that provides advanced care in the Northeast of Malaysia especially for cases requiring neurosurgical procedures. Our results showed these referrals were more likely when the stroke was haemorrhagic than for other types. The implication for our study of this referral pattern is the possibility of the introduction of two types of bias. a) Berkson's bias and b) centripetal bias. Both are classified under 'selection bias' (Delgado-Rodriguez and Llorca, 2004). Berkson's bias is produced when, for example, the probability of hospitalisation for two different diseases differ and centripetal bias means patients are drawn to a specific centre because of its reputation (Delgado-Rodriguez and Llorca, 2004). There is also a possibility that the higher fatality in haemorrhagic stroke could be due to the referral source or them being in worse clinical condition than those with ischaemic stroke. To provide more information on this a further analysis, such as a stratified analysis (stratified Cox or stratified logistic regression), could be performed in on each stratum (or level) of the referral source (Hosmer et al., 2011, Kleinbaum and Klein, 2012).

In our study, patients with haemorrhagic stroke had significantly lower GCS scores because the intracranial environment has been more severely compromised due to more severe pathologies. The pathologies include disruption of the normal brain anatomy due to the accumulation of haematoma, increased local pressure, increased blood toxicity and the accumulation of oxidative stress after cell death (Aronowski and Zhao, 2011).

Haemorrhage (bleeding) occurring from a focal blood vessel in the brain parenchyma gives rise to a condition known as ICH; if from sudden bleeding into the subarachnoid space, it causes a condition known as SAH, commonly due to a ruptured aneurysm leading to more global clinical manifestations (Giraldo, 2015c).

Our analyses show that all haemorrhagic stroke subtypes—SAH, ICH and other types of brain haemorrhage—pose significantly higher risk for fatality because of the extent of brain damage due to bleeding and rebleeding (Roos et al., 2000).

In cerebral infarction, the pathophysiology is more limited (localised) because the main culprit is occlusion of the brain arteries resulting in lack of blood supply, which is clinically manifested as sudden neurological deficit. In cerebral infarction, thrombolytic therapy is the standard treatment and only rarely is surgical intervention necessary (Giraldo, 2015b).

The mean age on admission for the cerebral infarction patients in this study is similar to those in other studies involving Asian populations (Wong, 1999, Basri and Azman Ali, 2003, Christopher et al., 2007). Our ICH patients were almost 5 years younger than ICH patients reported in a similar study (mean age = 61.3 years) pooled from 36 hospitals all over Asia (the Asian Acute Stroke Advisory Panels study) (Wong, 1999).

In our study, both the cerebral infarction and haemorrhagic stroke patients were younger on admission than those in studies in the high-income countries such as Korea (Kim, 2009), Finland (Huhtakangas et al., 2013), Canada (Saposnik et al., 2008), Sweden (Nedeltchev et al., 2010) and Denmark (Andersen et al., 2011). This is in line with the finding reported in a well-known review of stroke epidemiology (Feigin et al., 2009), where the younger age groups in the low- to middle-income countries were at higher risk of stroke than their counterparts in the high-income countries.

Our analyses show that, based on the GC scores, patients in the haemorrhagic stroke group had more impaired consciousness, which is consistent with the findings of others (Singh et al., 2000, Giraldo, 2015b, Giraldo, 2015c, Giraldo, 2015a). We also report that on admission, patients in the haemorrhagic stroke group presented with higher blood pressure. This

supports previous research reporting that blood pressure is an important risk factor for haemorrhagic stroke (Suzuki et al., 2011, Truelsen et al., 2007, Del Brutto et al., 2013).

In this study, the overall stroke case fatality in the cerebral infarction group was similar to that in an earlier report in Malaysia (Basri and Azman Ali, 2003) but was higher than that in a study based on other Asian populations (only 8.8%) (Wong, 1999). The 3-day, 7-day and 14-day stroke case fatality rates in the cerebral infarction group were all higher than that reported in Denmark (3-day fatality = 1.9%, 7-day fatality = 3.3%) (Andersen et al., 2011), Canada (7-day fatality = 6.9%) (Saposnik et al., 2008), Brazil (10-day fatality rates in three of four centres studied were between 4.9% and 7.5%) (Fernandes et al., 2012) and in a large international study—The International Stroke Trial—where the 14-day stroke fatality rate was only 4.6% (Czlonkowska et al., 2002).

The overall stroke case fatality in the haemorrhagic stroke group in this study matched that of a previous local study (Basri and Azman Ali, 2003) but was higher than that in studies elsewhere in Asia (Wong, 1999) and in Brazil (Fernandes et al., 2012). Our comparisons with other studies are limited because few studies focus on haemorrhagic stroke and because the use of different durations of observation (e.g. days) renders comparisons difficult.

The difference in the results of stroke fatality between our study and other studies could be due to: 1) the difference in the baseline characteristics such as age, sex, stroke severity and comorbidities, 2) the quality of treatment on admission, 3) the quality of care during admission, 4) the variations in the length of observation (e.g. days), and 5) the use of only in-hospital data (limited to only data during admission) or combination of in-hospital and out of hospital data (data during admission and after admission) (Saposnik et al., 2011, Saposnik et al., 2008, Sridharan et al., 2009, Thorvaldsen et al., 1995, Wahab et al., 2008, Zhou et al., 2013, Delbari et al., 2011, Feigin et al., 2009, Feigin et al., 2003, Xu et al., 2010, Yousuf et al., 2012).

On admission, there were few differences between patients admitted with cerebral infarction and haemorrhagic stroke when variables such as SBP or GCS score were examined. This highlights possible difficulties in clinical decision-making when there is a need to distinguish between cerebral infarction and haemorrhagic stroke (Donnan, 2011, Martins et al., 2011). Misdiagnosis may lead to underestimation of haemorrhagic stroke incidence (Keir et al., 2002, Shiber et al., 2010). In view of this difficulty, in practice the ability to distinguish cerebral infarction and haemorrhagic stroke must be improved at the earliest possible point of contact with physicians or emergency workers. The improvement may include the use of additional screening tools and advanced neuroimaging services to enable quick and accurate identification of stroke subtype and objective quantification of the severity of stroke (Runchey and McGee, 2010, Keir et al., 2002).

To manage stroke patients effectively, especially those with haemorrhagic stroke, hospitals should be equipped with easy access to neuroimaging services to enable prompt use of at least a CT scan and to provide aggressive interventions to control blood pressure, correct the underlying coagulopathy and to obliterate vascular lesions in patients with a high-risk of re-bleeding in the case of haemorrhagic stroke (Diamond et al., 2003, Lavados et al., 2005, Roos et al., 2000). Policy-makers should make available a stroke unit in every major hospital. Studies and systematic reviews have shown that stroke units providing integrated acute stroke care through neurologists, neurosurgeons and neuroradiologists have been effective in reducing stroke mortality and morbidity (Ingall et al., 2000, Stroke Unit Trialists, 2007, Saposnik et al., 2011). In Asia especially, new resources should be directed towards developing and implementing stroke training for health care workers (Donnan, 2011).

The main strength of the study is that we recruited and extracted both the routine hospital and clinical data of nearly 300 stroke patients in the setting of a major hospital in a region of Southeast Asia, from which such data are sparse. The stroke analysis is also

strengthened by the use of CT scan images—the investigation of choice in stroke (Falcone et al., 2013a)—or at least formal radiological reports, to establish stroke subtype. This reduces the classification bias of stroke subtype.

We are aware of a few limitations of the study. One of them is the restriction to hospital-based stroke patients. The severity and fatality in stroke patients in the community and in hospitals are likely to be different (Andersen et al., 2011), hence our inferences pertain to hospitalised stroke patients only. Second, we cannot generalise these results to stroke patients in all types of hospitals because our subjects were treated in a tertiary care hospital equipped with a neurology unit. Much of tertiary hospitals in the Asian countries, including Malaysia, still lack an integrated neurology care stroke unit. Thirdly, we assumed that one of the censoring processes was non-informative. We based this assumption on the evidence that the covariate of interest (stroke subtype) shares its dependency with the outcome based on the Cox regression and reversed Cox regression analyses (see Figure H.1). In both analyses, the estimates of stroke subtype remain plausible and significant as shown in Table H.3 and Table H.4. On the other hand, censored observations provide non-negligible information (informative censoring) when there is no exact independence between time and censoring. In cases where censoring is informative, the numerical estimates from Kaplan-Meier and Cox proportional hazards become biased. In this instance, the multiple imputation method can be performed, and this method replaces the missing observations. The imputation can be performed using different parameters to mimic different scenarios. In Appendix H, we outline the principle of multiple imputation for Cox regression and describe two recent methods (Gamma imputation and Risk Score imputation) that are available as software packages. A sensitivity analysis (a technique used to determine how different values impact a particular dependent) can then be performed.

4.6 Conclusion

This study has shown that on admission to hospital, haemorrhagic stroke patients present at a younger age and have lower GCS scores and higher blood pressure than cerebral infarction patients. Overall, 3-day, 7-day and 14-day in-hospital stroke case fatality are all higher in patients with haemorrhagic stroke—especially SAH—than in those with cerebral infarction. The median survival time is shorter in patients with haemorrhagic stroke (14 days) than in patients with cerebral infarction (28 days). Haemorrhagic stroke subtypes—intracerebral, subarachnoid and other type of brain haemorrhage—carry at least twice the risk (crude and adjusted) for in-hospital stroke fatality than cerebral infarction.

4.7 Acknowledgements

We thank Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia for funding the study (Research grant /304/PPSP/61312028). We appreciate the assistance of Ms Kartini Daud with data entry.

Chapter 5 Barthel Index Scores over Three Months are Related to Age and Stroke Subtype in Asian Stroke Patients

Kamarul Imran Musa¹, Peter J Diggle², Thomas J Keegan²

¹ Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kbg Kerian, Kelantan, Malaysia

² CHICAS, Lancaster Medical School, Lancaster University, LA1 4YM Lancaster, UK

Corresponding author:

Kamarul Imran Musa

Lancaster Medical School,

Lancaster

United Kingdom LA14YW

Email: k.musa@lancaster.ac.uk

Telephone: +44 1524594539

5.1 Abstract

Aims To describe the change in Barthel Index score at baseline (at discharge) to 1 month and then 3 months post-baseline and to explore the relationship between age, sex and stroke subtype with Barthel Index score over the 3-month period.

Method A total 98 in-hospital stroke patients were recruited and their Barthel Index scores were assessed on 3 measurement occasions (at baseline, 1 month after baseline, 3 months after baseline). The Barthel Index was scored by telephone interview. The Barthel Index has 10 items, with total scores ranging from 0 (worst) to 100 (best). Analysis was done using a random intercept model taking into consideration the individual subject random effect.

Results The mean age of the patients ($n = 98$) was 60.7 years ($SD = 13.6$); 65.3% (64/98) were female and 73.7% (70/95) had ischaemic stroke. The overall Barthel Index score increased from baseline (mean = 35.1, $SD = 39.4$) to 1 month (mean = 64.4, $SD = 39.5$) and to 3 months (mean = 78.0, $SD = 38.9$). Over the 3-month period from the baseline, 13 patients died; their mean baseline Barthel Index score was different from the mean score of patients at discharge (Kolmogorov–Smirnov test $p = 0.048$). Measurement occasion, age and stroke subtype were significantly related with the Barthel Index score. Over the same period, the adjusted Barthel Index mean score increased from 35.1 to 68.7. The Barthel Index score decreased as age increased, while haemorrhagic stroke patients consistently had lower Barthel Index scores than ischaemic stroke patients. There was no significant interaction between the covariates in the model.

Conclusion Overall, the Barthel Index score increases from baseline to 3 months post-baseline. Measurement occasion, age and stroke subtype but not sex are related with Barthel Index score over a 3-month period.

Keywords: Barthel Index, Mixed-models, Stroke

5.2 Introduction

Stroke is a major public health problem that causes a substantial global burden of health (Krishnamurthi et al., 2014). About 10% of all deaths around the world are due to stroke, with millions of survivors left disabled (Strong et al., 2007). In 2005 alone, 16 million people suffered from a first-ever stroke (Mukherjee and Patil).

Acute stroke results in neurological, functional and cognitive decline (Lisabeth et al., 2014). After surviving a stroke, functional status generally improves over time (Rachpukdee et al., 2013). It is useful to estimate this change in functional status for 3 principal reasons: a) for clinicians to provide effective stroke care, anticipate discharge planning and support patients and family (Park et al., 2013), b) for researchers to explore the prognostic factors and underlying pathophysiology of stroke (Bhalla et al., 2013), and c) to facilitate comparisons between interventions in clinical trials (Harrison et al., 2013). Studies have shown that functional status—as well as other outcomes—after acute stroke are associated with age (Sohrabji et al., 2013, van Almenkerk et al., 2013, Khan et al., 2012, Kong and Lee, 2014), stroke severity, size of the stroke (van Almenkerk et al., 2013) and stroke subtype (Di Carlo et al., 2006, Paci et al., 2011, McNaughton et al., 2001).

One of the most well-known measures of functional status after stroke is the Barthel Index. It is the most commonly cited measure of disability in stroke rehabilitation trials in the published literature (Sangha et al., 2005). The Barthel Index was initially developed to act as a simple index for independence of function in patients post-stroke and as a useful scoring assessment during rehabilitation (Quinn et al., 2011). It has excellent reliability (Duffy et al., 2013) and its 10-item version—with scores of 0 to 100 in 5-point increments—is the most often used (Quinn et al., 2011).

In this study, we aimed to answer the following questions: a) how do Barthel Index scores change between baseline and at 1 and 3 months post-baseline? and b) are the covariates age, sex and stroke subtype associated with change in the Barthel Index score over a 3-month period? Our objectives were first to describe the Barthel Index scores at baseline, after 1 month, then 3 months post-baseline. Subsequently, we explored the relationship between specific covariates (age, sex, stroke subtype) and the Barthel Index scores over this 3-month period.

5.3 Methods

5.3.1 Patients

A total of 108 consecutive stroke patients diagnosed and admitted between 1 July 2013 and 31 October 2014 from two major tertiary hospitals in Kelantan, Malaysia: a) Hospital Universiti Sains Malaysia (HUSM) and b) Hospital Raja Perempuan Zainab II (HRPZ) were recruited for this study.

The stroke patients were eligible if they met the following inclusion criteria: a) stroke was the primary diagnosis for hospital admission, b) age > 18 years, c) stroke diagnosis was established by the neurology team, d) diagnosed and admitted between 1 July 2013 and 31 October 2014 and e) stroke subtype was limited to: a) ischaemic stroke (infarction of the central nervous system) or b) haemorrhagic stroke (spontaneous non-traumatic haemorrhage) (Cioncoloni et al., 2012). We excluded stroke patients if: a) the primary diagnosis for admission was not stroke, or b) they required intensive care on the day of recruitment.

The clinical definition for stroke in this study was 'a clinical syndrome characterised by rapidly developing clinical symptoms and/or signs, focal, and at times global, loss of cerebral

function, with symptoms lasting more than 24 hours or leading to death, with no apparent cause other than a vascular one' (O'Donnell et al., 2010a, Hatano, 1976).

5.3.2 Barthel Index

The Barthel Index was designed to be a simple index of a patient's independence and is also used to measure the change in functional status during rehabilitation (Mahoney and Barthel, 1965). It is a scale that indicates the ability to perform activities of daily living. It has 10 items (tasks), with total scores ranging from 0 (worst mobility in activities of daily living) to 100 (full mobility in activities of daily living) (Quinn et al., 2011) and it has adequate clinimetric (quality of clinical measurements) properties in stroke rehabilitation (Cioncoloni et al., 2012). In the index, the 10 items have these scoring combinations: a) 0 and 5, b) 0, 5 and 10, or c) 0, 5, 10 and 15.

In this analysis, we treated the scores as numerical as previously done (Nakao et al., 2010, Shah et al., 1989, Harrison et al., 2013), although treating Barthel Index scores as categories: a) score > 80, independent; b) score < 40, very dependent; or a) score of 95/100, cut-off point for excellent outcome; and b) score of 75/100, a poor outcome, is also acceptable (Harrison et al., 2013).

5.3.3 Consent and interview

Informed consent was obtained from the patient, or if the patient was unable to give informed consent, from a proxy, usually the next of kin. Following that, one researcher (KIM) conducted all Barthel Index assessments.

Patients were deemed able to participate in the study if they had a Glasgow Coma Scale score of ≥ 15 and were orientated to time, place and person. Additionally, patients had to have been assessed as fit and reliable to participate in the interview by a neurologist.

Eliciting information from a proxy, usually the next of kin, in stroke studies has been proven to be accurate (Bruno et al., 2010) and has been used in a large multi-centre stroke study: the INTERSTROKE study (O'Donnell et al., 2010b).

Before discharge, Barthel Index scores were elicited in the ward by visual inspection and face-to-face interview. There are 10 items in the Barthel Index and it was administered by interview by KIM. Subsequently, the data were coded based on scores of 0, 5, 10 or 15 for each item. KIM then entered the recorded Barthel Index scores into a database.

The total 10 items yielded the overall Barthel Index score. The scale of the index ranges from 0 (worst mobility in activities of daily living) to 100 (full mobility in activities of daily living). The items in the Barthel Index address a patient's ability in feeding, bathing, grooming, dressing, bowel and bladder control, toileting, chair transfer, ambulation and stair climbing (Harrison et al., 2013). A full list of the items is included in the appendices (Appendix F (Malay version) and Appendix G (English version)).

For patients who had been discharged, commonly 1 and 3 months thereafter, we called the patients or the proxy for a telephone assisted interview. On these two measurement occasions, KIM conducted the telephone interview.

Using telephone interviews for Barthel Index assessment has many advantages: a) it is reliable and valid (Della Pietra et al., 2011, Go, 2008), and b) it is valid for proxies—such as carers—and lay persons (Korner-Bitensky and Wood-Dauphinee, 1995, Hoffmann et al., 2010, Janssen et al., 2010). This is especially true when stroke patients are burdened with cognitive impairment (up to 60% if they have intracerebral haemorrhage) (Tveiten et al., 2014) and dementia (Go, 2008), where the only reliable information is from their carers.

5.3.4 Measurement occasions and covariates

We conducted the interviews and constructed the Barthel Index at 3 time points: a) at baseline, b) at 1-month post-baseline and c) at 3 months post-baseline.

At baseline, most of the patients in this study were assessed after day 3 of admission (usually the day the acute stroke occurs). This was done because any Barthel Index assessment performed earlier than 3 days post-acute stroke has limitations (Kwakkel et al., 2010, Quinn et al., 2011).

We chose 3 months as the maximum length of follow-up because it has been reported that the greatest recovery from neurological deficits after stroke occurs during the first 3 months or 10 weeks (Van Peppen et al., 2004, Kwakkel and Kollen, 2013).

The information on the covariates sex, age and stroke subtypes were abstracted from the medical records. The other covariates such as education, income and occupation were dropped from the analysis because the information was not available in the medical records (because it could only be collected directly from patients, and many were too ill to interview for this information).

5.3.5 Statistical analysis

Of 108 stroke patients recruited, 10 patients were uncontactable during follow-up. We described the patients' characteristics and Barthel Index score at baseline and at 1 and 3 months post-baseline were described using mean (SD) and frequency (%) where appropriate for the remaining 98 patients.

To compare if the baseline Barthel Index scores between patients who were alive throughout the study and those died within the 3-month period followed the same continuous

distribution, the cumulative probabilities for each subject were plotted and compared using the Kolmogorov–Smirnov test ('ks.test' function in R software) (R Development Core Team, 2013).

The relationship between the Barthel Index scores at each measurement occasion— at baseline, 1 and 3 months post-baseline (treated as a dummy variable)—and other covariates: age, sex (male or female) and stroke subtype (haemorrhagic stroke or ischaemic stroke) were examined individually using simple linear regression.

Next, a random intercept model—a type of linear mixed effect ('lme') model— considered the stroke patient as the random component and analysed it using the 'nlme' package in R software (Pinheiro et al., 2015). The regression parameters were estimated using maximum likelihood estimates. First, we fitted a model containing only the main effect variables. Next, we tested the addition of the 2-way interaction terms in the model. All comparisons between models were performed using likelihood ratio (LR) tests.

A histogram of the distribution of residuals was plotted to check the assumption of normality, as was a plot of the standardised residuals against the fitted values for homogeneity of variance. This latter plot was used for model checking.

5.3.6 Ethical approval

We received ethical approval from the Medical and Research Ethics Committee (MREC), Ministry of Health, Malaysia (NMRR-12-471-12139), the USM Human Research Ethics Committee (HREC), Universiti Sains Malaysia (JEPeM [242.4.(1.4)]) and the Lancaster University Ethics Committee.

5.4 Results

Data from 98 stroke patients, with a mean age of 60.7 years, were eligible for analysis. Of the patients who were eligible and for whom a Barthel Index had been constructed, 34.7% (34/98) were male and 65.3% (64/98) were female. Female patients were slightly older than male patients; the mean age of the female patients was 61.2 years; that of the male patients was 59.8 years. Ischaemic stroke patients comprised of 73.7% (70/95) of all strokes patients in this study. The proportion of ischaemic stroke was bigger than the haemorrhagic stroke because ischaemic stroke is more prevalent (common) than haemorrhagic stroke.

Table 5-1 shows that during the entire follow-up period, 13 patients died and two (2) were lost to follow-up. Of those who died, 10 did so between baseline and 1 month post-baseline and 3 patients died between 1 and 3 months post-baseline. Because the number of deaths was small, this will limit the power of the subsequent analysis.

In Table 5-2, we show that the overall Barthel Index mean score at baseline was 35.1 (SD = 39.4). The mean score then increased to 64.4 (SD = 39.5) at 1 month and to 71.5 (SD = 38.9) at 3 months' post-baseline. The Barthel Index score was lower at baseline in females than in males (mean = 31.6 vs. 41.8), but at 1-month post-baseline, the scores for females and males were almost similar (mean = 63.9 vs. 65.6). The scores for both sexes improved at 3 months' post-baseline (with a greater improvement in the male patients). The baseline Barthel Index score for haemorrhagic stroke patients was lower than that for ischaemic stroke patients (mean = 15.0 vs. 41.9), but then it increased more than 4-fold between baseline and 1-month post-baseline (mean = 66.3 vs. 63.0). At 3 months' post-baseline, the Barthel Index scores were almost similar (mean = 72.0 vs. 70.3). The baseline Barthel Index score for those who died within the first month (n=10) was 4.0 (SD = 8.43), and those who died within the three months (n=13) was 8.46 (SD = 19.9).

Our analyses also show that for patients who survived, the Barthel Index score increased modestly from baseline until 3 months post-baseline. The mean Barthel Index score at baseline of those who survived up to 1 month post-baseline ($n = 10$) was 4.0 (SD = 8.4) and the mean Barthel Index score at baseline of those who survived up to 3 months post-baseline ($n = 13$) was 8.5 (SD = 19.9).

Figure 5-1 shows that more than 40% of patients had baseline scores of 0. The cumulative probability plots show that the distribution of the baseline Barthel Index scores ($n = 98$) was marginally different from the distribution of baseline scores of patients who died during the 3-month follow-up ($n = 13$) (Kolmogorov–Smirnov test $p = 0.0484$).

Table 5-3 shows the univariable linear regression parameters. At all measurement occasions, age was inversely related to Barthel Index score. The mean Barthel Index scores for stroke subtype and sex were not significant ($p \geq 0.05$) at every measurement occasion, except for haemorrhagic stroke, which had a significantly lower Barthel Index score at baseline.

In Table 5-4, the regression parameters produced by the *lme* models, the log-likelihood values and the p-values are shown. It shows that the best predictors for change in the Barthel Index over the 3-month interval come from model 4. The predictors are 1) measurement occasion (baseline and 1 month and 3 months post-baseline), 2) age (centred at 60.7 years) and 3) stroke subtype (haemorrhagic stroke vs. ischaemic stroke). Model 4 shows that at baseline the Barthel Index was 39.6, at 1 month after discharge 66.9 and at 3 months after discharge 73.2. Age has inverse relationship with the Barthel Index: a one-year increase in age, the Barthel Index reduces by 0.94 unit, indicative poorer functional health status with advancing age. Patients with haemorrhagic stroke (HS) has poorer Barthel Index score (16.8 unit lower) than patients with ischaemic stroke.

Table 5-5 also shows the results of fitting model 4 (covariate measurement occasion + age + stroke subtype) in Table 5-4 with the products of 2-way interaction: a) age × measurement occasion, and b) age × stroke subtype. Based on the LR test, both the interaction terms (model 5 and model 6) failed to significantly improve model 4 ($p = 0.123$ and 0.535 , respectively).

Table 5-1 Patient characteristics and distribution of patients at baseline, one month and three months' post baseline

Variables		n	Male	Female	All
Stroke subtype	Ischaemic Stroke	70	27 (81.8)	43 (69.4)	70 (73.7)
	Haemorrhagic Stroke	25	6 (18.2)	19 (30.6)	25 (26.3)
Mean Age (SD)	Years	98	59.8 (12.6)	61.2 (14.2)	60.7 (13.6)
At Baseline	Alive	98	34 (100.0)	64 (100.0)	98 (100)
Baseline to 1 month	Alive	87	28 (82.3)	59 (93.7)	87 (89.7)
	Dead	10	6 (17.7)	4 (6.3)	10 (10.3)
Baseline to 3 months	Alive	83	27 (79.4)	56 (90.3)	83 (86.5)
	Dead	13	7 (20.6)	6 (9.7)	13 (13.5)

Table 5-2 The Barthel Index score at baseline, 1-month post-baseline and 3 months' post-baseline by sex, major stroke subtype, and vital status

Patients	n	At baseline	n	At 1 month	n	At 3 months
		Mean (SD)		Mean (SD)		Mean (SD)
All	98	35.1 (39.4)	86	64.4 (39.5)	82	71.5 (38.9)
Male	34	41.8 (42.4)	27	65.6 (41.3)	27	78.0 (34.6)
Female	64	31.6 (37.5)	59	63.9 (39.0)	55	68.3 (40.8)
Ischaemic stroke	70	41.9 (40.7)	61	63.0 (40.6)	57	70.3 (39.3)
Haemorrhagic Stroke	25	15.0 (27.2)	23	66.3 (38.1)	23	72.0 (39.6)

Figure 5-1 shows that more than 40% of patients had baseline score equalled zero. The cumulative probability plots show the distribution of the baseline Barthel Index score ($n=98$) is marginally different from the distribution of baseline scores in patients who died during the three months' follow-up ($n=13$) (p -value for Kolmogorov-Smirnov test = 0.0484).

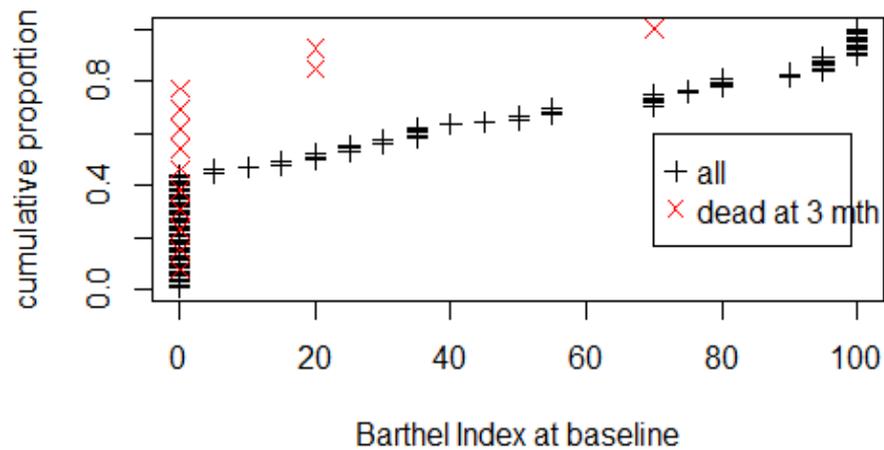


Figure 5-1 The cumulative proportions and the baseline Barthel Index scores for all patients (n=98) and for patients who died after baseline measurement up to 3 months post baseline (n=13). The cumulative proportions for Barthel Index were compared using the two-sample Kolmogorov-Smirnov test (2-sided test, $D = 0.4027$, $p\text{-value} = 0.0484$)

5.5 Discussion

This study has shown that overall, the crude (unadjusted) Barthel Index scores increased from baseline (mean = 35.1) to 1 month (mean = 64.4) and to 3 months post-baseline (mean = 71.5), implying that recovery had taken place. In particular, recovery took place between baseline and 1 month post-baseline (83.5% increase in the mean Barthel Index score). The longitudinal trend of the Barthel Index score is predicted by duration after stroke, age and stroke subtype. Our analyses show that increased age is associated with decreased Barthel Index score and that patients with haemorrhagic stroke have significantly lower Barthel Index scores than those with ischaemic stroke.

In this study, the crude mean Barthel Index score increased by about 83.5% (from 35.1 to 64.4) and 104% (35.1 to 71.5) from baseline to 1 month and baseline to 3 months, respectively, suggesting that substantial recovery had taken place during these periods, especially between baseline and 1 month. Early recovery has been shown to be mediated

through spontaneous neurological processes (Tveiten et al., 2014). Other factors such as socioeconomic processes can also influence recovery but this is more towards long-term recovery (Grube et al., 2012). Our result is consistent with previous studies reporting that the mean Barthel Index score in stroke patients improves over time (Rachpukdee et al., 2013, Kong and Lee, 2014, Hebel et al., 2014). The period during which the greatest recovery occurs is still in question but it has been shown that most recovery occurs by 3 months after stroke (Kong and Lee, 2014). Between 6 months and 2 years post-acute stroke, the Barthel Index score does not improve considerably (von Vogelsang et al., 2015).

Our results show that age is negatively associated with Barthel Index score during the first 3 months following a stroke. The inverse role of age on the Barthel Index can be due to various mechanism which can be grouped into a) selective survival and/or cohort effect, b) physiologic age-related phenomenon, and c) increasing level of comorbidity due to aging (Volpato et al., 2001, Al-Saeed et al., 2016). It is likely that with increasing age, the body becomes weaker, making recovery slower, brain tissue is damaged and the protective effect of the endothelium and astrocytes in the brain is dysregulated (Sohrabji et al., 2013) and a consequent negative effect on sensory-motor recovery (Coelho and Giraldo-Guimaraes, 2014). Our finding provides further evidence that age is a consistent predictor of stroke outcome (van Almenkerk et al., 2013), more precisely, it is a negative predictor of functionality, as shown in other studies (Gunathilake et al., 2014, Chindapasirt et al., 2013, Tveiten et al., 2014, Knoflach et al., 2012). It has been suggested that negative effect on the functionality could last up to 10 years after the acute stroke (Bhalla et al., 2013). To understand further the role of age, a few strategies can be used in future studies: a) model age of onset as the covariate, b) perform follow-up studies and c) use index of comorbidity such as Charlson Index (Al-Saeed et al., 2016, Volpato et al., 2001).

In this study, our results indicate that patients with haemorrhagic stroke have lower Barthel Index scores during follow-up compared to patients with ischaemic stroke. Stroke of the haemorrhagic subtype is more severe due to more extensive brain injury as a result of: a) the accumulation of blood and b) brain ischaemia following the haemorrhage (Frontera et al., 2015). In haemorrhagic stroke patients, more complications during in-hospital rehabilitation service (Di Carlo et al., 2006), a higher prevalence mental disorders (Li and Chen, 2014) and poorer cognitive status (Tveiten et al., 2014) render these patients more vulnerable to slower recovery. This result confirms the findings that stroke subtype plays important roles in stroke outcome (Paci et al., 2011, Di Carlo et al., 2006).

In our study, we found that that sex has no significant relationship with Barthel Index score. Studies on age and its effect on functionality report conflicting results but one review concluded that women have poorer functional outcomes after stroke (Reeves et al., 2008).

The study has its strengths and limitations. The strengths include the follow-up of up to 3 months after stroke—an appropriate period for showing changes in functional status—as it has been shown that most motor recovery in stroke is completed within 10 weeks of stroke (Kwakkel and Kollen, 2013). We employed the mixed effect model in the analysis instead of ordinary least square to the longitudinal data which can severely over- or underestimate the variance of the regression parameters, and analysis of variance methods are not feasible for longitudinal data analysis (Diggle et al., 2002). The mixed effect model takes into account the correlation and dependences of the observation and is an efficient method for estimation (Rabe-Hesketh and Skrondal, 2012).

We acknowledge three major limitations. First, there were only 3 measurements over the 3-month period. To better quantify the changes in the Barthel Index score after stroke, more than three (3) measurements are probably necessary during the follow-up because improvements are extended until 12 months post-stroke (Kong and Lee, 2014). Second, the

sample size should be larger to accommodate more covariates in the model. Third, other covariates such as psychological and employment variables were not readily available in the medical records (Rachpukdee et al., 2013, Haghgoo et al., 2013) . Lastly, the floor effect (tendency for minimum scores) and the ceiling effect (tendency for maximum scores) reduce the ability of the Barthel Index to detect changes in function during early and late phases of recovery (Ellul et al., 1998).

Table 5-3 The estimated regression parameters from univariate linear regression. The outcomes are Barthel Index scores at baseline, 1-month post-baseline and 3 months' post-baseline.

Variables	Parameters	Baseline		1 month		3 months	
		Beta (SE)	p-value	Beta (SE)	p-value	Beta (SE)	p-value
Age (n=98)	Constant	78.52 (17.79)	<0.001	110.22 (19.02)	<0.001	118.03 (19.14)	<0.001
	Age	-0.72 (0.29)	0.014	-0.76 (0.31)	0.016	-0.77 (0.31)	0.015
Stroke types (n=95)	Constant	41.93 (4.50)	<0.001	62.95 (5.12)	<0.001	70.26 (5.22)	<0.001
	Haemorrhagic Stroke ^a	-26.93 (8.78)	0.003	3.35 (9.78)	0.733	1.69 (9.73)	0.862
Sex (n=98)	Constant	41.77 (6.73)	<0.001	65.56 (7.65)	<0.001	77.96 (7.48)	<0.001
	Female ^b	-10.20 (8.33)	0.224	-1.66 (9.23)	0.858	-9.69 (9.14)	0.292

^a Haemorrhagic stroke vs Ischaemic stroke (reference) ^b Female vs male (reference)

Table 5-4 Results from linear mixed models performed using maximum likelihood estimation. The covariates and the assigned model are shown in the first column. The regression parameters estimated from these main effect models are shown with the p-values and log-likelihood.

Covariates (model)	Variable	Beta	SE	p-val	LL ^a		Likelihood ratio	p-value ^b
Occasion ^c , n=98(1)	baseline	35.10	3.95	<0.001	-1289.9			
	1 month	62.44	4.09	<0.001				
	3 months	68.82	4.14	<0.001				
Occasion + Age, n=98 (2)	baseline	35.11	3.81	<0.001	-1285.0	vs Occasion (1)	9.68	0.002
	1 month	62.37	3.96	<0.001				
	3 months	68.77	4.00	<0.001				
Occasion + Stroke type ^d , n=95 (3)	Age	-0.80	0.25	0.002				
	baseline	36.86	4.56	<0.001	-1256.6	vs Occasion (1)	0.64	0.424
	1 month	64.21	4.69	<0.001				
	3 months	70.45	4.75	<0.001				
Occasion + Age + Stroke type, n=95 (4)	HS ^d	-7.66	8.20	0.353				
	Baseline	39.64	4.44	<0.001	-1250.7	vs Occasion (1)	10.05	0.007
	1 month	66.89	4.57	<0.001				
	3 months	73.18	4.62	<0.001				
	Age	-0.94	0.27	0.001				
	HS ^d	-16.77	8.22	0.044				

^a LL=log-likelihood, ^b p-value obtained from comparing the log-likelihood between the current model and the previous one using complete cases (n=95)

^c Occasion is the three measurement occasions treated as a dummy variable; baseline Barthel Index (BI), BI at 1-month post-baseline and BI at 3 months' post-baseline.

^d HS=Haemorrhagic stroke vs Ischaemic stroke (reference)

Table 5-5 Results from linear mixed models performed using maximum likelihood estimation. The covariates contained the 2-way interaction term from the best model in table 4 (model 4). The regression parameters estimated from these main effect models are shown with the p-values and log-likelihood.

Covariates with interaction (model no)	Variable	beta	SE	p-val	LL ^a		Likelihood ratio	p-value ^b
occasion ^c + age + age*occasion, n=98 (5)	baseline	35.11	3.83	<0.001	-1237.7	vs occasion + age + stroke type (4)	2.38	0.123
	1 month	62.31	3.97	<0.001				
	3 months	68.69	4.02	<0.001				
	Age	-0.72	0.28	0.013				
	1 month * Age	-0.13	0.25	0.606				
	3 month * Age	-0.15	0.25	0.537				
occasion ^c + age + stroke type + age*stroke type, n=95 (6)	Baseline	39.90	4.43	<0.001	-1250.4	vs occasion + age + stroke type (4)	0.38	0.535
	1 month	67.16	4.56	<0.001				
	3 months	73.46	4.61	<0.001				
	HS ^d	-14.16	8.93	0.116				
	Age	-1.03	0.30	0.001				
	Age * HS	0.51	0.71	0.472				

^a LL=log-likelihood, ^b p-value obtained from comparing the log-likelihood between the current model and the previous one using complete cases (n=95)

^c Occasion is the three measurement occasions treated as a dummy variable; baseline Barthel Index (BI), BI at 1-month post-baseline and BI at 3 months' post-baseline.

^d HS=Haemorrhagic stroke vs Ischaemic stroke (reference)

Our findings suggest the functional status of after stroke improves and the improvement is related with a few predictors. This improvement can be captured by instruments such as the Barthel Index. Unfortunately, it is a routine practice in Malaysia to monitor the functional status in most hospitals in Malaysia. Our view is that the change in the functional status should be monitored and instruments like Barthel Index should be instituted as a standard functional outcome assessment to quantify this change.

We used the telephone interview at 1 month and 3 months after the baseline measurement. The limitations of the telephone interview include incomplete information, higher level of missing data, greater difficulty in achieving rapport and the lack of visual cues (Carr and Worth, 2001) but these limitations were minimised here because the Barthel Index questionnaire was concise, short and simple. The advantages of the telephone interview were that; 1) it was time-saving, 2) cost-effective, and 3) flexible (Carr and Worth, 2001, Barriball et al., 1996).

The implication of this study results on the clinical practice is that rehabilitation for stroke patients should begin early—as early as 48 hours—because early rehabilitation improves functional outcome (Liu et al., 2014).

5.6 Conclusion

The overall 10-item Barthel Index score increases consistently from baseline to 1 month post-baseline and then to 3 months post-baseline. The score largely improves between the measurement at baseline and at 1-month post-baseline (68.7% improvement) as compared to the period between 1 and 3 months post-baseline (9.4% improvement). In addition to measurement occasion, age is negatively related with Barthel Index score and haemorrhagic stroke patients consistently score lower on the Barthel Index than patients with ischaemic stroke over the 3-month period.

5.7 Acknowledgements

The authors thank Dr Ozgur Asar for the technical assistance in using 'linear mixed effect' function in R, Dr Rose Izura (neurologist, Hospital Raja Perempuan Zainab II (HRPZ II), Kota Bharu, Kelantan, Malaysia) and staff-nurse Wardah (HRPZ II) for assisting in patient recruitment. We also thank the Universiti Sains Malaysia for funding this project (304/PPSP/61312028).

Chapter 6 Conclusion

The PhD work undertaken for this thesis was motivated by a general epidemiological research question: how can we understand the natural history of stroke in the state of Kelantan, Malaysia? Understanding the natural history of a disease requires quantification of the risk, prevalence and outcome of the disease, which is important for preventing and controlling the disease (Gordis, 2009).

For this PhD work, we conducted 4 projects to answer 4 specific research questions on: a) the risk factors for stroke, b) the prognostic factors for stroke fatality, c) the clinical presentation and prognostic effect of cerebral infarction and cerebral haemorrhage and d) the functional outcome over a 3-month period after acute stroke.

The four projects were conducted in the setting of two hospitals in Kelantan: a) Hospital Raja Perempuan Zainab II (HRPZ) and b) Hospital Universiti Sains Malaysia (HUSM). Both hospitals are the two biggest hospitals and most important tertiary hospitals in Kelantan. They are hospitals under the care of the Ministry of Health, Malaysia. HRPZ is a public hospital while HUSM is a university hospital.

To meet the goals of this research, I designed the studies, collected the data, interviewed the patients and used a range of statistical modelling methods for the data analyses.

I faced several limitations, including difficulty in obtaining good-quality primary data because of: a) the absence of a stroke registry (at the time of the study), b) the limited number of variables recorded before and at arrival at the emergency departments of both hospitals, which were subsequently stored by the records office, c) the absence of routine stroke assessment at the emergency departments of both hospitals and d) the use of hospital-based

patients rather than population-based patients as recommended by prominent stroke scientists (Feigin et al., 2014). These limitations impeded the generalisability of the results.

Of equal, if not more, importance to the methodological and analytical issues that stemmed from these studies is the experience I have gained: from working with a multidisciplinary team, interacting with stroke patients, their treating physicians and nurses to the hands-on application of epidemiological and statistical methods to the data collected.

In this chapter, I reflect upon the challenges, successes and future work associated with this thesis and the experience I gained through the undertaking of this work.

6.1 Limitations and strengths

The project was motivated by the pragmatic goal of understanding the risk factors, clinical presentations, survival and functional recovery from stroke in an Asian population. We chose the state of Kelantan, Malaysia, for our project. We collected data personally from the hospital records offices and from the patients themselves and later performed the recommended epidemiological and statistical modelling.

All data in our project was obtained from two major hospitals in Kelantan: a) HRPZ and b) HUSM. The data collection and analytical methods used in the study required me to: a) design a suitable epidemiological study, b) obtain data from the relevant medical records offices, c) extract individual patient data from the clinical records, d) interview stroke patients over a 3-month period to assess their functional outcome status and lastly, e) apply the appropriate statistical analysis to the data obtained.

The main aim of the thesis was to understand the natural history of stroke—an important component in quantifying the risk and outcome of the disease (Gordis, 2009)—in Kelantan. This course ranged from: a) identifying the risk factors for stroke, b) assessing on-

admission characteristics of stroke patients at the emergency department, c) identifying prognostic variables for stroke fatality during hospitalisation, d) comparing the prognostic effect between cerebral infarction and haemorrhagic stroke, to e) assessing functional status using the Barthel Index at discharge (baseline) and at 1 and 3 months thereafter.

The first major limitation was the scarcity of electronically recorded demographic, clinical and outcome variables of stroke patients in both hospitals. In practice, major hospitals in Malaysia use the International Classification of Disease version 10 (ICD-10) to code the diagnosis at discharge. They record basic baseline demographic variables, on-admission variables, and status at discharge.

HUSM uses an online system they developed to record these data. HRPZ, on the other hand, uses a system provided by the Ministry of Health, Malaysia, to record the variables. The use of extensive electronic medical records is not available to almost all hospitals in Malaysia—only 4 hospitals in Malaysia to date have been gazetted as emergency medical record (EMR)-compliant hospitals (Noraziani K et al., 2013). The limited uptake of EMR is probably due to EMR being considered ineffective and that it would escalate the cost of health care in Malaysia (Health Technology Assessment Unit, 2006).

The current practice in Malaysia is that doctors at the emergency department examine stroke patients brought in by relatives (brought by own transport or ambulance) or referred from other hospitals or clinics (brought by ambulance). In most cases in Malaysia, doctors are not present at the scene of an acute stroke. Thus, the care received before reaching a hospital (pre-hospital care) could be hampered by inadequate assessment, recording and patient stabilisation. In Malaysia, the ambulance services are manned by nursing staff and non-medical ambulance drivers (Hisamuddin et al., 2007), and the quality of service is limited by slow ambulance response times (Shah et al., 2008).

Only at the emergency department will the attending doctor obtain the full clinical history, perform a physical examination and obtain baseline blood samples. In major hospitals in Malaysia, a computed tomography (CT) scan investigation is performed after consulting the neurology team on patients brought to hospital with the suspicion of stroke. Unfortunately, except for the Glasgow Coma Scale (GCS), other stroke-specific assessments and stroke screening tools are not routinely used. These stroke assessment tools include the National Institute of Health Stroke Scale (NIHSS) (Seki et al., 2014, Nilanont et al., 2010, Dawodu and Danesi, 2008), the Scandinavian Stroke Scale (SSS) (Luvizutto et al., 2012, Christensen et al., 2009, Barber and Stott, 2004) and the Canadian Neurological Scale (Nilanont et al., 2010, Seki et al., 2014, MacKay et al., 2007, Brass and Kernan, 1989), which are sensitive and valid tools in the acute stroke setting. Assessment of stroke dependency such as with the modified Rankin scale (MRS) (Banks and Marotta, 2007, Balu, 2009, Cincura et al., 2009) and functional status such as with the Barthel Index (Quinn et al., Balu, 2009, Cincura et al., 2009) are also not included as standard stroke assessments in almost all hospital settings in Malaysia.

The second limitation is the possibility of informative censoring from the survival data in paper 2 and paper 3. In both papers, we analysed data for which the outcome of interest is the time-to-event data. In the data, the event was in-hospital fatality, which was defined as stroke death (fatality) due to all causes during the period of hospital admission at HUSM. In paper 2, stroke patients who were discharged alive from the HUSM were taken as censored observations. In paper 3, stroke patients who were discharged alive either by the treating doctors or at their own-risk (or discharge) were considered as censored observations. Such time-to-event study design has the bias that the censored observations could be informative, that is that the censoring process ($\delta = 0$) has significant dependence with time (T). If the censoring is informative, the numerical estimates from Kaplan-Meier and Cox regression methods are biased. We performed the Cox and reversed Cox regression (outcome was reversed) to assess if there are shared dependencies between important covariates and the

outcome (see Figure H.1 in Appendix H). Based on the consistent significant effect between the covariates, such as stroke subtypes and Glasgow Coma Scale shown in Table H.1, Table H.2, Table H.3 and Table H.3, we felt that the assumption of non-informative censoring was met. In Appendix H, we also describe the principle to handle informative censoring using two multiple imputation methods: a) Gamma imputation and b) Risk Score imputation.

The third limitation of this study is that all our study data was from 2 major hospitals serving the population of Kelantan and there were no data from smaller hospitals such as district hospitals. There are 8 district hospitals in the 11 districts in Kelantan, which act as the gatekeepers to primary care service in Kelantan. A portion of patients with acute stroke will receive treatment from these smaller hospitals, and when they are suspected of having had a stroke, these patients will immediately be referred to larger hospitals such as HRPZ and HUSM. The rest, consisting of a very small percentage of patients, either have very mild stroke that requires admission and subsequent discharge, or too severe a stroke, to which they will succumb or be brought back home at their own risk.

Using data from major hospitals such as HRPZ and HUSM limits inference of the results to stroke patients treated at other major hospitals in Malaysia. To be able to generalise the results to all stroke patients in the community, a population-based study with data from all hospitals in the community is required (Feigin et al., 2014, Feigin et al., 2003, Feigin et al., 2009, Burke and Venketasubramanian, 2006).

The fourth limitation is the small sample sizes for Papers 2, 3 and 4. This was unavoidable due to logistic factors and the limited availability of CT scan imaging in hospitals in Kelantan. For Paper II and Paper III, we collected data from HUSM because only HUSM, and not HRPZ, stores CT scan images in a picture archiving and communication system (PACS). The availability of this service permitted our radiologist colleague to review CT scan images to verify stroke diagnosis and stroke subtype. For Paper III, given the prospective nature of the

study and the fact that only KIM performed the interview—to minimise inter-rater bias—and follow-up, only 98 patients who were eligible for analysis were recruited in the space of the 1 year of follow-up.

The fifth limitation was that we worked with data collected at the earliest at the emergency department by the emergency medical team or the neurology team. There were no data from the scene where the acute stroke took place and the medical team recorded no data while transporting patients from the scene to the hospital. In addition, the data collected at the emergency department lacked many clinical parameters or variables. In Malaysia, most hospitals, including HUSM and HRPZ, do not have specific assessment tools, such as standard observational and clinical assessment tools, when examining patients with acute stroke. Many of the variables specific to stroke and are useful for developing prognostic and risk factor models are not collected at the earliest possible time. We would recommend that all hospitals in Malaysia, especially the general tertiary hospitals such as HUSM and HRPZ, have routine clinical assessments and tools and laboratory tests for assessing the symptoms, signs and severity of stroke. The establishment of a stroke registry is paramount to achieving this objective (Shigematsu et al., 2013, Sridharan et al., 2009, Burd et al., 2006, Sia et al., 2007, Shiber et al., 2010, Hong et al., 2013, Kita et al., 2009, Saposnik et al., 2008).

The sixth limitation is that we did not have the residential coordinates for the stroke patients in Paper I. Unlike developed countries such as the UK and Japan, residential postcodes are not yet available in Malaysia. The available postcodes in Malaysia only serve as indicators of the servicing post office. The best we could do was perform spatial analysis using the centroid coordinates of the sub-subdistricts ($n = 288$) as the patients' coordinates. Given this approach, it is possible that several patients would share the same coordinates, as these patients stay in the same sub-subdistricts, and there would only be 288 points (coordinates) eligible for spatial autocorrelation analysis.

The seventh limitation is that we did not have information on an acute stroke until the patient arrives at the emergency department. Patients are usually brought to the emergency department in a private vehicle or an ambulance. From the acute stroke until arrival, we presumed that active clinical intervention is given to the patient while in the vehicle. In the ambulance, paramedics provide basic life support to prolong and increase the survival of stroke patients. This is done using ambulance call. However, there has been no standard data recording to collect information on the timing of the stroke, where the acute stroke took place and the basic clinical and physical assessment and emergency treatment administered on the way to the hospital. Studies have shown that the ambulance response time in Kelantan is far slower than the international standard (Shah et al., 2008, Hisamuddin et al., 2007).

The first strength of this study is that we used the ICD-10—the standard diagnostic tool for epidemiology, health management and clinical purposes—for a working definition for stroke (World Health Organization, 2015). The use of ICD-10 diagnosis (a type of administrative diagnosis) makes our study comparable to other stroke studies elsewhere that use similar ICD-10 codes in their study design. The ICD-10 is an internationally accepted classification developed by the World Health Organisation (WHO). The ICD-10 provides a standard measure used by WHO member states. It has several limitations such as errors in the coding process (O'Malley et al., 2005), but studies have shown the benefit of using the ICD-10 in stroke studies, including its high predictive value (Olson et al., 2014), high detection rate (McCormick et al., 2015) and its ability to identify stroke and its risk factors (Kokotailo and Hill, 2005).

The second strength was that we designed all 4 parts of the project, from data collection to data analysis. In designing the study, we considered the local setting of interest, international importance and comparability. Working with data from studies we designed maximised the utilisation of data in our project. In the process of designing our project, we looked at the potential meaningful interpretation of the results and their clinical importance

specifically for local physicians, neurologists, neurosurgeons, radiologists and public health workers in the management and prevention of stroke.

The third strength was that we collected, cleaned and analysed the data ourselves. The use of primary data is paramount in any research (Verpoorte, 2012), while the use of existing data provides little or no hands-on experience in an epidemiologic study (Buring, 2008). Working with primary data enabled us to match our collected data with our study clinical definition and the inclusion and exclusion criteria. Data from Paper 1 were obtained first-hand from the medical records office. For Paper 2 and 3, we extracted data from each patient's medical notes. In Paper 4, we also extracted data from the medical notes; in addition, we personally interviewed patients on their progress using the valid and reliable Barthel Index from admission until 3 months after discharge.

The fourth strength is that in Paper 2 and Paper 3, we verified the ICD-10 diagnosis (the administrative diagnosis) using CT scan images stored in the HUSM PACS server. A CT scan is the investigation of choice to verify stroke and further subtype it (Kalantri and Kalantri, 2010, Falcone et al., 2013b, Runchey and McGee, 2010). We provided the ICD-10 diagnoses for all data used in Paper 2, 3 and 4 and our radiologist colleague reviewed and verified them against the radiological diagnosis based on the CT scan images. In cases where CT scan images were not available because the CT scan had been performed in the referring hospitals, we verified the ICD-10 diagnosis using the radiological diagnosis in the clinical notes that accompanied the patient to HUSM during the transfer or referral.

6.2 Experience gained from collaboration

The experience I gained from Paper I includes working with the Malaysian Centre for Geospatial Data Infrastructure (MaCGDI, <http://www.mygeoportal.gov.my/>). The centre provided a list of all registered villages in Kelantan with the corresponding geospatial

coordinates of the villages. The list was presented in the form of spatial data in shapefile format. To match the addresses of the stroke patients in our study with the addresses of the villages in the database, I used ArcGIS software (ESRI, 2010).

I also communicated with the Department of Statistics, Malaysia (<https://www.statistics.gov.my/>), the official statistics agency of the government of Malaysia. They are responsible for conducting a nationwide population survey every 10 years (a decennial survey). For this project, they provided us with spatial data containing area-level population data. The data also came in shapefile format readable by ArcGIS software. There were 3 levels to the area: a) district (*'jajahan'*), b) subdistrict (*'mukim'*) and c) sub-subdistrict (*'mukim kecil'*). Given its sensitivity, the spatial data consisted of: a) population number by area, age group, sex and race, and b) the unique identification number for each district, sub-district and sub-subdistrict. Other socioeconomic indicator data were provided to us because these variables might be important factors related to having stroke.

I also established a connection with each records office manager at HUSM and HRPZ. I developed an understanding of the electronic data registration for the patients at each hospital. As a university hospital, HUSM perhaps has a superior system. It uses its own 'personalised' electronic registration, but HRPZ, being a government hospital, uses a 'standard' system provided by the Ministry of Health, Malaysia. This helped me understand the limitation and quality of data managed by both records offices, and the connection made me aware of the limitations of the systems in both hospitals. While the quality of patient data based on the ICD-10 is good, the staff who code the diagnosis (coders) need rigorous training in the use of the ICD-10 to minimise bias (misclassification bias) (Kokotailo and Hill, 2005, World Health Organization, 2010, World Health Organization, O'Malley et al., 2005).

I also connected with a neurologist at HRPZ and neurosurgeons, a neurologist and a radiologist at HUSM. This allowed me to understand their workload and how data can help

them understand stroke and manage patients with stroke. The present study helped them recognise the importance of stroke as a public health problem in Malaysia (Feigin et al., 2003, Feigin et al., 2009), understand the role of stroke prevention and control to reduce its massive burden on health (Feigin et al., 2014, Krishnamurthi et al., 2014) and the need to establish a stroke registry, especially in the Asian population, where stroke data are very scarce (Venketasubramanian et al., 2015).

6.3 Further research

To understand stroke among the Asian population in Malaysia, a larger longitudinal study with data from the population or community is required. Such study is still lacking in the Asian countries (Feigin et al., 2009). The PEARL study was the first and the only stroke population-based incidence study in Malaysia. But the study was conducted in a much smaller scale and focused on the population in the South-Western District of the Penang Island, Malaysia which has different socio-cultural practices (Neelamegam et al., 2013). A larger population-based study that focuses on the population in the east coast of Peninsular Malaysia is needed. The prospective population-based study would be ideal to provide more accurate estimates of the incidence, improve our understanding of stroke determinants and burden and help with the development and monitoring of the effectiveness of stroke prevention and management and rehabilitation strategies (Feigin et al., 2014) . Such a study can be conducted in particular in two states that are in proximity to one another on the east coast of Peninsular Malaysia—Kelantan and Terengganu, which are particularly different from other states in Malaysia because their populations are predominantly of Malay ethnicity. The future population-based cohort study must include more stroke-specific variables and capitalise on many widely known stroke assessment tools. We will acquire data from all major hospitals and from the smaller—but equally important, being the first gatekeepers to health care—district hospitals in both states. The population-based study will provide information on stroke incidence and outcome

and their relationship with environmental, social, economic and societal variables. This is important for the control and prevention of stroke on the ground (primary and secondary prevention).

We propose a study that uses patients' residential locations as the preferred coordinates. This future prospective study should be able to collect the exact locations (coordinates) with two attributes: a) location of the residence, and b) location where a stroke occurs, for better spatial analysis. The availability of such data enables the identification of the presence of spatial correlation between: a) residential address and risk for stroke and its outcome, and b) the geographical location of stroke and its relationship with the outcome and quality of stroke care. Information and understanding from such projects will be valuable for establishing intervention programmes and allocating the appropriate resources for acute stroke care and stroke rehabilitation in the community.

6.4 Closing remarks

This thesis gave me the opportunity to learn important concepts, build networks with stroke clinicians and acquire the necessary skills for designing an epidemiological study, collecting data, managing sensitive data and modelling data to address a given research question. The studies in this PhD project have enhanced my awareness of the difficulty in conducting research in the community, managing and analysing data and the limitation of generalising the results to the intended population. I greatly appreciate the importance and the richness that stems from strong professional relationships across different disciplines. The main message I have learnt from completing this project is that, provided there is adequate understanding of the natural history of a disease, all necessary variables are satisfactorily collected, all participants have agreed upon the direction of the undertaken research and that the results are clearly communicated to the appropriate audiences, epidemiological and statistical

methods in disease modelling continue to play an ever-increasing role in supporting and directing decisions in public health. Through the relationships I have forged and the skills I have learned, I hope to be able to continue my contributions to this field.

Chapter 7 Appendices

Appendix A: Additional results for Chapter 2

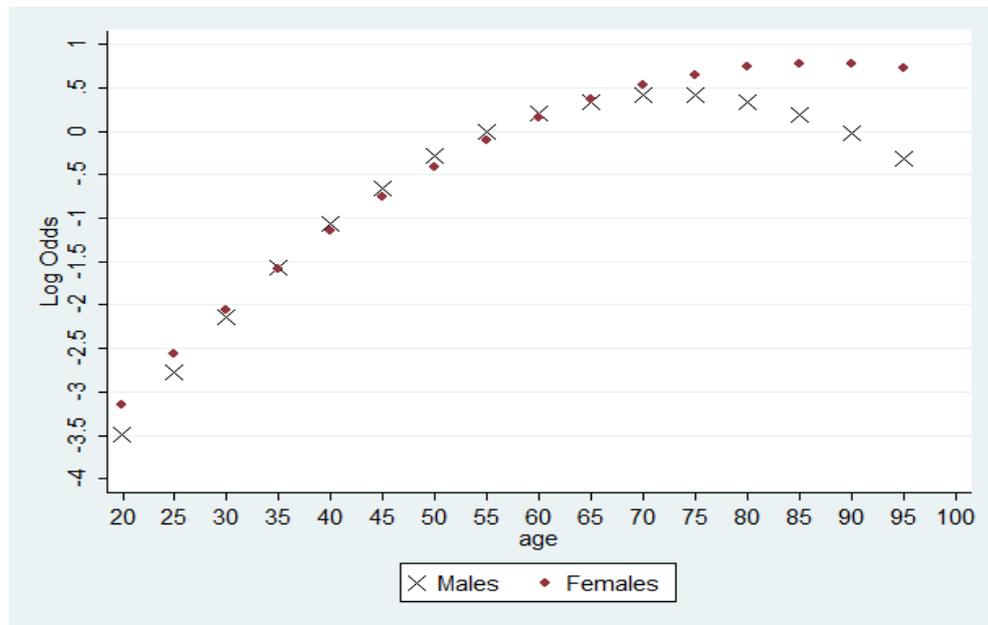


Figure A.1: The non-linear relationship between age (years) and the log odds for having stroke for males and females. The log odds for males increase from the age of 20 years, plateau between the age of 70 and 75 years, then on the decreasing trend from the age 75 years. The log odds for females starts to plateau at the later age than males (between 75 to 80 years) but never slow down until the age 95 years.

Table A.1: The classification table showing the distribution of patients with stroke (cases) and without stroke (controls) based on multivariable logistic model, the diagnostic accuracy and the predictive values of the logistic model (to predict stroke)

Classified (predicted) as stroke or no stroke based on model	Observed data		
	Had stroke	No stroke	Total
Stroke	979	657	1636
No stroke	390	1092	1482
Total	1369	1749	3118
Diagnostic accuracy and predictive values of the model to predict stroke			
Sensitivity	979/1369	Pr(+ D)	71.51%
Specificity	1092/1749	Pr(-~D)	62.44%
Positive predictive value	979/1636	Pr(D +)	59.84%
Negative predictive value	1092/1482	Pr(~D -)	73.68%
False + rate for true ~D	657/1749	Pr(+~D)	37.56%
False - rate for true D	390/1369	Pr(- D)	28.49%
Correctly classified	2071/3118		66.42%

Note: Classified as having stroke if predicted $\text{Pr}(D=\text{stroke}) \geq .5$

Table A.2: Pearson goodness-of-fit of the final multivariable logistic model for predicting stroke. The goodness of fit is assessed over the fitted values determined by the covariates in the model. For Pearson goodness-of-fit test, the difference in the observed and fitted values uses Pearson residuals

number of observations =	3118
number of covariate patterns =	2834
Pearson $\chi^2(2825)$ =	2883.99
Prob > χ^2 =	0.2153

Table A.3: Hosmer-Lemeshow goodness-of-fit (HL-gof) of the multivariable logistic model for predicting stroke. The goodness of fit is assessed over the fitted values determined by the covariates in the model but for HL-gof, groupings are created based on the estimated probabilities. The default in Stata is 10 groups ($g=10$) based on the deciles of the estimated risk. The HL-gof statistic, C , is obtained by calculating the Pearson chi-square statistics from $g \times 2$ table of observed and estimated expected frequencies of cases (strokes) and controls (no-strokes)

number of observations	3118
number of groups	10
Hosmer-Lemeshow $\chi^2(8)$	6.51
Prob > χ^2	0.5908

Table A.4: The deciles (the first column) of risk for having stroke based on the multivariable logistic model. Each row is collapsed according to the estimated probabilities for having stroke. The groupings were created based on the estimated probabilities. The default in Stata is to have 10 groups (g=10) based on the deciles of the estimated risk.

Group (decile)	Probability	Obs=1	Exp=1	Obs=0	Exp=0	Total
1	0.0733	16	13.5	296	298.5	312
2	0.2217	37	42.2	275	269.8	312
3	0.3625	94	92.3	218	219.7	312
4	0.4504	134	127.8	178	184.2	312
5	0.5132	148	150.4	163	160.6	311
6	0.5522	160	166.7	152	145.3	312
7	0.5809	187	176.5	125	135.5	312
8	0.6155	177	186.8	135	125.2	312
9	0.6565	193	197.9	119	114.1	312
10	0.8124	223	214.9	88	96.1	311
TOTAL		1369	1369	1749	1749	3118

*Obs = number of observed stroke cases; Exp=number of expected stroke cases

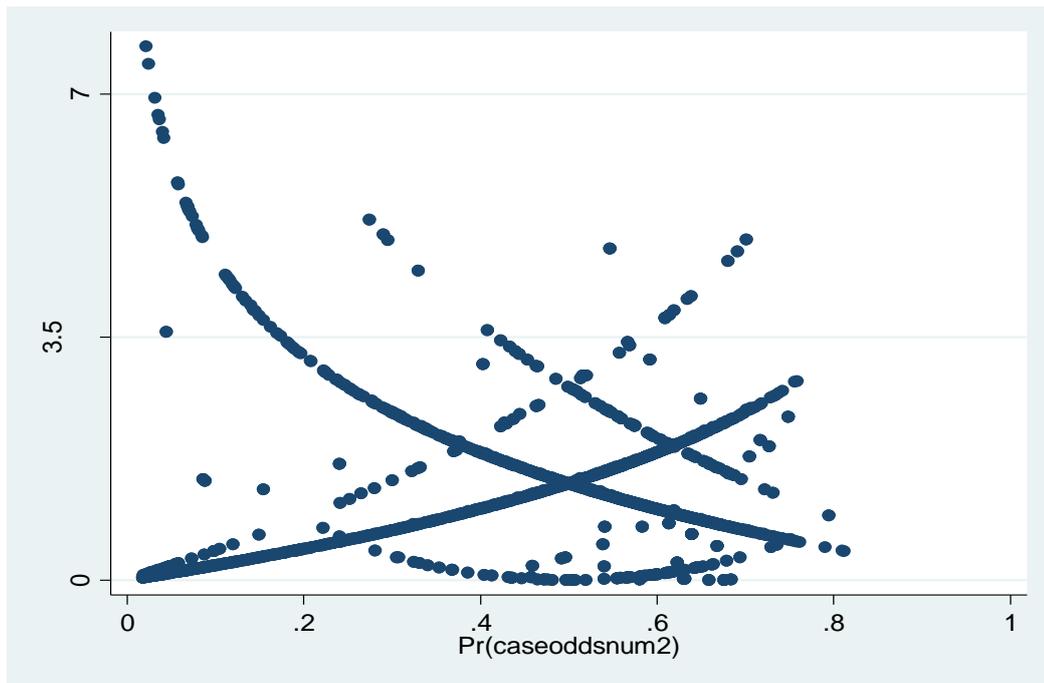


Figure A.2: The Hosmer-Lemeshow delta deviance (HL dD) against the probability of having stroke (Pr(caseoddsnum2)). The probability of having stroke was predicted using Stata.

Appendix B: Additional results for Chapter 3

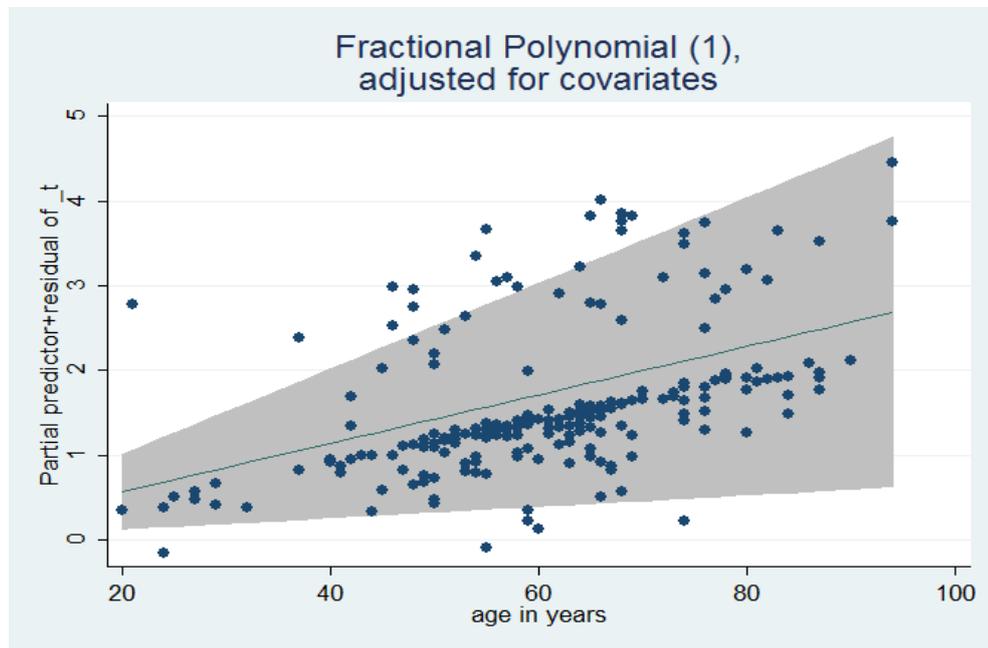


Figure B.1: The partial predicted values estimated from the fractional polynomial (fp) function during the Cox proportional hazard regression analysis in Stata are plotted against age (years) which is a numerical variable. Overall, the functional relationship between the variable 'age' and stroke case-fatality was linear based on the fitted straight line and the distribution of points in the plot. Hence, variable 'age' was modelled without any transformation during the Cox regression analysis.

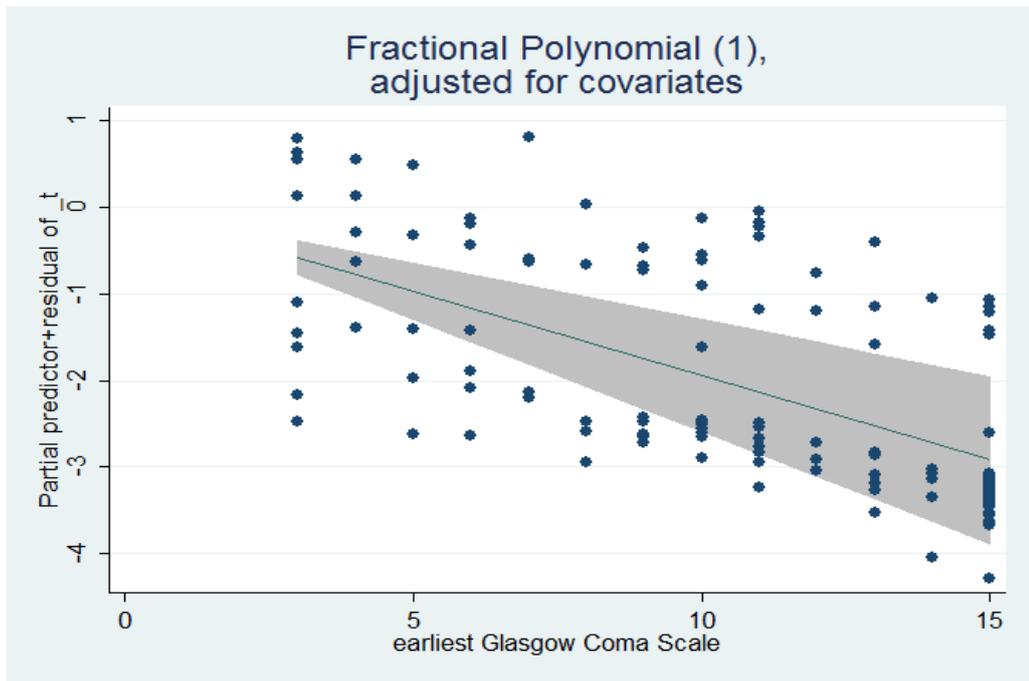


Figure B.2: The partial predicted values from the fractional polynomial (fp) function during the Cox proportional hazard regression analysis in Stata are plotted against the Glasgow Coma Scale (GCS) which is a numerical variable. Overall, the functional relationship between the Glasgow Coma Scale (on-admission) and stroke case-fatality was linear and was modelled without any transformation during the Cox regression analysis.

Table B.1: Results for the test of proportional hazard assumptions for the Cox proportional hazard regression model. The test is based on the Schoenfeld's residuals. It tests the overall model for the proportionality assumption (global test) and for each covariate in the Cox model (individual test). The test of non-zero slope was based on the Therneau and Grambsch idea with slightly different algorithm in Stata ('estat phtest' function).

Time scaling		rho	chi2	df	Prob>chi 2
Time	gcs	0.00275	0	1	0.9858
	age2	-0.06765	0.31	1	0.5763
	global test		0.31	2	0.8549
log(time)	gcs	0.18343	1.42	1	0.2341
	age2	-0.06443	0.28	1	0.5945
	globaltest		1.59	2	0.4516
Kaplan-Meier estimates	gcs	0.05928	0.15	1	0.7006
	age2	-0.08352	0.48	1	0.4902
	global test		0.58	2	0.7493
Rank	gcs	0.20012	1.69	1	0.1942
	age2	-0.08626	0.51	1	0.4761
	global test		2.03	2	0.3623

Hypothesis: null hypothesis of zero slope, which is equivalent to testing that the log hazard-ratio function is constant over time. Thus, rejection of the null hypothesis of a zero slope indicates deviation from the proportional-hazards assumption

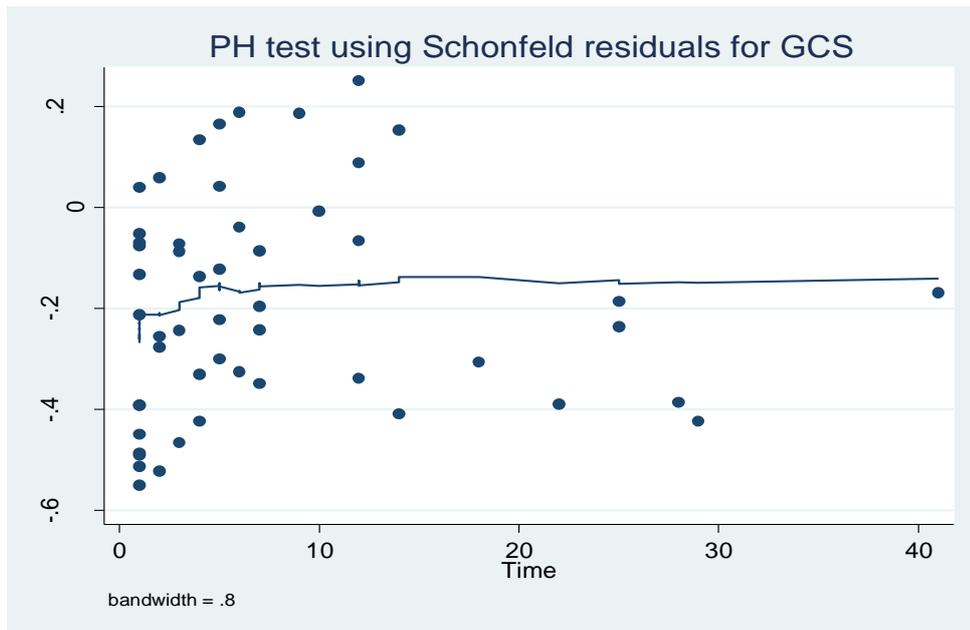


Figure B.3: The lowess smooth plot of Schoenfeld's residuals from the Glasgow Coma Scale (GCS) estimated in the Cox hazard regression model against the follow up time (days). The slight rise and the straight line afterwards line indicate non-significant departure from a slope of zero. In Stata, this plot is produced using the 'estat phtest' function and 'plot' argument.

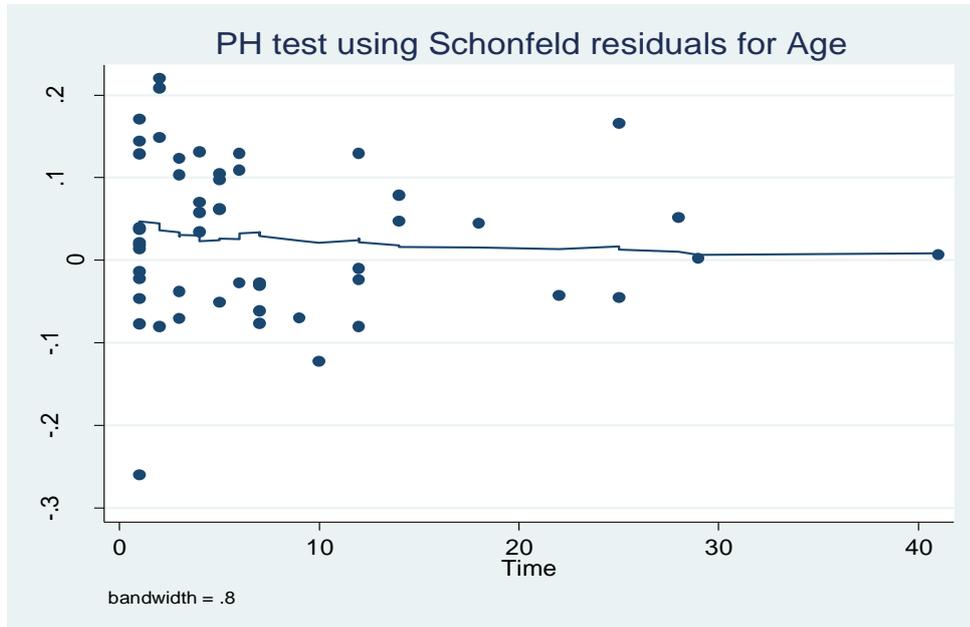


Figure B.4: The lowess smooth plot of Schoenfeld's residuals for variable 'age' generated from the Cox hazard regression model against the follow up time (days). The slight rise and the straight line afterwards line indicate non-significant departure from a slope of zero. In Stata, this plot is produced using the 'estat phtest' function and 'plot' argument.

Appendix C: Additional results for Chapter 4

Table C.1: The overall characteristics of in-hospital stroke recruited and analysed for survival analysis. There were 297 stroke patients (n=297) with 85 died during hospitalization (n=85)

Category	total	mean	Per subject		
			min	median	max
no. of subjects	297				
no. of records	297	1	1	1	1
(first) entry time		0	0	0	0
(final) exit time		6.521886	1	4	81
subjects with gap	0				
time on gap if gap	0				
time at risk	1937	6.521886	1	4	81
failures	85	0.2861953	0	0	1

Table C.2: The incidence for stroke fatality based on stroke subtypes. The 50% survival is also known as 'the median survival time' ^a

Stroke types	Time at risk	Incidence rate	Subjects (n)	Survival		
				25%	50%	75%
Cerebral Infarction	719	0.025	150	12	28	41
SAH	332	0.060	30	3	14	24
ICB	514	0.064	84	4	14	25
Other Hg	278	0.0468	28	1	9	.
All	1843	0.0456	292	6	15	29

*SAH = subarachnoid haemorrhage, ICB= intracerebral haemorrhage, Other Hg = other types of cerebral haemorrhage

$$^a \hat{t}_{50} = \min\{t: \hat{S}(t) \leq 0.50\}$$

Table C.3: The incidence (stroke fatality) based on stroke subtypes. The 50% survival is also known as 'the median survival time'

Stroke types	Time at risk	Incidence rate	Subjects (n)	Survival		
				25%	50%	75%
Cerebral infarction	719	0.0250348	150	12	28	41
Haemorrhagic stroke	1218	0.0550082	147	4	14	29
All	1937	0.0438823	297	6	15	29

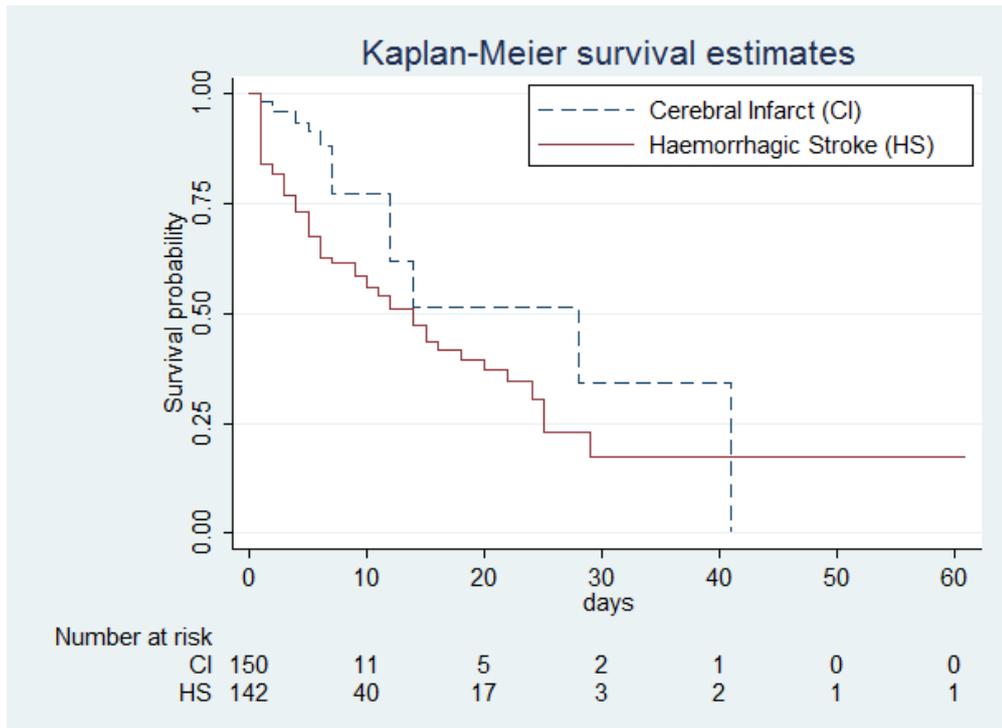


Figure C.1: The Kaplan-Meier survival curves for stroke patients with cerebral infarction (CI, n=150) and haemorrhagic stroke (HS, n=142). They show the survival probability after acute stroke against time (in days). In Stata, the function 'sts graph' plots the survival function.

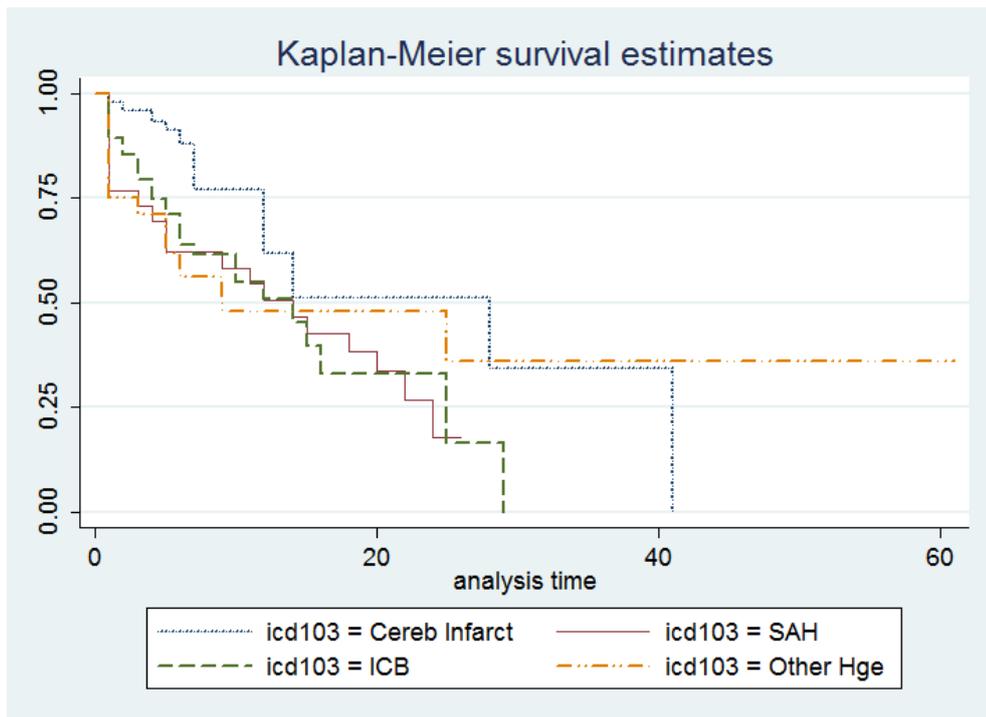


Figure C.2: The Kaplan-Meier survival curves for stroke patients suffering from different stroke subtypes: cerebral infarction (CI), subarachnoid haemorrhage (SAH), intra-cerebral haemorrhage (ICH) and Others (other types of haemorrhagic stroke) showing the survival probability against time (in days). In Stata, the function 'sts graph' plots the survival function.

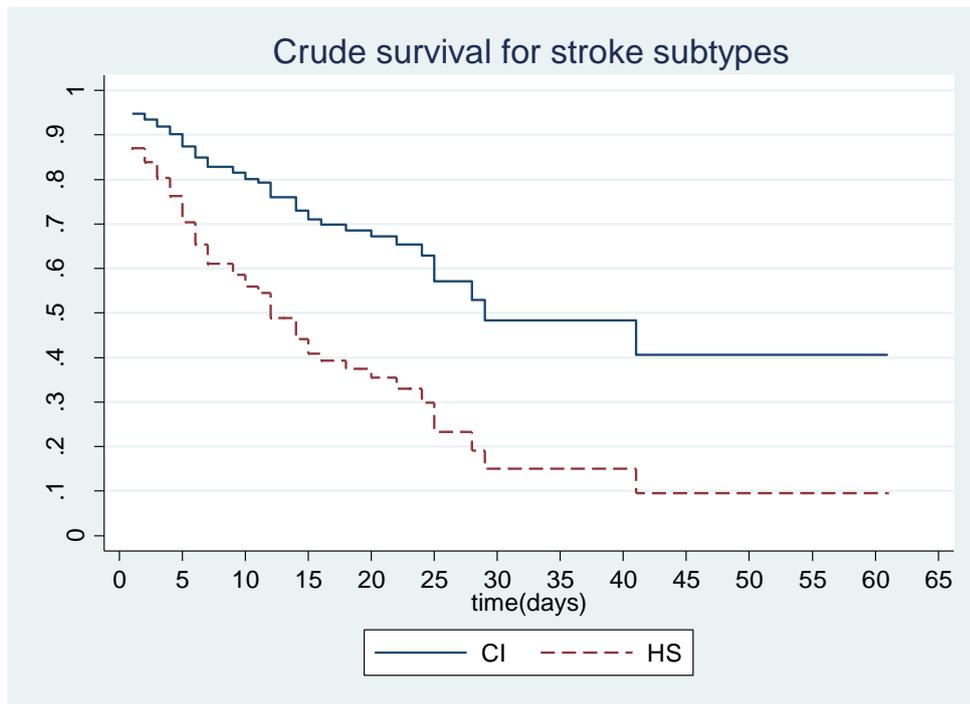


Figure C.3: The crude survival curves for stroke patients with cerebral infarction (CI) and haemorrhagic stroke (HS). It shows the survival probability after acute stroke against time (in days). In Stata, the function 'stcurve' plots the survival function after running the Cox proportional hazard regression using 'stcox' function.

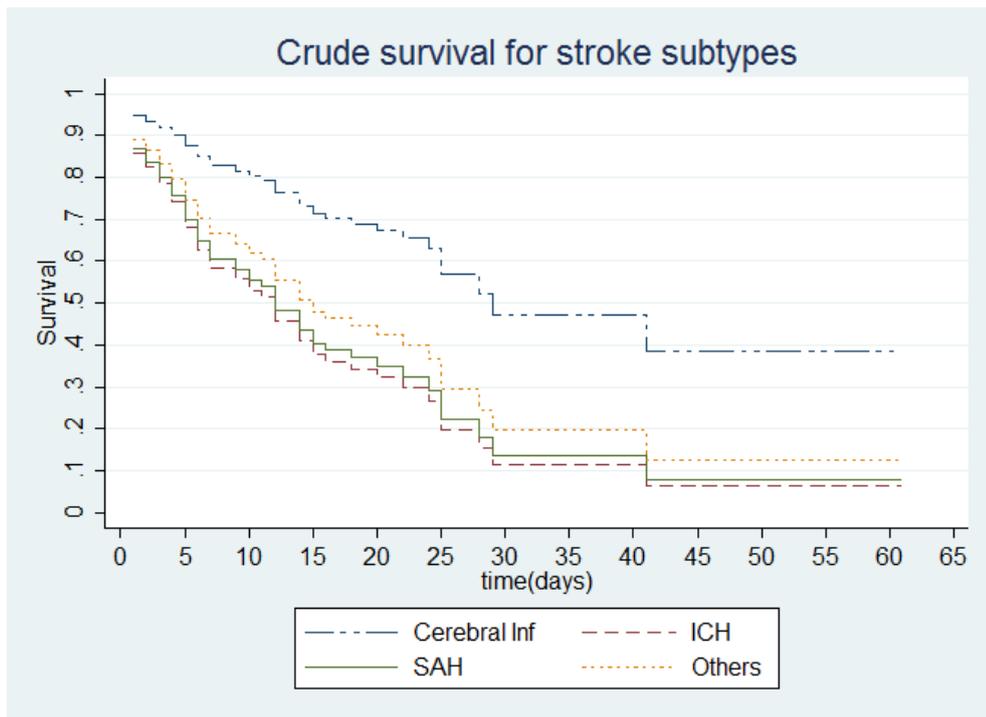


Figure C.4: The crude survival curves for stroke patients with different stroke subtypes: cerebral infarction (CI), subarachnoid haemorrhage (SAH), intra-cerebral haemorrhage (ICH) and Others (other types of haemorrhagic stroke) showing the survival probability against time (in days). In Stata, the function 'stcurve' plots the survival function after running the Cox proportional hazard regression using 'stcox' function.

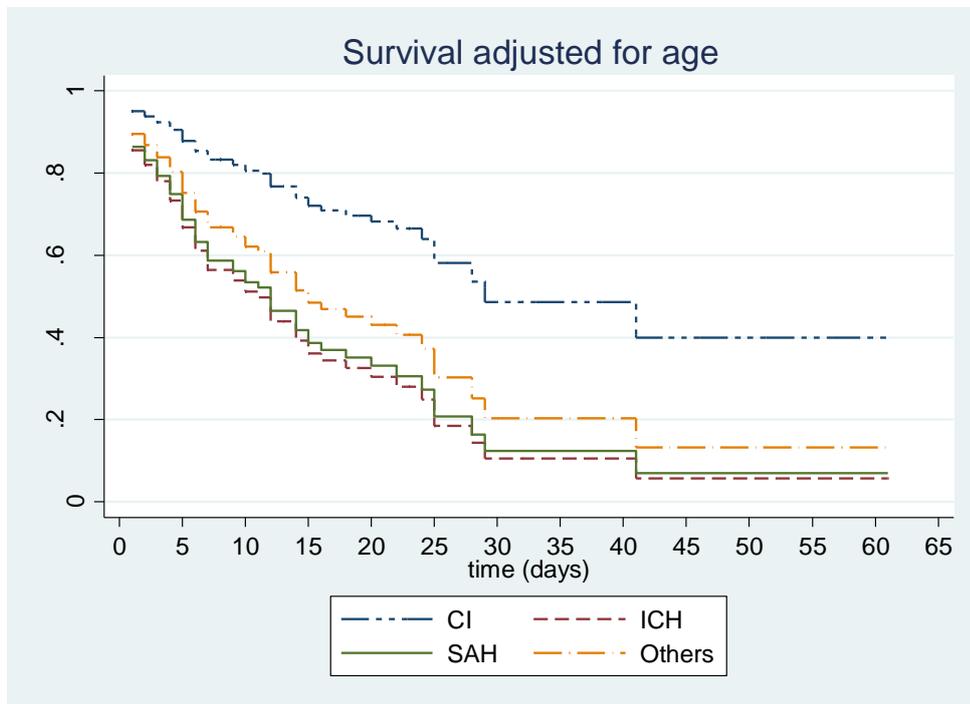


Figure C.5: The adjusted survival curves for stroke patients suffering from different stroke subtypes: cerebral infarction (CI), subarachnoid haemorrhage (SAH), intra-cerebral haemorrhage (ICH) and Others (other types of haemorrhagic stroke) showing the survival probability against time (in days) adjusting for age. In Stata, the function 'stcurve' plots the survival probability after running the Cox proportional hazard regression using 'stcox' function. In this model, the covariates are 'stroke subtypes' and 'age'.

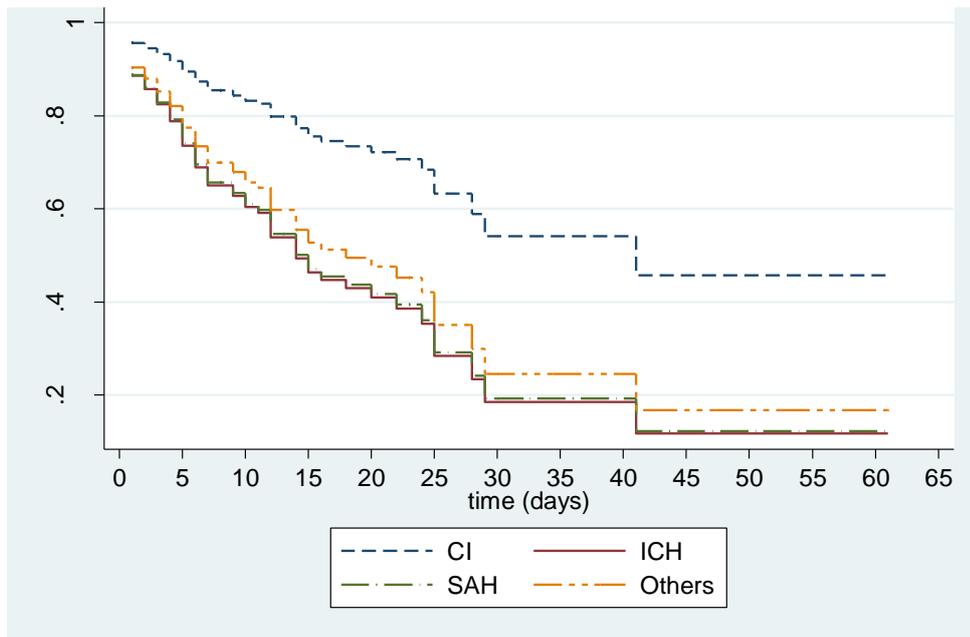


Figure C.6: The estimated survival probability for male stroke patients from a model with covariates 'stroke subtype' and 'sex'. The probability was estimated from the Cox proportional hazard regression using the 'stcox' function in Stata. In the model, there are stroke subtypes - cerebral infarction (CI), subarachnoid haemorrhage (SAH), intra-cerebral haemorrhage (ICH) and Others (other types of haemorrhagic stroke) - and sex.

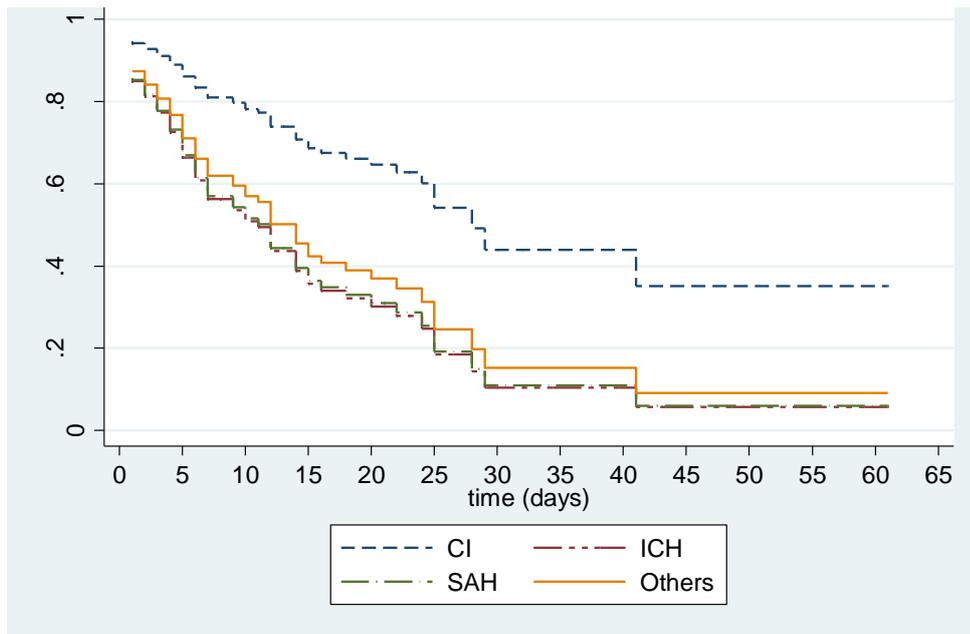


Figure C.7: The estimated survival probability for female stroke patients based on a model with covariates 'stroke subtype' and 'sex'. The probability was estimated from the Cox proportional hazard regression using the 'stcox' function in Stata. In the model, there are stroke subtypes - cerebral infarction (CI), subarachnoid haemorrhage (SAH), intra-cerebral haemorrhage (ICH) and Others (other types of haemorrhagic stroke) - and sex.

Appendix D: Additional results for Chapter 5

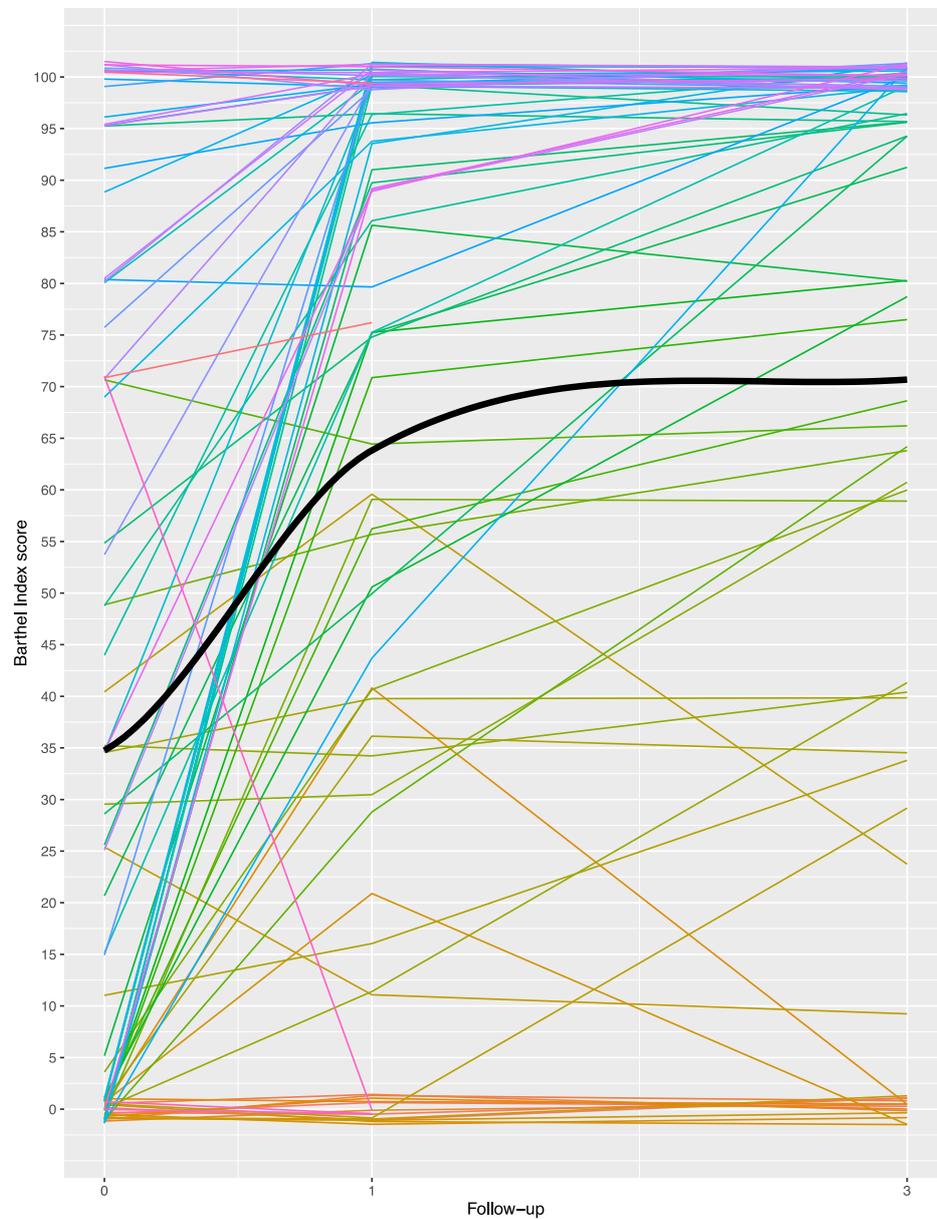


Figure D.1: A spaghetti plot showing the Barthel Index scores (y-axis) from baseline (zero), to one-month post baseline (1) and to three-month post baseline (3) (x-axis). The dark line is the smoothed conditional mean using the default argument in 'ggplot2'. No I score was taken on the second month. The scores were jittered to show more clearly the individual scores. There are two types of strokes: a) haemorrhagic stroke (bleeding inside the brain) in red lines and b) ischaemic stroke (inadequate blood flow to the brain) in green lines. A total of 98 (n=98) post acute stroke patients were recruited and 13 died between the baseline and the third month.

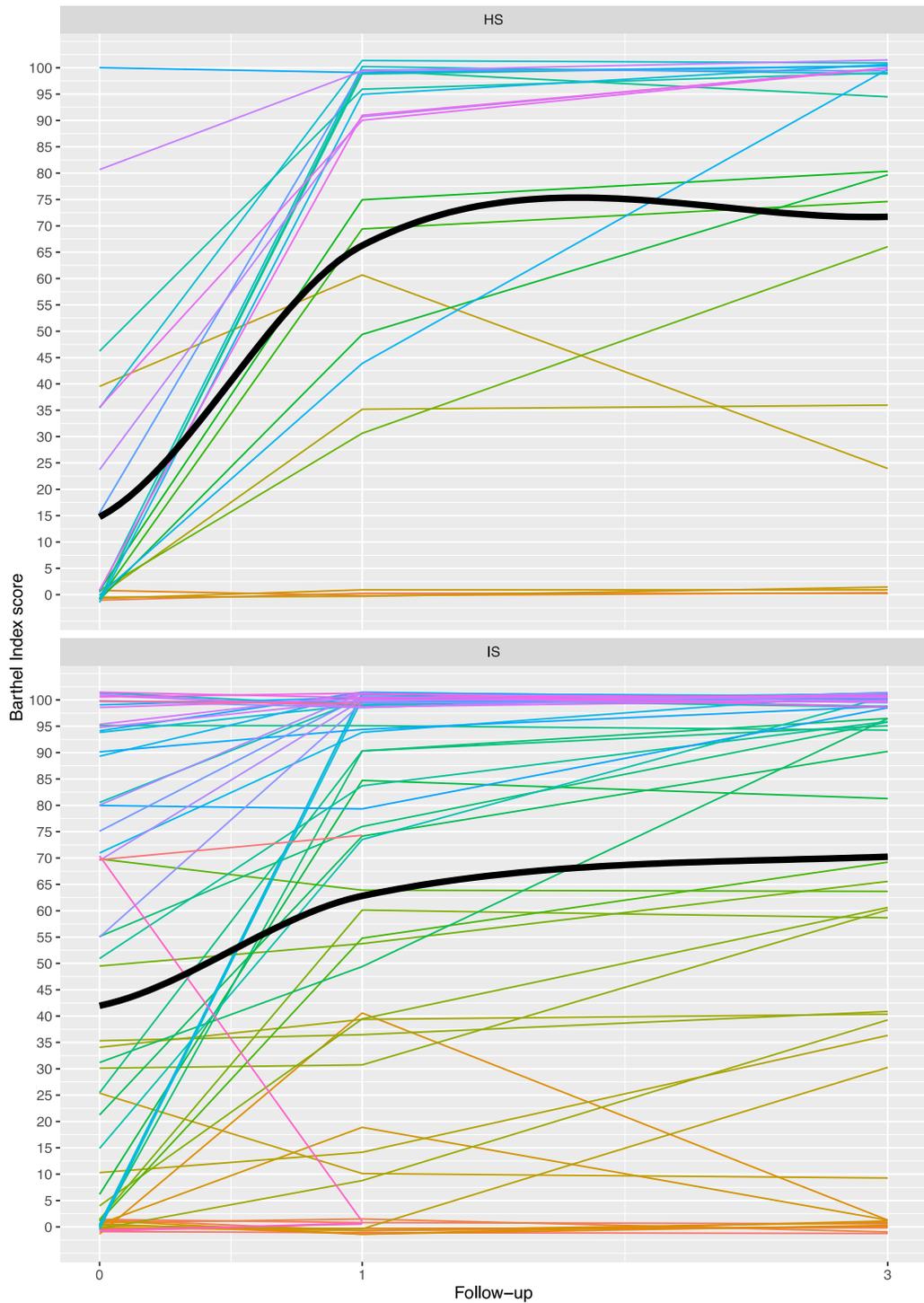


Figure D.2: Spaghetti plots showing the Barthel Index (BI) scores (y-axis) from baseline (zero), to one-month post baseline (1) and to three-month post baseline (3) (y-axis) for post acute stroke patients with haemorrhagic stroke (HS) and ischaemic stroke (IS). The dark lines are estimated from the smoothed conditional means using the default arguments in 'ggplot2'. No BI score was taken on the second month. The scores were jittered to show more clearly the individual scores.

Table D.1: The estimated regression parameters based on Generalized Estimating Equation (GEE) with the outcome of Barthel Index scores in post acute-stroke patients (n=98). Age was centred at 60.7 years. GEE using 'exchangeable' correlation structure

Covariates		Est	naïve SE	naïve Z	Robust SE	Robus z
Age	cons	53.65	3.52	15.23	3.53	15.22
	c.age	-0.83	0.26	-3.18	0.26	-3.25
Age + stroke	cons	58.04	4.21	13.80	4.15	13.99
	c.age	-0.97	0.28	-3.47	0.25	-3.83
	HS	-16.00	8.43	-1.90	7.26	-2.20
Age+stroke+sex	cons	57.34	6.26	9.16	6.54	8.77
	c.age	-0.97	0.28	-3.45	0.25	-3.85
	HS	-16.24	8.61	-1.89	7.50	-2.17
	female	1.15	7.65	0.15	7.90	0.15

Note: Est=estimated beta, SE=standard error, cons=constant, c.age=centred age, HS=Haemorrhagic stroke vs ischaemic stroke, female=female vs male

Table D.2: The estimated regression parameters based on random intercept linear mixed model (lme) with the outcome of Barthel Index scores in post acute-stroke patients (n=98). Age was centred at 60.7 years. The last column is the results for the log-likelihood estimates.

Covariates		Value	SE	DF	t-value	p-value	ll
Age	Cons	53.61	3.56	168.00	15.07	0.00	-1331.347
	c.age	-0.83	0.26	96.00	-3.15	0.00	
Age + stroke	Cons	57.70	4.38	163.00	13.18	0.00	-1295.29
	c.age	-0.95	0.30	90.00	-3.20	0.00	
	HS	-15.54	8.71	90.00	-1.78	0.08	
Age+stroke+sex	Cons	57.34	6.26	164.00	9.16	0.00	-1295.278
	c.age	-0.97	0.28	91.00	-3.45	0.00	
	HS	-16.24	8.61	91.00	-1.89	0.06	
	female	1.15	7.65	91.00	0.15	0.88	

Note: ll=log-likelihood, cons=constant, c.age=centred age, HS=Haemorrhagic stroke vs ischaemic stroke, female=female vs male

Model

The simplest model contains only measurement occasions:

$$Y_{ij} = \gamma_{00} + \alpha_i + U_{0j} + \epsilon_{ij}$$

Where Y_{ij} is the measurement for individual j at time i

U_{0j} is the random effect for individual j , α_i is the fixed effect of time i and ϵ_{ij} is a random error component specific to individual j at time i . The assumption are that U_{0j} are independent $N(0, \sigma^2)$, ϵ_{ij} are independent $N(0, \sigma_\epsilon^2)$. The fixed part, α_i , does not contain a constant term, but is based on M dummies for the M measurement occasions. The fixed part is written as:

$$\alpha_i = \sum_{h=1}^M \alpha_h d_{h,i}$$

Our model where the dependent variable is the Barthel Index score and the covariates are measurement occasions, age and stroke subtypes can be written as:

$$\text{score}_{ij} = a_1 d_{1ij} + a_2 d_{2ij} + a_3 d_{3ij} + \beta_{1j}(\text{age}_j) + \beta_{2j}(\text{HS}_j) + U_{0j} + \epsilon_{ij}$$

Which is a random intercept model.

Model comparison

When parameters of the statistical model are estimated by the maximum likelihood (ML) method, the estimation also provides the likelihood, which can be transformed into the deviance defined as minus twice the natural logarithm of the likelihood. The deviance can be regarded as a measure of lack of fit between model and data. The deviance is written as:

$$\begin{aligned} D &= 2[l(\text{model}_2) - l(\text{model}_1)] \\ &= 2\text{Log} - \text{likelihood}(\text{model}_2) - 2\text{Log} - \text{likelihood}(\text{model}_1) \end{aligned}$$

Which under certain conditions approximately follows a chi-square distribution with $t-r$ degrees of freedom.

Residuals

The level-2 residual can be predicted by the posterior means:

$$\check{U}_{0j} = E(U_{0j}|Y, X, \theta)$$

And the level-1 residuals are:

$$\hat{\varepsilon}_{ij} = \varepsilon_{ij} - \check{U}_{0j}$$

Model Diagnostics

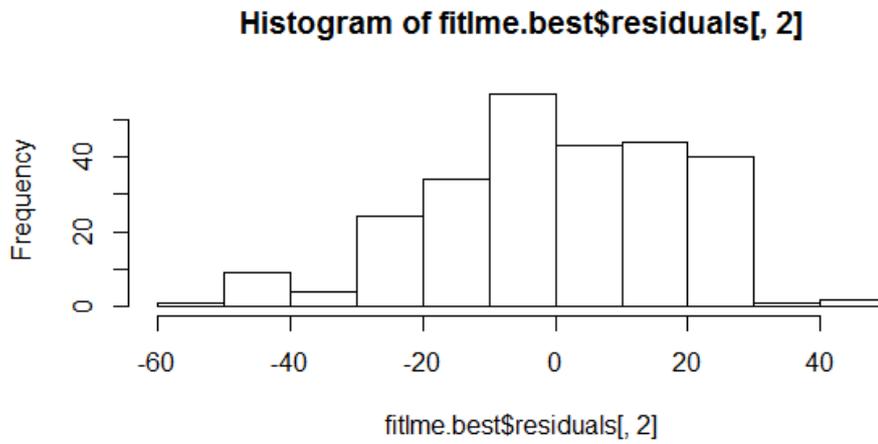


Figure D.1: Histogram of residuals based on the final mixed model with covariates 'time' (baseline, 1 month and 3 months post baseline), 'age (centred)' and 'stroke subtypes' (HS vs IS). The random effect was subjects (variable id). The residuals were approximately normally distributed.

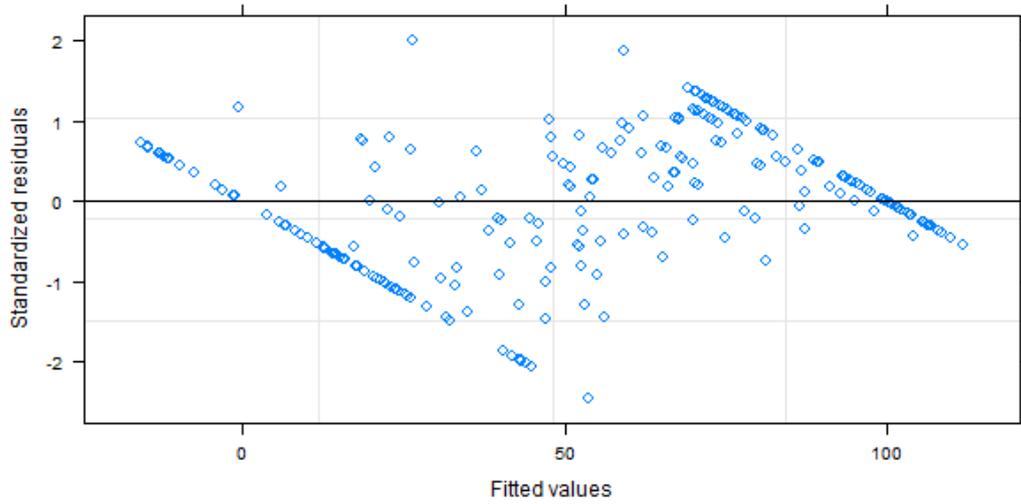


Figure D.2: Scatterplot of standardized residuals for the mixed model with covariates ‘time of follow up’, ‘age (centred)’ and ‘stroke subtypes’ (HS vs IS). The random effect was subjects (variable ‘id’). The Barthel Index scores range from zero (0) to 100. There were 18 out of 259 measurements with score less than 0, and 31 out of 259 measurements above 100. The non-linear distribution of residuals observed. The flat oblique lines reflect the nature of Barthel Index scores with the presence of floor and ceiling effect (above 40% with observed score of 0 at baseline)

Appendix E: Data abstraction form

Data abstraction - Incidence and Outcome of Stroke and Their Related Population and Patient Factors in Kelantan, Malaysia

Instructions:

1. These section is to be filled by a medical doctor
2. Fill in the blank or mark (v) where appropriate

A) BIODATA		Label	
Full name of patient (<i>namemr</i>)			
	Example: Muhd Karim		
Full number of the hospital registration number (<i>rnmr</i>)			
	Example: AB1234 (without dash)		
Full number of the identification card (<i>icmr</i>)			
	Example: 010160031234 (without dash)		
Study ID - First 6 numbers of IC and first 4 hospital registration numbers- (<i>idmr</i>)			
	Example: 010160AB12		
Village/Street name (<i>addmr</i>)			
District (<i>distr</i>)	<input type="checkbox"/> Kota Bharu=1 <input type="checkbox"/> Pasir Mas=2 <input type="checkbox"/> Tumpat=3 <input type="checkbox"/> Bachok=4 <input type="checkbox"/> Pasir Putih=5 <input type="checkbox"/> Machang=6 <input type="checkbox"/> Tanah Merah=7 <input type="checkbox"/> K Krai=8 <input type="checkbox"/> Jeli=9 <input type="checkbox"/> Gua Musang=10 <input type="checkbox"/> Not Recorded=11 <input type="checkbox"/> Missing =12		
Sex (<i>sexmr</i>)	<input type="checkbox"/> Male=1 <input type="checkbox"/> Female=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9		
Date of birth - dd/mm/yy - (<i>dobmr</i>)		Age in years (<i>agemr</i>)	<input type="checkbox"/> Not Recorded=99 <input type="checkbox"/> Missing =999
Race (<i>racemr</i>)	<input type="checkbox"/> Malay=1 <input type="checkbox"/> Chinese=2 <input type="checkbox"/> Indian=3 <input type="checkbox"/> Others=4 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing =9		
Date of admission (<i>doamr</i>)		Date of discharge (<i>dodmr</i>)	<input type="checkbox"/> Not Recorded=99 <input type="checkbox"/> Missing =999
	dd/mm/yy		
Status at discharge (<i>sad</i>)	<input type="checkbox"/> Alive=1 <input type="checkbox"/> Dead=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing =9		
If dead, state date of death (<i>datedied</i>)	dd/mm/yy		

Is this the first ever stroke? (<i>first</i>)	<input type="checkbox"/> Yes =1 Jump to B) if 'Yes' <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing =9	
Date of the last stroke (<i>datestroke</i>)	dd/mm/yy	

B) CLINICAL PROFILES			Label
Earliest Glasgow Coma Scale (<i>gcs</i>)	over 15		<input type="checkbox"/> Not Recorded=88 <input type="checkbox"/> Missing =99
Earliest Systolic Blood Pressure in mmHg (<i>sbp</i>)		Earliest Diastolic Blood Pressure in mmHg (<i>dbp</i>)	<input type="checkbox"/> Not Recorded=888 <input type="checkbox"/> Missing =999

C) MEDICAL AND SURGICAL HISTORY				
Diabetes (<i>dm</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Hypertension (<i>hpt</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Chronic kidney disease (<i>ckd</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Atrial fibrillation (<i>af</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Heart failure/IHD (<i>hd</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Dyslipidemia (<i>dyslipid</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Previous TIA (<i>tia</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9
Smoker (<i>smoker</i>)	<input type="checkbox"/> Yes=1	<input type="checkbox"/> No=2	<input type="checkbox"/> Not Recorded=8	<input type="checkbox"/> Missing =9

D) EARLIEST LABORATORY RESULTS				Label
Haemoglobin (<i>hb</i>)		Blood glucose (<i>gluc</i>)		<input type="checkbox"/> Not Recorded=888 <input type="checkbox"/> Missing=999
Platelet count (<i>plt</i>)		Capillary Blood Sugar (<i>cbs</i>)		
White cell count (<i>wbc</i>)		Cholesterol (<i>chol</i>)		
Na (<i>na</i>)		Triglyceride (<i>tg</i>)		
Potassium (<i>pt</i>)		Urea (<i>urea</i>)		

E) REFERRAL AND TRANSPORT	
Referral (<i>referral</i>)	<input type="checkbox"/> From a tertiary hospital=1 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> From a general practice=2 <input type="checkbox"/> From a district hospital=3 <input type="checkbox"/> Missing=9 <input type="checkbox"/> Direct from the scene/home=4
Transport (<i>transp</i>)	<input type="checkbox"/> by ambulance=1 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> by a private/personal transport=2 <input type="checkbox"/> Missing=9

F) DIAGNOSIS			
Diagnosis based on ICD-10 (<i>icd10disc</i>)			
<input type="checkbox"/> Sub-arachnoid haemorrhage=1 <input type="checkbox"/> Intra-cerebral Haemorrhagic=2 <input type="checkbox"/> Cerebral infarction=3 <input type="checkbox"/> Stroke, not specified as haemorrhage or infarction=4 <input type="checkbox"/> Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction=5	<input type="checkbox"/> Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction=6 <input type="checkbox"/> Other cerebrovascular diseases=7 <input type="checkbox"/> Cerebrovascular disorders in diseases classified elsewhere=8 <input type="checkbox"/> Sequela of cerebrovascular disease=9 <input type="checkbox"/> Not Recorded=99 <input type="checkbox"/> Missing =999		
G) EARLIEST MRI SCAN			Label
Date of MRI (<i>datemri</i>) dd/mm/yy			
Time of MRI (<i>timemri</i>) mm:hh			
		<input type="checkbox"/> Not done (jump to H) <input type="checkbox"/> Not Recorded=99 <input type="checkbox"/> Missing =999	
Hyper-intense area ** not suppressed by FLAIR (<i>hyperintmri</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9	Mass effect (<i>mass</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9
Hypo-intense area (<i>hypointmri</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9	Multiplicity (<i>multimri</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9
H) EARLIEST CT SCAN			Label
Date of CT scan (<i>detect</i>) dd/mm/yy			
Time of CT scan (<i>timect</i>) mm:hh			
		<input type="checkbox"/> Not done <input type="checkbox"/> Not Recorded=99 <input type="checkbox"/> Missing =999	
Insular ribbon sign (<i>insular</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9	Hypo-dense *lesion feature of ischaemic stroke (<i>hypodens</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9
Hyper-dense MCA (<i>mca</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9	Effacement of basal cistern (<i>efface</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9
Multiplicity (<i>multict</i>)	<input type="checkbox"/> Yes=1 <input type="checkbox"/> No=2 <input type="checkbox"/> Not Recorded=8 <input type="checkbox"/> Missing=9		

Name of the doctor:

Contact number :

Date :

Appendix F: Interview Form (Malay version)

BAHAGIAN B: TEMURAMAH

Arahan kepada penemubual: Anda perlu memperkenalkan diri dan tujuan anda menemubual kepada peserta kajian iaitu pesakit atau ahli keluarga yang menjawab panggilan telefon. Contoh:

“Selamat tengahari. Assalamualaikum, nama saya <nama anda> dan saya bekerja di Jabatan Perubatan Masyarakat, Hospital Universiti Sains Malaysia.. Adakah ini Encik/Puan.....?”

➔ Jika jawapan [Ya] : terus ke (PESAKIT)

➔ Jika jawapan [Tidak] : tanya nama beliaudan terus ke (WARIS)

(PESAKIT)

Katakan, “Encik/Puan <nama pesakit>, saya nak maklumkan tujuan saya. Untuk makluman, tuan/puan dijemput untuk menjawab soalan-soalan dalam projek kajian serangan strok yang tuan/puan alami pada <tarikh>. Sekarang saya akan terangkan beberapa perkara penting mengenai projek ini. Projek ini telah mendapat kelulusan Jawatankuasa Etika Kajian Manusia, Universiti Sains Malaysia. Semua maklumat yang diberi adalah terpelihara dan tidak akan didedahkan kepada orang lain. Projek ini bertujuan mengenal pasti faktor-faktor penting berkenaan strok di Kelantan. Temubual ini mengambil masa lebih kurang 15 minit.

Penyertaan dalam temubual ini sepenuhnya sukarela. Maksudnya tuan/puan tidak perlu mengambil bahagian jika tidak ingin berbuat demikian. Jika tuan/puan bersetuju untuk menyertai temubual ini, tuan/puan bebas menjawab hanya soalan-soalan yang tuan/puan sedia untuk jawab sahaja. Risiko dalam kajian ini adalah minimum sekali. Kerahsiaan maklumat tuan/puan dalam projek ini adalah keutamaan kami paling tinggi. Tuan/puan bebas menghentikan penyertaan bila-bila masa sahaja semasa temubual ini.

Jika tuan/puan mempunyai soalan-soalan atau pertanyaan mengenai projek ini, tuan/puan boleh hubungi Ketua Penyelidik, Dr Mohd Nazri Shafei di 0199761992 atau Puan Mazlita Zainal Abidin, Setiausaha, Jawatankuasa Etika Kajian Manusia, Universiti Sains Malaysia (Tel: 09-7672355). Jawatankuasa ini menjaga kepentingan dan kebajikan peserta-peserta kajian di Universiti Sains Malaysia.

Keputusan kajian mungkin diterbitkan dalam jurnal-jurnal saintifik untuk meningkatkan kefahaman mengenai penyakit strok dan memperbaiki pencegahan dan pengawalan penyakit ini. Kami akan maklumkan pautan ke bahagian keputusan kajian sebaik projek dihabiskan jika diminta.”

“Adakah tuan/puan mempunyai soalan?”

"Adakah tuan/puan bersetuju untuk menyertai temubual ini dengan sukarela?"

[Jika persetujuan diberi, TERUSKAN, if tidak, HENTIKAN temubual]

Boleh kita mula sekarang?

- ➔ Jika jawapan [Ya] : Terus ke **(B2)**
- ➔ Jika jawapan [Tidak] : Katakan, “Malang sekali kerana kita tidak dapat meneruskan temuramah ini sekarang. Bila masa yang sesuai untuk saya hubungi anda semula?
Tarikh (.....) Masa (.....)

(WARIS)

Katakan, “Untuk makluman, Encik/Puan <nama pesakit> dijemput untuk menjawab soalan-soalan dalam projek kajian serangan strok yang beliau alami. Sekarang saya akan terangkan beberapa perkara penting mengenai projek ini. Projek ini telah mendapat kelulusan Jawatankuasa Etika Kajian Manusia, Universiti Sains Malaysia. Semua maklumat yang diberi adalah terpelihara dan tidak akan didedah kepada orang lain. Projek ini bertujuan mengenal pasti faktor-faktor penting berkenaan strok di Kelantan. Temubual ini mengambil lebih kurang 15 minit.

Penyertaan dalam temubual ini sepenuhnya sukarela. Maksudnya tuan/puan tidak perlu mengambil bahagian jika tidak ingin berbuat demikian. Jika tuan/puan bersetuju untuk menyertai temubual ini, tuan/puan bebas menjawab hanya soalan-soalan yang tuan/puan sedia untuk jawab sahaja. Risiko dalam kajian ini adalah minimum sekali. Kerahsiaan maklumat tuan/puan dalam projek ini adalah keutamaan kami paling tinggi. Tuan/puan bebas menghentikan penyertaan bila-bila masa sahaja semasa temubual ini.

Jika tuan/puan mempunyai soalan-soalan atau pertanyaan mengenai projek ini, tuan/puan boleh hubungi Ketua Penyelidik, Dr Mohd Nazri Shafei di 0199761992 atau Puan Mazlita Zainal Abidin, Setiausaha, Jawatankuasa Etika Kajian Manusia, Universiti Sains Malaysia (Tel: 09-7672355). Jawatankuasa ini menjaga kepentingan dan kebajikan peserta-peserta kajian di Universiti Sains Malaysia.

Keputusan kajian mungkin diterbitkan dalam jurnal-jurnal saintifik untuk meningkatkan kefahaman mengenai penyakit strok dan memperbaiki pencegahan dan pengawalan penyakit ini. Kami akan maklumkan pautan ke bahagian keputusan kajian sebaik projek dihabiskan jika diminta.”

“Adakah tuan/puan mempunyai soalan?”

"Adakah tuan/puan bersetuju untuk menyertai temubual ini dengan sukarela?"

[Jika persetujuan diberi, TERUSKAN, if tidak, HENTIKAN temuramah]

Boleh kita mula sekarang?

- ➔ Jika jawapan [Ya] : Terus ke **(B1)**

➔ Jika jawapan [Tidak] : Katakan, “malang sekali kerana kita tidak dapat meneruskan temuramah ini sekarang. Bila masa yang sesuai untuk saya hubungi anda semula?

Tarikh (.....) Masa (.....)

Name of the person giving consent	:.....
Relation to patient	: Patient / Next of kin (specify:)
Contact number	:.....
Date and time of consent given	:.....

Informer	<input type="checkbox"/> Pesakit=0 <input type="checkbox"/> Waris=1	informer
----------	---	----------

B1) ALAMAT TEMPAT TINGGAL KETIKA BERLAKUNYA STROK	Katakan: “Sekarang, saya akan tanya mengenai, alamat lengkap tempat tinggal tuan/puan ketika berlakunya strok tempoh hari.”	
1. Nama taman Contoh: Taman Aman	taman
2. Nama kampung Contoh: Kampung Sentosa	vil
3. Nama jalan Contoh: Jalan Meranti	street
4. Poskod Contoh: 16150	postcode
5. Bandar Contoh: Kota Bharu	town
6. Daerah Contoh: Kota Bharu	district
B2) ALAMAT TEMPAT TINGGAL KINI	Katakan: “Sekarang, saya akan tanya alamat penuh, tempat kediaman dalam setahun ini.”	
	<input type="radio"/> Alamat sama seperti di atas	
	<input type="radio"/> Alamat baru	

1 Nama taman Contoh: Taman Uda	taman2
2 Nama kampung Contoh: Kampung Bahagia	vil2
3 Nama jalan Contoh: Jalan Mawar	street2
4 Poskod Contoh: 16150	postcode2
5 Bandar Contoh: Kota Bharu	town2
6 Daerah Contoh: Kota Bharu	district2

C) DEMOGRAFI	Katakan: "Sekarang, saya ingin bertanya mengenai status sosio-ekonomi dan sejarah penyakit yang lampau."	code
1. Tuan/puan berkerja sekarang?	<input type="checkbox"/> Ya=0 <input type="checkbox"/> Sudah berhenti=1 <input type="checkbox"/> Tidak pernah bekerja=2	empstat
2. Apakah nama pekerjaan tuan/puan? Contoh: guru	empname
3. Berapa anggaran pendapatan tuan/puan sebulan? Contoh: RM2500	incmonth
4. Berapa anggaran pendapatan sebulan isi rumah (semua yang tinggal sebumbung)? Contoh: RM5000	incmonth2
5. Apakah tahap tertinggi pendidikan yang tuan/puan terima?	<input type="checkbox"/> Ijazah dan lebih tinggi=0 <input type="checkbox"/> Diploma=1 <input type="checkbox"/> Sekolah menengah=2	educ

	<input type="checkbox"/> Sekolah rendah=3 <input type="checkbox"/> Tiada pendidikan formal=4	
6. Tuan/puan pernah alami penyakit darah tinggi	<input type="checkbox"/> Tidak=0 <input type="checkbox"/> Ya=1	hhpt
7. Tuan/puan pernah alami sakit jantung	<input type="checkbox"/> Tidak=0 <input type="checkbox"/> Ya=1	hheart
8. Tuan/puan pernah alami penyakit kencing manis?	<input type="checkbox"/> Tidak=0 <input type="checkbox"/> Ya=1	hdm
9. Ibumama tuan/puan pernah kena strok?	<input type="checkbox"/> Tidak=0 <input type="checkbox"/> Ya=1	hstroke
10. Pernah ada sejarah strok di kalangan adik-beradik tuan/puan?	<input type="checkbox"/> Tidak=0 <input type="checkbox"/> Ya=1	hstroke2
D) LAIN-LAIN	Katakan: "Sekarang, dua soalan ini adalah sensitif tetapi penting dalam kajian kami."	code
1. Adakah tuan/puan minum arak?	<input type="checkbox"/> Tidak pernah=0 <input type="checkbox"/> Ya, tetapi telah berhenti lebih dari 6 bulan yang lalu=1 <input type="checkbox"/> Ya=2	alcohol
2. Adakah tuan/puan merokok?	<input type="checkbox"/> Tidak pernah=0 <input type="checkbox"/> Ya, tetapi telah berhenti lebih dari 6 bulan yang lalu=1 <input type="checkbox"/> Ya=2	smoke

Tarikh temuramah pertama (semasa discaj) : dd/mm/yy
Date of the first interview (at discharge)

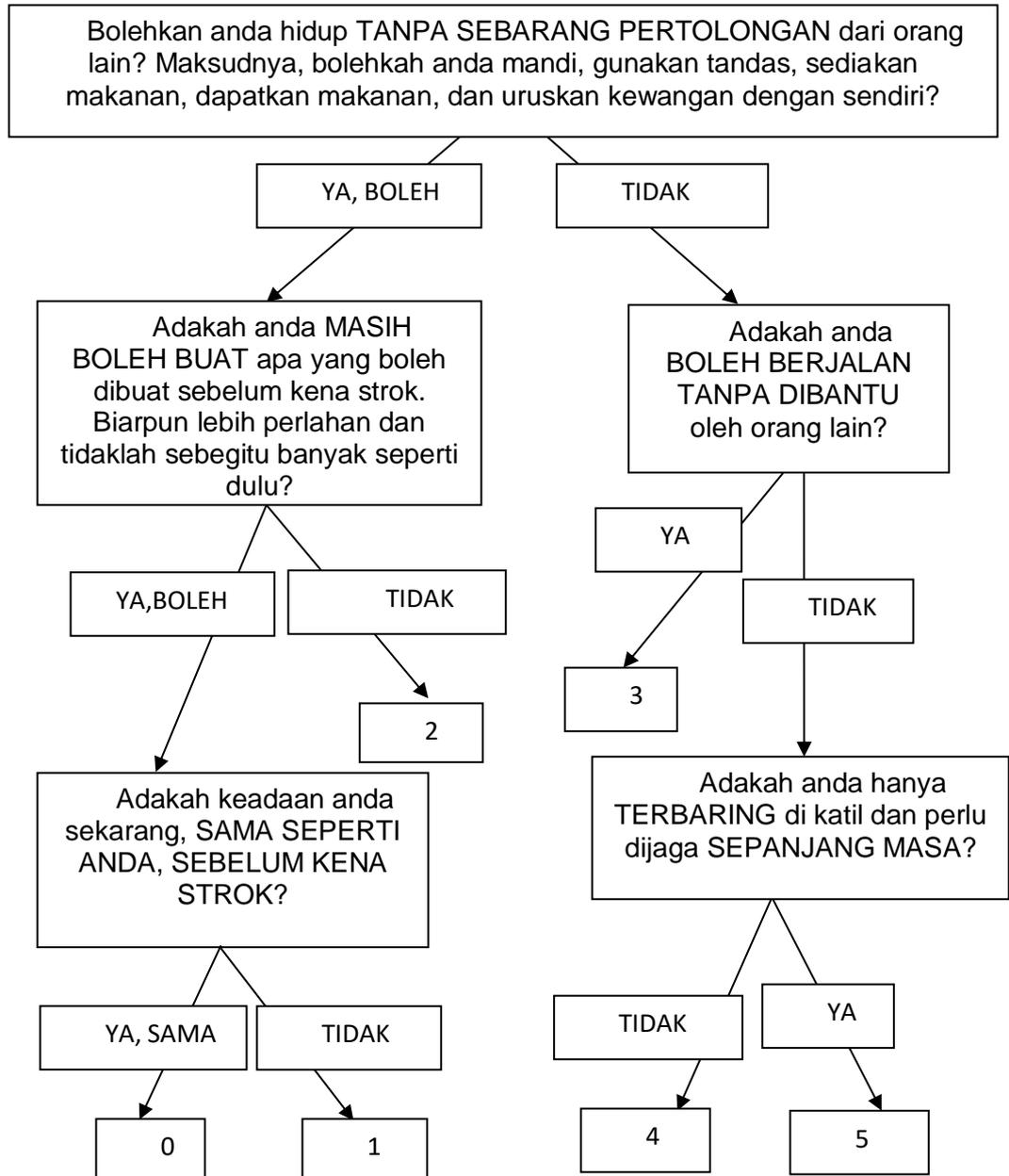
Tarikh temuramah kedua (30 hari selepas discaj) : dd/mm/yy
Date of the second interview (at day 30 post discharge)

Tarikh temuramah ketiga (90 hari selepas discaj) : dd/mm/yy
Date of the third interview (at day 90 post discharge)

Jika temuramah tidak dapat diteruskan kerana pesakit telah meninggal, tuliskan tarikh meninggal:
Tarikh kematian *Date of the death* : dd/mm/yy

BAHAGIAN E: MODIFIED RANKIN SCALE

Arahan: Tanyakan pesakit atau waris, carta-alir di bawah dan bulatkan SATU respon yang relevan. Katakan kepada pesakit atau waris "Kami akan tanyakan keadaan tuan/puan atau pesakit sekarang. Sebutlah pilihan yang paling hampir dengan keadaan sekarang. Ini soalnya"



Mas Timing	Semasa Discaj On discharge	30 hari selepas discaj Day 30 post discharge	90 hari selepas discaj Day 90 post discharge
Code	mrsdis	mrs1	mrs3
Date of assessment dd/mm/yy dd/mm/yy dd/mm/yy
Modified Rankin Scale (Tick where appropriate)	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5

BAHAGIAN F: BARTHEL INDEX

Arahan kepada penemubual: Barthel Index hanya ditanya kepada pesakit-pesakit yang masih hidup. Tanyalah pesakit sendiri atau ahli keluarga terdekat. Tandakan (v) di respon yang paling relevan dengan pesakit dalam tempoh seminggu ini. Jawap setiap soalan:

Katakan kepada pesakit "Tuan/Puan, soalan-soalan berikut bertanya mengenai penjagaan diri Tuan/Puan. Tolong maklumkan apa yang Tuan/Puan berupaya buat lebih kurang seminggu terakhir ini."

D) BARTHEL INDEX	TANDAKAN (v) PADA RESPON PALING RELEVAN			code
	Semasa Discaj <i>On discharge</i>	30 hari selepas discaj <i>Day 30 post discharge</i>	90 hari selepas discaj <i>Day 90 post discharge</i>	
 dd/mm/yy dd/mm/yy dd/mm/yy	
1. Bagaimana anda menggunakan tandas atau alatan untuk membuang air besar?	<input type="radio"/> Perlu ditolong sepenuhnya =0 <input type="radio"/> Perlu ditolong sedikit =5 <input type="radio"/> Boleh buat sendiri =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	toilet
2. Bagaimana anda makan?	<input type="radio"/> Perlu banyak ditolong =0 <input type="radio"/> Perlu ditolong sebahagiannya seperti untuk mengerat makanan =5 <input type="radio"/> Tak perlu ditolong =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	feed
3. Bagaimana anda berpindah dari katil ke kerusi?	<input type="radio"/> Tidak boleh langsung =0 <input type="radio"/> Perlu banyak ditolong dari satu atau dua orang =5 <input type="radio"/> Perlu ditolong sedikit dari satu orang =10 <input type="radio"/> Boleh berpindah sendiri =15	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10 <input type="radio"/> =15	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10 <input type="radio"/> =15	move
4. Bagaimana anda bergerak dari satu	<input type="radio"/> Tidak mampu bergerak langsung =0	<input type="radio"/> =0	<input type="radio"/> =0	getabo

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

D) BARTHEL INDEX	TANDAKAN (√) PADA RESPON PALING RELEVAN			code
tempat ke tempat lain?	<input type="radio"/> Boleh menolak sendiri kerusi roda tanpa di tolong =5 <input type="radio"/> Boleh berjalan, tapi ada seorang menolong atau memerhati =10 <input type="radio"/> Boleh berjalan sendiri walaupun menggunakan tongkat atau frame =15	<input type="radio"/> =5 <input type="radio"/> =10 <input type="radio"/> =15	<input type="radio"/> =5 <input type="radio"/> =10 <input type="radio"/> =15	
5. Adakah anda perlu ditolong bila memakai pakaian?	<input type="radio"/> Ya, perlu ditolong hampir sepenuhnya =0 <input type="radio"/> Ya, saya mampu pakai pakaian sendiri lebih kurang separuh =5 <input type="radio"/> Tidak perlu. Saya boleh pakai pakaian sendiri sepenuhnya =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	dress
6. Bagaimana anda naik atau turun tangga?	<input type="radio"/> Tidak boleh langsung =0 <input type="radio"/> Perlu ditolong samada secara lisan, fizikal =5 <input type="radio"/> Boleh naik turun tangga sendiri =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	stairs
7. Adakah anda perlu ditolong bila mandi atau mencuci badan?	<input type="radio"/> Ya =0 <input type="radio"/> Tidak =5	<input type="radio"/> =0 <input type="radio"/> =5	<input type="radio"/> =0 <input type="radio"/> =5	bath
8. Adakah anda boleh kawal buang air besar (berak)?	<input type="radio"/> Tidak boleh kawal buang air besar =0 <input type="radio"/> Sekali sekala (sekali seminggu) terbuang air besar dalam pakaian=5 <input type="radio"/> Boleh. Dapat kawal buang air besar sepenuhnya =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	bowel

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

D) BARTHEL INDEX	TANDAKAN (√) PADA RESPON PALING RELEVAN			code
9. Adakah anda boleh kawal kencing?	<input type="radio"/> Tidak boleh kawal kencing sendiri. Atau saya memakai tiub kencing =0 <input type="radio"/> Sekali sekala (paling kerap pun sekali sehari) terkencing dalam pakaian =5 <input type="radio"/> Boleh kawal kencing sepenuhnya =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	urine

Katakan “Encik/Puan <nama>, temubual kita telah berakhir. Kami akan meneliti respon temuramah ini dan akan menghubungi Tuan/Puan jika perlu. Jika ada persoalan, silalah hubungi kami. Kami ingin mengucapkan sebanyak terima kasih terhadap kerjasama Tuan/Puan. Terima kasih”

Nama penemubual :

.....

Nombor untuk dihubungi :

.....

TEMURAMAH TAMAT

Appendix G: Interview Form (English version)

PART B: INTERVIEW

Instruction to interviewer: You must introduce yourself and tell patient or relative the reasons for the interview. For example:

“Good afternoon. My name is <your name> and I am working at the Department of Community Medicine, Hospital Universiti Sains Malaysia. Are you Mr/Mrs?”

- ➔ If the answer is [Yes] : proceed to (PATIENT)
- ➔ If the answer is [No] : ask their name and proceed to (RELATIVE)

(PATIENT)

Say, “Mr/Mrs <name>, I like to inform you the reasons for this interview. For your information, you are invited to answer a set of questions related with stroke that you had on <date>. Now, I will tell you important information about this project. This project has obtained the ethical approval from the Human Research Ethics Committee, Universiti Sains Malaysia. All information in this study is confidential and will not be disclosed to others. This project aims to identify important factors related with stroke in Kelantan. This interview takes about 15 minutes.

Your participation in this interview is fully voluntary. It means that you may decide not to take part if you do not want to. If you agree to participate in this interview, you have the freedom to only answer questions that you like to. The risk in this study is minimal. The confidentiality of the information in this study is our utmost concern. You are free to stop anywhere and anytime during this interview.

If you do have questions or queries with regards to this project, you may contact the Principal Investigator, Dr Mohd Nazri Shafei at 0199761992 or Puan Mazlita Zainal Abidin, Secretary, Human Research Ethics Committee, Universiti Sains Malaysia (Tel: 097672355). This committee is responsible for the safety and welfare of all participants in studies done by Universiti Sains Malaysia.

The results of the study may be published in scientific journals to improve the understanding of stroke and to improve the control and prevention of this disease. We will inform you the link to the results once this project has finished, if requested.”

“Do you have any question?”

"Do you agree to voluntarily participate in this interview?"

[If consent given, PROCEED, if not, STOP the interview]

Can we start now?

- ➔ If the answer is [Yes] : Proceed to **(B2)**
- ➔ If the answer is [No] : Say, “It is unfortunate that we will not be able to continue with the interview now. When is the appropriate date so I can call you again?
Date (.....) Time (.....)

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

(RELATIVE)

Say, "For your information, Mr/Mrs <name> is invited to answer a set of questionnaire related with stroke that he/she had before. Now, I will tell you important information about this project. This project has obtained the ethical approval from the Human Research Ethics Committee, Universiti Sains Malaysia. All information in this study is confidential and will not be disclosed to others. This project aims to identify important factors related with stroke in Kelantan. This interview takes about 15 minutes.

Your participation in this interview is fully voluntary. It means that you may decide not to take part if you do not want to. If you agree to participate in this interview, you have the freedom to only answer questions that you like to. The risk in this study is minimal. The confidentiality of the information in this study is our utmost concern. You are free to stop anywhere and anytime during this interview.

If you do have questions or queries with regards to this project, you may contact the Principal Investigator, Dr Mohd Nazri Shafei at 0199761992 or Puan Mazlita Zainal Abidin, Secretary, Human Research Ethics Committee, Universiti Sains Malaysia (Tel: 097672355). This committee is responsible for the safety and welfare of all participants in studies done by the Universiti Sains Malaysia.

The results of the study may be published in scientific journals to improve the understanding of stroke and to improve the control and prevention of this disease. We will inform you the link to the results once this project has finished if requested."

"Do you have any question?"

"Do you agree to voluntarily participate in this interview?"

[If consent given, PROCEED, if not, STOP the interview]

- ➔ If the answer is [Yes] : Proceed to **(B1)**
- ➔ If the answer is [No] : Say, "It is unfortunate that we will not be able to continue with the interview now. When is the appropriate date so I can call you again?
Date (.....) Time (.....)

Name of the person giving consent	:.....
Relation to patient	: Patient / Next of kin (specify:)
Contact number	:.....
Date and time of consent given	:.....

Informer	<input type="checkbox"/> Patient=0 <input type="checkbox"/> Relative=1	informer
----------	--	----------

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

B1) ADDRESS WHEN STROKE HAPPENED		Say : "Now, I will ask the full address of your residence at the time you had the stroke."	
7.	Name of taman Example: Taman Aman	taman
8.	Name of village Example: Kampung Sentosa	vil
9.	Nama of street Example: Jalan Meranti	street
10.	Poscode Example: 16150	postcode
11.	Town Example: Kota Bharu	town
12.	District Example: Kota Bharu	district
B2) CURRENT RESIDENTIAL ADDRESS		Say : "Now, I will ask the full address of your residence since last year."	
		<input type="radio"/> Same address as above <input type="radio"/> New address	
7	Name of taman Example: Taman Uda	taman2
8	Name of village Example: Kampung Bahagia	vil2
9	Name of street Example: Jalan Mawar	street2
10	Postcode Example: 16150	postcode2
11	Town Example: Kota Bharu	town2
12	District Example: Kota Bharu	district2

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

C) DEMOGRAPHY	Say: "Now, I would like to ask about your socio-economy profiles and past medical history."	code
11. Are you currently employed?	<input type="checkbox"/> Yes=0 <input type="checkbox"/> Stopped working=1 <input type="checkbox"/> Never had jobs=2	empstat
12. What is your occupation? Example: teacher	empname
13. What is the estimated income every month? Example: RM2500	incmonth
14. What is the estimated household income every month (including all living in the same house)? Example: RM5000	incmonth2
15. What is the highest education attained by you?	<input type="checkbox"/> Degree or higher=0 <input type="checkbox"/> Diploma=1 <input type="checkbox"/> Secondary school=2 <input type="checkbox"/> Primary school=3 <input type="checkbox"/> No formal education=4	educ
16. Do you have history of high blood pressure?	<input type="checkbox"/> No=0 <input type="checkbox"/> Yes=1	hhpt
17. Do you have history of heart attack?	<input type="checkbox"/> No=0 <input type="checkbox"/> Yes=1	hheart
18. Do you have history of diabetes mellitus?	<input type="checkbox"/> No=0 <input type="checkbox"/> Yes=1	hdm
19. Did your parents ever have stroke?	<input type="checkbox"/> No=0 <input type="checkbox"/> Yes=1	hstroke
20. Do you have history of stroke among your siblings?	<input type="checkbox"/> No=0 <input type="checkbox"/> Yes=1	hstroke2
D) OTHERS	Say : "Now, these 2 questions are sensitive however, they are important in this study."	code
3. Do you drink alcohol?	<input type="checkbox"/> Never=0 <input type="checkbox"/> Yes but stopped more than 6 months ago=1	alcohol

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

	<input type="checkbox"/> Yes=2	
4. Do you smoke?	<input type="checkbox"/> Never=0 <input type="checkbox"/> Yes but stopped more than 6 months ago=1 <input type="checkbox"/> Yes=2	smoke

Tarikh temuramah pertama (semasa discaj) : dd/mm/yy
Date of the first interview (at discharge)

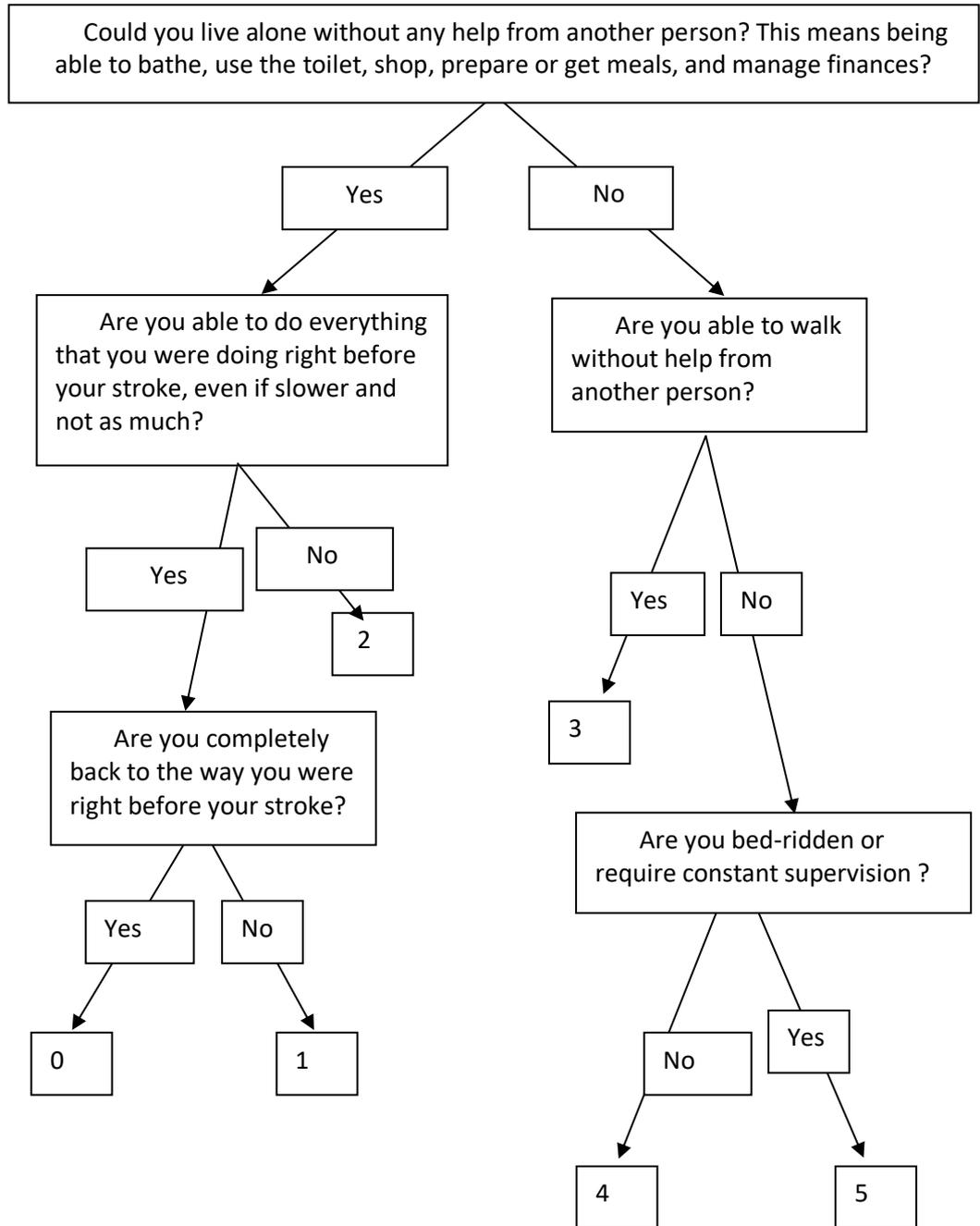
Tarikh temuramah kedua (30 hari selepas discaj) : dd/mm/yy
Date of the second interview (at day 30 post discharge)

Tarikh temuramah ketiga (90 hari selepas discaj) : dd/mm/yy
Date of the third interview (at day 90 post discharge)

Jika temuramah tidak dapat diteruskan kerana pesakit telah meninggal, tuliskan tarikh meninggal:
 Tarikh kematian : dd/mm/yy
Date of the death

SECTION E: MODIFIED RANKIN SCALE

Instruction: Ask patient or relative using flow-chart below and circle the MOST relevant response. Say to patient or relative "We will ask about your current situation. Choose the one that most resemble your current situation. This is the question"



Mas Timing	Semasa Discaj On discharge	30 hari selepas discaj Day 30 post discharge	90 hari selepas discaj Day 90 post discharge
---------------	-------------------------------	---	---

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

Code	Mrsdis	mrs1	mrs3
Date of assessment dd/mm/yy dd/mm/yy dd/mm/yy
Modified Rankin Scale (Tick where appropriate)	<input type="radio"/> 0	<input type="radio"/> 0	<input type="radio"/> 0
	<input type="radio"/> 1	<input type="radio"/> 1	<input type="radio"/> 1
	<input type="radio"/> 2	<input type="radio"/> 2	<input type="radio"/> 2
	<input type="radio"/> 3	<input type="radio"/> 3	<input type="radio"/> 3
	<input type="radio"/> 4	<input type="radio"/> 4	<input type="radio"/> 4
	<input type="radio"/> 5	<input type="radio"/> 5	<input type="radio"/> 5

SECTION F: BARTHEL INDEX

Instruction to interviewer:

This index should be asked if the patient is still ALIVE. Ask the participant or the next of kin. Tick (v) the most relevant option with the patient condition during this week. Answer each question.

Say to patient “Mr/Mrs <name>, the following questions ask about how you look after yourself. Please, tell me which of the following, have you actually done in the last week or so.”

D) BARTHEL INDEX	TICK ONE FOR EACH QUESTION			code
	Semasa Discaj <i>On discharge</i>	30 hari selepas discaj <i>Day 30 post discharge</i>	90 hari selepas discaj <i>Day 90 post discharge</i>	
 dd/mm/yy dd/mm/yy dd/mm/yy	
1. Do you wash your own face, brush your teeth and hair (For men, shave)?	<input type="radio"/> With help?=0 <input type="radio"/> Without help?=5	<input type="radio"/> =0 <input type="radio"/> =5	<input type="radio"/> =0 <input type="radio"/> =5	wash
2. Do you use the toilet (or commode)?	<input type="radio"/> With a lot of help?=0 <input type="radio"/> With a little help?=5 <input type="radio"/> On your own?=10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	toilet
3. Do you feed yourself?	<input type="radio"/> With major help?=0 <input type="radio"/> With some help? E.g. cutting=5 <input type="radio"/> Without any help?=10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	feed
4. How do you move from bed to the chair?	<input type="radio"/> Not at all =0 <input type="radio"/> With a lot of help from one or two people=5 <input type="radio"/> With a little help from one person?=10 <input type="radio"/> On your own?=15	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10 <input type="radio"/> =15	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10 <input type="radio"/> =15	move
5. How do you get about?	<input type="radio"/> No at all=0 <input type="radio"/> Propelling yourself independently in a wheelchair?=5	<input type="radio"/> =0 <input type="radio"/> =5	<input type="radio"/> =0 <input type="radio"/> =5	getabout t

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

	<input type="radio"/> Walking with the help or supervision of one person=10 <input type="radio"/> Walking with no one helping (even if you use a stick or frame)=15	<input type="radio"/> =10 <input type="radio"/> =15	<input type="radio"/> =10 <input type="radio"/> =15	
6. Do you need any help with dressing?	<input type="radio"/> Yes, I need help for almost everything=0 <input type="radio"/> Yes, I am able to do about half unaided=5 <input type="radio"/> No, I can do everything=10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	dress
7. How do you get up and down the stairs?	<input type="radio"/> No at all=0 <input type="radio"/> With help (either supervision or assistance)=5 <input type="radio"/> Without any help=10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	stairs
8. Do you need help with bathing or showering?	<input type="radio"/> Yes=0 <input type="radio"/> No=5	<input type="radio"/> =0 <input type="radio"/> =5	<input type="radio"/> =0 <input type="radio"/> =5	bath
9. Are you incontinent of bowels?	<input type="radio"/> Yes, incontinent=0 <input type="radio"/> Occasional accident (once per week)=5 <input type="radio"/> No, continent=10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	bowel
10. Are you of incontinent of urine?	<input type="radio"/> Yes, incontinent or I have or the participant has a catheter fixed=0 <input type="radio"/> Occasional accident (maximum once per 24 hours)=5 <input type="radio"/> No, continent=10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	<input type="radio"/> =0 <input type="radio"/> =5 <input type="radio"/> =10	urine

Say "Mr/Mrs, this is the end of the interview. We will review the response and will get back to you if necessary. All the information given here is strictly confidential and is accessible only to the researchers in this project. We would like to say thank you very much for your cooperation"

Name of the interviewer :

Date :

INCIDENCE AND OUTCOME OF STROKE AND THEIR RELATED POPULATION AND PATIENT FACTORS IN KELANTAN,
MALAYSIA:
INTERVIEW

END OF INTERVIEW

Appendix H: Methods to deal with informative censoring

Gamma imputation method

The gamma imputation method quantifies the sensitivity of the conclusion from a fitted Cox proportional hazards model when the independent censoring is in doubt, rather than focusing on the reasons why the assumption of independent censoring may be false (Jackson et al., 2014). It uses intuition that censoring associates with the severity of a person's condition. By modelling the resulting association between censoring and failure, investigators gain an informed view of the model's assumptions and the plausible range of sensitivity parameter or parameters (Burkoff et al., 2016a).

Consider a time to event data set where subject i has time to event T_i and is censored at time C_i . Each subject has an observed time $Y_i = \min(C_i; T_i)$ and event indicator δ_i where $\delta_i = 1$ when $T_i < C_i$ and $\delta_i = 0$ otherwise. The multiple imputation procedure generates M imputed datasets where subjects who were censored now have an imputed time $T_i^m \geq Y_i$ and a new event indicator $\delta_i^m = 1$. The data $\{ T_i^m, \delta_i^m \}$ are imputed under the assumption that at the point of censoring the log-hazard function jumps by a constant denoted by γ_i . With $\gamma_i \neq 0$, the independent censoring assumption has been relaxed. The user can vary the size and magnitude of γ_i to run analyses on each γ_i value. Based on the differences amongst parameter estimates from different γ_i , the user can then assess the importance of these differences. For example, standard survival analysis can be applied to each data-set from M imputed data sets. Finally, the resulting parameters estimates are combined using Rubin's rules.

The approach assumes that the hazard for failure, given that censoring has not yet occurred, is equal to,

$$h(t | C_i > t, Z_i, S_i) = h_{(0, S_i)}(t) \exp(\beta Z_i)$$

where Z_i are time-independent covariates for subject i , S_i is the stratum for subject i , $h_{(0, S_i)}(t)$ is the baseline hazard function for the stratum denoted by S_i and β are the regression coefficients. This model can be fitted to the observed data using partial likelihood in the standard way.

After censoring has occurred it is assumed that,

$$h(t | C_i < t, Z_i, S_i) = h_{(0, S_i)}(t) \exp(\beta Z_i) \exp(\gamma_i),$$

therefore if $\gamma_i > 0$ there is an elevated risk for failure after censoring and if $\gamma_i < 0$ there is a decreased risk after censoring (Jackson et al., 2014, Burkoff et al., 2016a). The parameter γ_i represents the change in log-hazard for failure following censoring, conditional on the covariates in the imputation model.

The user must specify the number of imputations to be generated. Too small a number will lead to non-negligible simulation variability. The user hence may use his/her personal judgment to determine acceptable precision. The Cox model will then be fitted to a bootstrap sample of the original data set.

Given that a subject in stratum S_i and with covariates Z_i , was censored at time C_i so $\delta_i = 0$, an imputed failure time ($> C_i$) is sampled from the model

$$h(t | C_i, Z_i, S_i) = h_{j(0, S_i)}(t) \exp(\bar{\beta}_j Z_i + \gamma_i)$$

Where $h_{j(0,s_i)}$ and $\bar{\beta}_j$ are from the model fit associated with the bootstrapped sample j . If the imputed time is $> F_i$ (F_i is the maximum follow up period) then a time of F_i and event indicator $\delta_i = 0$ are imputed.

Risk score multiple imputation method

Consider a two-arm time-to-event data-set where subject i has event time T_i and potential censoring time C_i . For each subject we observe a time $X_i = \min(C_i; T_i)$ and event indicator Δ_i which = 1 if the subject was observed to have had an event at X_i and = 0 otherwise. The independent censoring assumption states that T_i and C_i are independent, and when this assumption is violated, standard methods for inference are in general invalidated.

The risk score imputation approach creates multiple imputations of event times for those subjects whose event times were censored. The imputation procedure utilizes subject-level covariates, including time-varying covariates, which are known or believed to be related to the hazard of failure and/or the hazard of censoring.

In risk score imputation, the method works by generating m imputed event times $Y_i^m \geq X_i$ and event indicators Δ_i^m . The data $\{ Y_i^m, \Delta_i^m \}$ is imputed by creating a risk set of similar subjects to subject i and then using a procedure called Kaplan-Meier imputation (KMI). Once the M data sets have been imputed, standard time to event statistical analyses (for example the log rank test) can be applied to each data set and the results combined (as described below) to produce point estimates of model parameters or to perform a hypothesis test (e.g. of equality of survivor functions).

Generally, this method involves these 3 steps: a) calculation of risk set, b) Kaplan-Meier imputation and c) Bootstrapped data to fit the models.

When calculating the risk set, we first calculate (for each censored subject) a risk set $R(i+, NN)$ which contains the nearest NN subjects to subject i (in the same treatment group) with event/censoring time greater than X_i where NN is a user specified number. In the calculation of the risk sets, the separate treatment groups are considered independently.

Next, we use Kaplan-Meier estimates to impute $\{ Y_i^m, \Delta_i^m \}$. By sampling from the Kaplan-Meier estimator of the survivor function for subjects in the risk set.

To ensure that the multiple imputations are proper, we then perform a bootstrapping step. For each imputed data set, proportional hazard models are fitted to a bootstrapped data set and risk scores are calculated. These scores can then be used for risk set identification for subject i .

Software package for multiple imputation in time-to-event data

Standard software assumes independent censoring, conditional on the covariates in the analysis model. This assumption is untestable and doubtful for individuals censored before the scheduled end of the study. Procedures that relax this assumption will often be useful.

In R, the '*InformativeCensoring*' package perform methods of multiple imputation for time-to-event data when the non-informative censoring assumption is violated. This might be true if the reason for censoring is related to the failure process. The package provides two methods of multiple imputation:

- a) Non-parametric multiple imputation that enables nonparametric comparison of two survival functions with dependent censoring (Hsu and Taylor, 2009) – the risk-score imputation method

- b) Multiple imputation that allows the Cox proportional hazard method to be used with independent censoring assumption (Burkoff et al., 2016a, Jackson et al., 2014) – the gamma-imputation method

In both methods, the user relaxes the assumption of non-informative censoring to one of being non-informative conditional on the imputation model. To maintain the robustness and valid inferences, the Cox models should of course be correctly specified, and the conditional non-informative censoring assumption satisfied (Bartlett, 2016).

Assessing non-informative censoring in our data

One of several possible models for informative censoring is to assume that both the censoring and failure times, C and F , for a patient depend on an unobserved variable U , but are conditionally independent given U (see Figure H-1). If we can know the value of U for each patient, *i.e.* if it becomes an observed covariate, C and F are independent and an analysis that includes this covariate the censoring is no longer informative.

Next, we will show (through Cox Proportional Hazard (PH) and reversed Cox PH) how each of C and F depends on our observed covariate, e.g. Glasgow Coma Scale and stroke subtypes. We aim to estimate the relationship between the covariates with the censored event (C) and failure events (F) and assess the shared dependency.

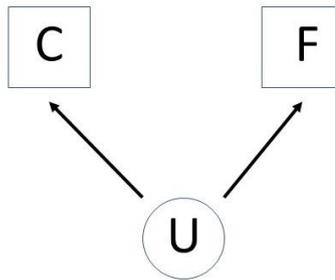


Figure H.1. C=censoring time, F=failure time. Censoring is informative if *U* is unobserved, non-informative if *U* is observed covariate.

In Table H.1, we model our data using the Cox Proportional Hazard (PH) model and in Table H.2, we run the reversed Cox proportional hazard model in which the roles of the censoring and failure times are reversed. In the Cox PH analysis, patients who died during admission (n= 53) were treated as the failure cases and those alive at discharge (n= 173) as censored cases. In the reverse Cox PH model, we run the analysis with patients who died during admission treated as the censored cases and patients who were discharged alive as failures. In both Cox PH models, The Glasgow Coma Scale (GCS) and the stroke subtype remain as significant and important predictors for in-hospital stroke fatality (both p-values < 0.001). Age however was significant at 5% level in the Cox PH model but not in the reverse Cox PH model. This does make sense because between the three (GCS, stroke subtype and age), GCS and stroke are the two most clinically plausible predictors for in-hospital stroke fatality. It also underlines the importance of including both covariates to make the assumption of non-informative censoring more plausible.

We used a similar procedure to compare the prognostic effect between Ischaemic Stroke (IS) and Haemorrhagic Stroke (HS). In this analysis, we have three categories of outcome: 1) died during admission (n= 84), 2) discharged alive (n= 206) and 3) discharged at-own-risk (aor) (n= 2). In the third category (discharged aor), the patients or family members

requested hospital discharge against the physicians or surgeons' advice. But both had full Glasgow Coma Scale (full consciousness) based on our data.

We ran the Cox PH model first with patients who died during admission set as failure cases and patients either discharged alive or aor set as censor cases. We show the results in Table H.3. Next, we ran the reversed Cox PH model, treating patients who were discharged alive and aor discharge as failure cases and patients who died during admission as censored cases. The results are shown in Table H.4.

The results in Table H.4 are in reverse of that of Table H.3. In both tables, stroke subtype remains as an important prognostic variable for in-hospital stroke fatality and all p-values are still less than 0.001.

In all the four analysis that we have performed above, we found that the censored events (C) and failure events (F) depends on the observed covariates (U), especially the Glasgow Coma Scale (GCS) and stroke subtypes (see Figure H-1). In daily clinical practice, GCS and stroke subtypes are known as the two most important predictors in stroke fatality. Physicians have long used them to guide in their stroke care. We believe the shared dependency shown in our Cox PH and reverse Cox PH models increases the plausibility (but does not and cannot prove) that including such covariates in our analysis makes the censoring non-informative.

Table H.1 The regression coefficients and adjusted hazard ratios estimated from the Cox proportional hazard model.

	Bet	Adj HR	SE ^b	Lowe	Uppe	p-
<u>Best model (Model 1)</u>						
n=225						
Glasgow Coma Scale (GCS)	-	0.82	0.0	0.77	0.88	<0.001
Age (years)	0.03	1.03	0.0	1.01	1.05	0.011
<u>Alternative model (Model 2)</u>						
n=226						
Haemorrhagic	0.88	2.40	0.7	1.30	4.44	0.005
Non-haemorrhagic	0	1.00				
Age (years)	0.03	1.03	0.0	1.01	1.05	0.014

^a hazard ratios obtained from the multivariable Cox PH regression ^b standard error for hazard ratio.

^c confidence interval. Alive = censor cases (coded = 0), Dead = failure cases (coded = 1)

Table H.2 The regression coefficients and adjusted hazard ratios estimated from the reversed Cox proportional hazard model.

	Beta	Adj HR ^a	Lower	Upper	p-value
<u>Best model (Model 1)</u>					
n=225					
Glasgow Coma Scale (GCS)	0.18	1.20	1.13	1.27	<0.001
Age (years)	0.01	1.01	1.00	1.02	0.107
<u>Alternative model (Model 2)</u>					
n=226					
Haemorrhagic	-1.19	0.31	0.21	0.44	<0.001
Non-haemorrhagic	0	1.00			
Age (years)	0.005	1.01	1.00	1.02	0.369

^a hazard ratios obtained from the inverse multivariable Cox PH regression ^b standard error for hazard ratio. ^c confidence interval. Alive = failure cases (coded = 1), Dead = censor cases (coded = 0)

Table H.3 The crude hazard ratios (HR) and adjusted HR (adjusted for age then sex), standard errors and 95% confidence intervals for HR estimated using the Cox proportional hazard regression model. The haemorrhagic stroke (HS) was compared against the cerebral infarction (CI as the baseline group). Only results for stroke subtypes are shown

Covariates		n	B	H	Lower 95% CI ^c		p-
stroke	H	1	0.	2.	1.53	4.44	<0.
	C	1	0	1.			
stroke	H	1	1.	2.	1.62	4.76	<0.
	C	1	0	1.			
stroke	H	1	0.	2.	1.53	4.44	<0.
	C	1	0	1.			

^a HR=Hazard ratio ^b SE=Standard error for HR ^c CI=confidence interval for HR ^d HS=Haemorrhagic stroke ^e CI=cerebral infarction. In Cox model, the dead = failure cases (coded =1) and the alive or at-own-risk (aor) discharge = censor cases (coded = 0)

Table H.4 The crude hazard ratios (HR) and adjusted HR (adjusted for age then sex), standard errors and 95% confidence intervals for HR estimated using the reversed Cox proportional hazard regression model. The haemorrhagic stroke (HS) was compared against the cerebral infarction (CI as the baseline group). Only results for stroke subtypes are shown

Covariates		n	B	H	Lower 95% CI ^c		p-
stroke	H	1	-	0.	0.30	0.55	<0.
	C	1	0	1			
stroke	H	1	-	0.	0.31	0.56	<0.
	C	1	0	1			
stroke	H	1	-	0.	0.30	0.52	<0.
	C	1	0	1			

^a HR=Hazard ratio ^b SE=Standard error for HR ^c CI=confidence interval for HR ^d HS=Haemorrhagic stroke ^e CI=cerebral infarction. In inverse Cox model, the alive or at-own-risk (aor) discharge = failure cases (coded =1) and the dead = censor cases (coded = 0)

Chapter 8 References

- ABOA-EBOULE, C., MENGUE, D., BENZENINE, E., HOMMEL, M., GIROUD, M., BEJOT, Y. & QUANTIN, C. 2013. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. *J Neurol*, 260, 605-13.
- ADAMS, H. 2007. *Principles of Cerebrovascular Disease*: Mc Graw Hill: New York.
- AHMED ABD RABOU & FRANK GALLARD. 2012. *Ischaemic stroke* [Online]. Available: <http://radiopaedia.org/articles/ischaemic-stroke> [Accessed 12 Dec 2012].
- AKHAVAN HEJAZI, S. M., MAZLAN, M., ABDULLAH, S. J. & ENKASAN, J. P. 2015. Cost of post-stroke outpatient care in Malaysia. *Singapore Med J*, 56, 116-9.
- AL-SAEED, A. H., CONSTANTINO, M. I., MOLYNEAUX, L., D'SOUZA, M., LIMACHER-GISLER, F., LUO, C., WU, T., TWIGG, S. M., YUE, D. K. & WONG, J. 2016. An Inverse Relationship Between Age of Type 2 Diabetes Onset and Complication Risk and Mortality: The Impact of Youth-Onset Type 2 Diabetes. *Diabetes Care*, 39, 823-9.
- ALI, M. F., ABDUL AZIZ, A. F., RASHID, M. R., CHE MAN, Z., AMIR, A. A., LIM, Y. S., RAMLI, N. S. & ZAINAL ABIDIN, N. A. 2015. Usage of Traditional and Complementary Medicine (T & CM): Prevalence, Practice and Perception among Post Stroke Patients Attending Conventional Stroke Rehabilitation in A Teaching Hospital in Malaysia. *Med J Malaysia*, 70, 18-23.
- AMARENCO, P., BOGOUSSLAVSKY, J., CAPLAN, L. R., DONNAN, G. A. & HENNERICI, M. G. 2009. Classification of stroke subtypes. *Cerebrovasc Dis*, 27, 493-501.
- ANDERSEN, K. K., ANDERSEN, Z. J. & OLSEN, T. S. 2011. Predictors of early and late case-fatality in a nationwide Danish study of 26,818 patients with first-ever ischemic stroke. *Stroke*, 42, 2806-12.
- ANDERSEN, K. K., OLSEN, T. S., DEHLENDORFF, C. & KAMMERGAARD, L. P. 2009. Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. *Stroke*, 40, 2068-72.
- ANDERSEN, M. N., ANDERSEN, K. K., KAMMERGAARD, L. P. & OLSEN, T. S. 2005. Sex differences in stroke survival: 10-year follow-up of the Copenhagen stroke study cohort. *J Stroke Cerebrovasc Dis*, 14, 215-20.

- APPELROS, P., NYDEVIK, I. & VIITANEN, M. 2003. Poor outcome after first-ever stroke: predictors for death, dependency, and recurrent stroke within the first year. *Stroke*, 34, 122-6.
- APPELROS, P., STEGMAYR, B. & TARENT, A. 2009. Sex differences in stroke epidemiology: a systematic review. *Stroke*, 40, 1082-90.
- ARBOIX, A. 2015. Cardiovascular risk factors for acute stroke: Risk profiles in the different subtypes of ischemic stroke. *World J Clin Cases*, 3, 418-29.
- ARONOWSKI, J. & ZHAO, X. 2011. Molecular pathophysiology of cerebral hemorrhage: secondary brain injury. *Stroke*, 42, 1781-6.
- ASIAN ACUTE STROKE ADVISORY PANEL. 2000. Stroke epidemiological data of nine Asian countries. Asian Acute Stroke Advisory Panel (AASAP). *J Med Assoc Thai*, 83, 1-7.
- ASPLUND, K. 1996. Stroke in Europe: widening gap between East and West. *Cerebrovascular Diseases*, 6, 3-6.
- AUCHINCLOSS, A. H., GEBREAB, S. Y., MAIR, C. & DIEZ ROUX, A. V. 2012. A review of spatial methods in Epidemiology, 2000-2010. *Annual Rev Public Health*, 33, 107-22.
- AZIZ, Z. A., LEE, Y. Y., NGAH, B. A., SIDEK, N. N., LOOI, I., HANIP, M. R. & BASRI, H. B. 2015. Acute Stroke Registry Malaysia, 2010-2014: Results from the National Neurology Registry. *J Stroke Cerebrovasc Dis*, 24, 2701-9.
- AZIZ, Z. A., LEE, Y. Y., SIDEK, N. N., NGAH, B. A., LOOI, I., HANIP, M. R. & BASRI, H. B. 2016. Gender disparities and thrombolysis use among patient with first-ever ischemic stroke in Malaysia. *Neurol Res*, 38, 406-13.
- BABAR, Z. D. & IZHAM, M. I. 2009. Effect of privatization of the drug distribution system on drug prices in Malaysia. *Public Health*, 123, 523-33.
- BALU, S. 2009. Differences in psychometric properties, cut-off scores, and outcomes between the Barthel Index and Modified Rankin Scale in pharmacotherapy-based stroke trials: systematic literature review. *Curr Med Res Opin*, 25, 1329-41.
- BANDERA, E. V., CHANDRAN, U., ZIRPOLI, G., MCCANN, S. E., CIUPAK, G. & AMBROSONE, C. B. 2013. Rethinking sources of representative controls for the conduct of case-control studies in minority populations. *BMC Med Res Methodol*, 13, 71, doi: 10.1186/1471-2288-13-71.

- BANKS, J. L. & MAROTTA, C. A. 2007. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. *Stroke*, 38, 1091-6.
- BARBER, M. & STOTT, D. J. 2004. Validity of the Telephone Interview for Cognitive Status (TICS) in post-stroke subjects. *Int J Geriatr Psychiatry*, 19, 75-9.
- BARRIBALL, K. L., CHRISTIAN, S. L., WHILE, A. E. & BERGEN, A. 1996. The telephone survey method: a discussion paper. *J Adv Nurs*, 24, 115-21.
- BARTLETT, J. 2016. *Multiple imputation for informative censoring R package* [Online]. Available: <http://thestatsgeek.com/2016/07/28/multiple-imputation-for-informative-censoring-r-package/> [Accessed 26 Jan 2017].
- BASRI, H. & AZMAN ALI, R. 2003. Predictors of in-hospital mortality after an acute ischaemic stroke. *Neurol J Southeast Asia*, 8, 5-8.
- BASU, S., ANDREWS, J., KISHORE, S., PANJABI, R. & STUCKLER, D. 2012. Comparative performance of private and public healthcare systems in low- and middle-income countries: a systematic review. *PLoS Med*, 9, e1001244.
- BERNARDINI, G. L. & DESHAIES, E. M. 2001. Critical care of intracerebral and subarachnoid hemorrhage. *Curr Neurol Neurosci Rep*, 1, 568-76.
- BHALLA, A., WANG, Y., RUDD, A. & WOLFE, C. D. 2013. Differences in outcome and predictors between ischemic and intracerebral hemorrhage: the South London Stroke Register. *Stroke*, 44, 2174-81.
- BHATTACHARJEE, M., AHMED, S., RAHMAN, K. M., UDDIN, M. J., DEY, S. K., ISRAIL, A., MONDAL, G., MIAH, A. H. & ROY, G. C. 2011. Sex differentiation and risk factor evaluation in stroke patients. *Mymensingh Med J*, 20, 371-6.
- BLAND, J. M. & ALTMAN, D. G. 1998. Survival probabilities (the Kaplan-Meier method). *BMJ*, 317, 1572.
- BONITA, R. 1992. Epidemiology of stroke. *Lancet*, 339, 342-4.
- BRASS, L. M. & KERNAN, W. N. 1989. Canadian Neurological Scale. *Neurology*, 39, 1556-8.
- BROTT, T., BRODERICK, J., KOTHARI, R., BARSAN, W., TOMSICK, T., SAUERBECK, L., SPILKER, J., DULDNER, J. & KHOURY, J. 1997. Early hemorrhage growth in patients with intracerebral hemorrhage. *Stroke*, 28, 1-5.

- BROTT, T., THALINGER, K. & HERTZBERG, V. 1986. Hypertension as a risk factor for spontaneous intracerebral hemorrhage. *Stroke*, 17, 1078-1083.
- BROWN, J. R., COCHRAN, R. P., LEAVITT, B. J., DACEY, L. J., ROSS, C. S., MACKENZIE, T. A., KUNZELMAN, K. S., KRAMER, R. S., HERNANDEZ, F., JR., HELM, R. E., WESTBROOK, B. M., DUNTON, R. F., MALENKA, D. J. & O'CONNOR, G. T. 2007. Multivariable prediction of renal insufficiency developing after cardiac surgery. *Circulation*, 116, 1139-43.
- BRUNO, A., SHAH, N., LIN, C., CLOSE, B., HESS, D. C., DAVIS, K., BAUTE, V., SWITZER, J. A., WALLER, J. L. & NICHOLS, F. T. 2010. Improving modified Rankin Scale assessment with a simplified questionnaire. *Stroke*, 41, 1048-50.
- BURD, R. S., JANG, T. S. & NAIR, S. S. 2006. Predicting hospital mortality among injured children using a national trauma database. *J Trauma*, 60, 792-801.
- BURING, J. E. 2008. Primary data collection: what should well-trained epidemiology doctoral students be able to do? *Epidemiology*, 19, 347-9.
- BURKE, T. A. & VENKETASUBRAMANIAN, R. N. 2006. The epidemiology of stroke in the East Asian region: A literature-based review. *Int J Stroke*, 1, 208-15.
- BURKOFF, N. S., METCALFE, P., BARTLETT, J. & RUAU, D. 2016a. Gamma imputation tutorial (Jackson 2014). Available: https://cran.rproject.org/web/packages/InformativeCensoring/vignettes/gamma_imputation_Jackson_2014.pdf
- BURKOFF, N. S., METCALFE, P., BARTLETT, J. & RUAU, D. 2016b. Risk Score Imputation tutorial (Hsu 2009). Available: https://cran.rproject.org/web/packages/InformativeCensoring/vignettes/risk_score_imputation_Hsu_2009.pdf.
- BURNHAM, K. P. & ANDERSON, D. R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Fort Collins, Colorado, USA.: Springer.
- BUSHNELL, C. D. 2008. Stroke in women: risk and prevention throughout the lifespan. *Neurol Clin*, 26, 1161-76, xi.
- CAPLAN, L. R. 2011. Stroke classification: A personal view. *Stroke*, 42, S3-6.
- CARR, E. C. J. & WORTH, A. 2001. The use of the telephone interview for research. *Nursing Times Research*, 6, 511-24.

- CARSON, A. P., MUNTNER, P., KISSELA, B. M., KLEINDORFER, D. O., HOWARD, V. J., MESCHIA, J. F., WILLIAMS, L. S., PRINEAS, R. J., HOWARD, G. & SAFFORD, M. M. 2012. Association of prediabetes and diabetes with stroke symptoms: the REasons for Geographic and Racial Differences in Stroke (REGARDS) study. *Diabetes Care*, 35, 1845-52.
- CHEN, Y. D., LI, S. J., SUN, F. H., LIU, Y. Y. & HU, W. L. 2011. Monitoring of medical complications after acute ischemic stroke in a neurological intensive care unit. *Eur Neurol*, 66, 204-9.
- CHENG, Y., CHEN, K.-J., WANG, C.-J., CHAN, S.-H., CHANG, W.-C. & CHEN, J.-H. 2005. Secular trends in coronary heart disease mortality, hospitalization rates, and major cardiovascular risk factors in Taiwan, 1971-2001. *Int J Cardiol*, 100, 47-52.
- CHINDAPRASIRT, J., SAWANYAWISUTH, K., CHATTAKUL, P., LIMPAWATTANA, P., TIAMKAO, S., AOUNTRI, P. & CHOTMONGKOL, V. 2013. Age predicts functional outcome in acute stroke patients with rt-PA treatment. *ISRN Neurol*, 2013, doi: 10.1155/2013/710681.
- CHRISTENSEN, M. C., BRODERICK, J., VINCENT, C., MORRIS, S. & STEINER, T. 2009. Global differences in patient characteristics, case management and outcomes in intracerebral hemorrhage: the Factor Seven for Acute Hemorrhagic Stroke (FAST) trial. *Cerebrovasc Dis*, 28, 55-64.
- CHRISTOPHER, R., NAGARAJA, D. & SHANKAR, S. K. 2007. Homocysteine and cerebral stroke in developing countries. *Curr Med Chem*, 14, 2393-401.
- CINCURA, C., PONTES-NETO, O. M., NEVILLE, I. S., MENDES, H. F., MENEZES, D. F., MARIANO, D. C., PEREIRA, I. F., TEIXEIRA, L. A., JESUS, P. A., DE QUEIROZ, D. C., PEREIRA, D. F., PINTO, E., LEITE, J. P., LOPES, A. A. & OLIVEIRA-FILHO, J. 2009. Validation of the National Institutes of Health Stroke Scale, modified Rankin Scale and Barthel Index in Brazil: the role of cultural adaptation and structured interviewing. *Cerebrovasc Dis*, 27, 119-22.
- CIONCOLONI, D., PIU, P., TASSI, R., ACAMPA, M., GUIDERI, F., TADDEI, S., BIELLI, S., MARTINI, G. & MAZZOCCHIO, R. 2012. Relationship between the modified Rankin Scale and the Barthel Index in the process of functional recovery after stroke. *NeuroRehabilitation*, 30, 315-22.
- CLEVES, M., GOULD, W., GUTIERREZ, R. & MARCHENKO, Y. 2010. *An Introduction to Survival Analysis Using Stata*, Texas, USA: StataCorp LP.
- COELHO, B. P. & GIRALDI-GUIMARAES, A. 2014. Effect of age and gender on recovery after stroke in rats treated with bone marrow mononuclear cells. *Neurosci Res*, 88, 67-73.

- COGGON, D., BARKER, D. & ROSE, G. 2003. *Epidemiology For The Uninitiated*, Chennai, India: BMJ books.
- COX, A. M., MCKEVITT, C., RUDD, A. G. & WOLFE, C. D. A. 2006. Socioeconomic status and stroke. *Lancet Neurol*, 5, 181-8.
- COX, D. 1986. Citation-Classic - Regression-Models and Life-Tables. *Current Contents/Agriculture Biology & Environmental Sciences*, 16-16.
- CZLONKOWSKA, A., NIEWADA, M., SALEH EL-BARONI, I., MENDEL, T., RYGLEWICZ, D., SANDERCOCK, P. & LEWIS, S. 2002. High early case fatality after ischaemic stroke in Poland: exploration of possible explanations in the International Stroke Trial. *J Neurol Sci*, 202, 53-7.
- DANS, A., NG, N., VARGHESE, C., TAI, E. S., FIRESTONE, R. & BONITA, R. 2011. The rise of chronic non-communicable diseases in Southeast Asia: Time for action. *Lancet*, 377, 680-9.
- DAWODU, C. O. & DANESI, M. A. 2008. Relationship of National Institute of Health Stroke Scores [NIHSS] to 90 days mortality in Africa. *Niger Postgrad Med J*, 15, 259-63.
- DEL BRUTTO, O. H., MONTALVAN, M., TETTAMANTI, D., PENAHERRERA, E., SANTIBANEZ, R., POW-CHON-LONG, F. & DEL BRUTTO, V. J. 2013. The "know your numbers" program in Atahualpa--a pilot study aimed to reduce cardiovascular diseases and stroke burden in rural communities of developing countries. *Int J Cardiol*, 168, 3123-4.
- DELBARI, A., SALMAN ROGHANI, R., TABATABAEI, S. S., RAHGOZAR, M. & LOKK, J. 2011. Stroke epidemiology and one-month fatality among an urban population in Iran. *Int J Stroke*, 6, 195-200.
- DELGADO-RODRIGUEZ, M. & LLORCA, J. 2004. Bias. *J Epidemiol Community Health*, 58, 635-41.
- DELLA PIETRA, G. L., SAVIO, K., ODDONE, E., REGGIANI, M., MONACO, F. & LEONE, M. A. 2011. Validity and reliability of the Barthel index administered by telephone. *Stroke*, 42, 2077-9.
- DEPARTMENT OF STATISTIC. 2015. *Kelantan @ a Glance* [Online]. Department of Statistic, Malaysia. Available: https://www.statistics.gov.my/index.php?r=column/cone&menu_id=RU84WGQxYkVPeVpodUZtTkpPdnBmZz09 [Accessed 5 July 2014].
- DEPARTMENT OF STATISTICS. 2015. *Malaysia @ a Glance* [Online]. Department of Statistics, Malaysia. Available:

https://www.statistics.gov.my/index.php?r=column/cone&menu_id=dDM2enNvM09oTGtQemZPVzRTWENmZz09 [Accessed 5 July 2014].

- DEVEBER, G. 2003. Arterial ischemic strokes in infants and children: An overview of current approaches. *Semin Thromb Hemost*, 29, 567-73.
- DEVRIES, D., ZHANG, Y., QU, M., MA, J. & LIN, G. 2013. Gender difference in stroke case fatality: an integrated study of hospitalization and mortality. *J Stroke Cerebrovasc Dis*, 22, 931-937.
- DI CARLO, A., LAMASSA, M., BALDERESCHI, M., PRACUCCI, G., CONSOLI, D., WOLFE, C. D., GIROUD, M., RUDD, A., BURGER, I., GHETTI, A., INZITARI, D. & EUROPEAN, B. S. O. S. C. G. 2006. Risk factors and outcome of subtypes of ischemic stroke. Data from a multicenter multinational hospital-based registry. The European Community Stroke Project. *J Neurol Sci*, 244, 143-50.
- DIAMOND, P., GALE, S. & STEWART, K. 2003. Primary intracerebral haemorrhage--clinical and radiologic predictors of survival and functional outcome. *Disabil Rehabil*, 25, 689-98.
- DIGGLE, P. J., HEAGERTY, P., LIANG, K. & ZEGER, S. 2002. *Analysis of Longitudinal Data*, Oxford, UK: Oxford University Press.
- DIGGLE, P. J. & RIBEIRO, P. J. 2007. *Model-based Geostatistics*: Springer, London
- DOBSON, A. J. 2002. *An Introduction To Generalized Linear Models*: Chapman & Hall/CRC, New York
- DONABEDIAN, A. 1988. The quality of care. How can it be assessed? *JAMA*, 260, 1743-8.
- DONNAN, G. A. 2011. Stroke in developing countries: The looming epidemic. *Int J Stroke*, 6, 97.
- DREFAHL, S., LUNDSTRÖM, H., MODIG, K. & AHLBOM, A. 2012. The era of centenarians: Mortality of the oldest old in Sweden. *J Intern Med*, 272, 100-2.
- DRIVER, J. A., DJOUSSÉ, L., LOGROSCINO, G., GAZIANO, J. M. & KURTH, T. 2008. Incidence of cardiovascular disease and cancer in advanced age: prospective cohort study. *BMJ*, 337, doi:10.1136/bmj.a2467.
- DUFFY, L., GAJREE, S., LANGHORNE, P., STOTT, D. J. & QUINN, T. J. 2013. Reliability (inter-rater agreement) of the Barthel Index for assessment of stroke survivors: systematic review and meta-analysis. *Stroke*, 44, 462-8.

- DURAI PANDIAN, J., PADMA, V., VIJAYA, P., SYLAJA, P. N. & MURTHY, J. M. 2007. Stroke and thrombolysis in developing countries. *Int J Stroke*, 2, 17-26.
- EASTERN STROKE CORONARY HEART DISEASE COLLABORATIVE RESEARCH GROUP. 1998. Blood pressure, cholesterol, and stroke in eastern Asia. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. *Lancet*, 352, 1801-7.
- ELLEKJAER, H., HOLMEN, J., KRUGER, O. & TERENT, A. 1999. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. *Stroke*, 30, 56-60.
- ELLUL, J., WATKINS, C. & BARER, D. 1998. Estimating total Barthel scores from just three items: the European Stroke Database 'minimum dataset' for assessing functional status at discharge from hospital. *Age Ageing*, 27, 115-22.
- ERIKSSON, M., ASPLUND, K., VAN ROMPAYE, B. & ELIASSON, M. 2013. Differences in cardiovascular risk factors and socioeconomic status do not explain the increased risk of death after a first stroke in diabetic patients: results from the Swedish Stroke Register. *Diabetologia*, 56, 2181-6.
- ESRI 2010. ArcGIS 10.2. 10.2 ed. Redlands, CA: ESRI
- FALCONE, G. F., BROUWERS, H. B., BIFFI, A., ANDERSON, C. D., BATTEY, T. W. K., AYRES, A. M., VASHKEVICH, A., SCHWAB, K., ROST, N. S., GOLDSTEIN, J. N., VISWANATHAN, A., GREENBERG, S. M. & ROSAND, J. 2013a. Predictors of hematoma volume in Primary Infratentorial Intracerebral Hemorrhage. *Cerebrovascular Diseases*, 35, 807.
- FALCONE, G. J., BIFFI, A., BROUWERS, H. B., ANDERSON, C. D., BATTEY, T. W., AYRES, A. M., VASHKEVICH, A., SCHWAB, K., ROST, N. S., GOLDSTEIN, J. N., VISWANATHAN, A., GREENBERG, S. M. & ROSAND, J. 2013b. Predictors of hematoma volume in deep and lobar Supratentorial Intracerebral Hemorrhage. *JAMA Neurol*, 70, 988-94.
- FARAWAY, J. F. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Non-parametric Regression Models*, FL 33487–2742, USA: Chapman & Hall/CRC.
- FEIGIN, V. L. 2007. Stroke in developing countries: can the epidemic be stopped and outcomes improved? *Lancet Neurol*, 6, 94-7.
- FEIGIN, V. L., FOROUZANFAR, M. H., KRISHNAMURTHI, R., MENSAH, G. A., CONNOR, M., BENNETT, D. A., MORAN, A. E., SACCO, R. L., ANDERSON, L., TRUELSEN, T., O'DONNELL, M., VENKETASUBRAMANIAN, N., BARKER-COLLO, S., LAWES, C. M., WANG, W., SHINOHARA, Y., WITT, E., EZZATI, M., NAGHAVI, M., MURRAY, C., GLOBAL BURDEN OF DISEASES, I., RISK FACTORS, S. & THE, G. B. D. S. E. G. 2014. Global and regional burden

of stroke during 1990-2010: Findings from the Global Burden of Disease Study 2010. *Lancet*, 383, 245-54.

FEIGIN, V. L., LAWES, C. M. M., BENNETT, D. A. & ANDERSON, C. S. 2003. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol*, 2, 43-53.

FEIGIN, V. L., LAWES, C. M. M., BENNETT, D. A., BARKER-COLLO, S. L. & PARAG, V. 2009. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: A systematic review. *Lancet Neurol*, 8, 355-69.

FEIGIN, V. L., RINKEL, G. J., LAWES, C. M., ALGRA, A., BENNETT, D. A., VAN GIJN, J. & ANDERSON, C. S. 2005. Risk factors for subarachnoid hemorrhage: an updated systematic review of epidemiological studies. *Stroke*, 36, 2773-80.

FERNANDES, T. G., GOULART, A. C., CAMPOS, T. F., LUCENA, N. M. G., FREITAS, K. L. A., TREVISAN, C. M., BENSEÑOR, I. M. & LOTUFO, P. A. 2012. Early stroke case-fatality rates in three hospital registries in the Northeast and Southeast of Brazil. *Arq Neuropsiquiatr*, 70, 869-73.

FRONTERA, J. A., AHMED, W., ZACH, V., JOVINE, M., TANENBAUM, L., SEHBA, F., PATEL, A., BEDERSON, J. B. & GORDON, E. 2015. Acute ischaemia after subarachnoid haemorrhage, relationship with early brain injury and impact on outcome: a prospective quantitative MRI study. *J Neurol Neurosurg Psychiatry*, 86, 71-8.

FUENTES, B. & TEJEDOR, E. D. 2014. Stroke: The worldwide burden of stroke--a blurred photograph. *Nat Rev Neurol*, 10, 127-8.

GAZIANO, J. M. 2010. Fifth phase of the epidemiologic transition: the age of obesity and inactivity. *JAMA*, 303, 275-6.

GENELETTI, S., RICHARDSON, S. & BEST, N. 2009. Adjusting for selection bias in retrospective, case-control studies. *Biostatistics*, 10, 17-31.

GIRALDO, E. A. 2015a. *Intracerebral Hemorrhage* [Online]. Available: <http://www.merckmanuals.com/professional/neurologic-disorders/stroke-cva/intracerebral-hemorrhage> [Accessed 31 May 2015].

GIRALDO, E. A. 2015b. *Ischemic Stroke* [Online]. Merck Available: <http://www.merckmanuals.com/professional/neurologic-disorders/stroke-cva/ischemic-stroke> [Accessed 31 May 2015].

- GIRALDO, E. A. 2015c. *Subarachnoid Hemorrhage (SAH)* [Online]. Merck. Available: <http://www.merckmanuals.com/professional/neurologic-disorders/stroke-cva/subarachnoid-hemorrhage-sah> [Accessed 31 May 2015].
- GLANTZ, S. A. 2005. *Primer of Biostatistics*, New York, USA: The McGraw-Hill Companies. New York.
- GO, S. 2008. Validity of telephone interview for functional outcome in post-stroke subjects *Alzheimer's and Dementia*, 4, T452.
- GOEL, M. K., KHANNA, P. & KISHORE, J. 2010. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res*, 1, 274-8.
- GOOD, D. C. 1990. Cerebrovascular diseases. In: WALKER, H. K., HALL, W. D. & HURST, J. W. (eds.) *Clinical Methods: The History, Physical, and Laboratory Examinations*. 3rd ed. Boston, Butterworths.
- GORDIS, L. 2009. *Epidemiology*, Philadelphia, USA: Elsevier Saunders.
- GORGUI, J., GORSHKOV, M., KHAN, N. & DASKALOPOULOU, S. S. 2014. Hypertension as a Risk Factor for Ischemic Stroke in Women. *Canadian Journal of Cardiology*, 30, 774-782.
- GRIMES, D. A. & SCHULZ, K. F. 2005. Compared to what? Finding controls for case-control studies. *Lancet*, 365, 1429-33.
- GROPEN, T., MAGDON-ISMAIL, Z., DAY, D., MELLUZZO, S., SCHWAMM, L. H. & GROUP, N. E. C. C. A. 2009. Regional implementation of the stroke systems of care model: recommendations of the northeast cerebrovascular consortium. *Stroke*, 40, 1793-802.
- GRUBE, M. M., KOENNECKE, H. C., WALTER, G., THUMMLER, J., MEISEL, A., WELLWOOD, I., HEUSCHMANN, P. U. & BERLIN STROKE, R. 2012. Association between socioeconomic status and functional impairment 3 months after ischemic stroke: the Berlin Stroke Register. *Stroke*, 43, 3325-30.
- GUNATHILAKE, R., KRISHNAMURTHY, V., OLDMEADOW, C., KERR, E., PADMAKUMAR, C., ATTIA, J., PARSONS, M. & LEVI, C. 2014. Relationships between age, other predictive variables, and the 90-day functional outcome after intravenous thrombolysis for acute ischemic stroke. *Int J Stroke*, 9, E36-7.
- HACKING, C. & GAILLARD, F. 2012a. *Acute subdural haemorrhage* [Online]. Available: <http://radiopaedia.org/articles/subdural-haemorrhage> [Accessed 13 April 2012].

- HACKING, C. & GAILLARD, F. 2012b. *Extradural haemorrhage* [Online]. Available: <http://radiopaedia.org/articles/extradural-haemorrhage> [Accessed 11 Feb 2012].
- HAGHGOO, H. A., PAZUKI, E. S., HOSSEINI, A. S. & RASSAFIANI, M. 2013. Depression, activities of daily living and quality of life in patients with stroke. *J Neurol Sci*, 328, 87-91.
- HAMIDON, B. B. & RAYMOND, A. A. 2003. Risk factors and complications of acute ischaemic stroke patients at Hospital Universiti Kebangsaan Malaysia (HUKM). *Med J Malaysia*, 58, 499-505.
- HANLEY, J. A. & MCNEIL, B. J. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 143, 29-36.
- HAO, Z., WU, B., WANG, D., LIN, S., TAO, W. & LIU, M. 2013. A cohort study of patients with anemia on admission and fatality after acute ischemic stroke. *J Clin Neurosci*, 20, 37-42.
- HARRISON, J. K., MCARTHUR, K. S. & QUINN, T. J. 2013. Assessment scales in stroke: clinimetric and clinical considerations. *Clin Interv Aging*, 8, 201-11.
- HASTIE, T. J. & TIBSHIRANI, R. J. 1990. *Generalized Additive Models*, New York: Chapman & Hall.
- HATA, J. & KIYOHARA, Y. 2013. Epidemiology of stroke and coronary artery disease in Asia. *Circ J*, 77, 1923-32.
- HATANO, S. 1976. Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ*, 54, 541-53.
- HAVULINNA, A. S., PÄÄKKÖNEN, R., KARVONEN, M. & SALOMAA, V. 2008. Geographic patterns of incidence of ischemic stroke and acute myocardial infarction in Finland during 1991-2003. *Ann Epidemiol*, 18, 206-13.
- HEALTH TECHNOLOGY ASSESSMENT UNIT. 2006. Technology Report: Electronic Medical Record. Available: <http://www.moh.gov.my/attachments/6392.pdf>.
- HEBEL, K., BIENIASZEWSKI, L. & KOWALEWSKI, W. 2014. Health education for stroke patient carers: does it affect functional status improvement in patients after ischemic stroke? *Appl Nurs Res*, 27, e7-e12.

- HERNAN, M. A., HERNANDEZ-DIAZ, S., WERLER, M. M. & MITCHELL, A. A. 2002. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*, 155, 176-84.
- HISAMUDDIN, N. A., HAMZAH, M. S. & HOLLIMAN, C. J. 2007. Prehospital emergency medical services in Malaysia. *J Emerg Med*, 32, 415-21.
- HOFFMANN, T., WORRALL, L., EAMES, S. & RYAN, A. 2010. Measuring outcomes in people who have had a stroke and their carers: Can the telephone be used? *Topics in Stroke Rehabilitation*, 17, 119-27.
- HONG, C. Y., CHIA, K. S., HUGHES, K. & LING, S. L. 2004. Ethnic differences among Chinese, Malay and Indian patients with type 2 diabetes mellitus in Singapore. *Singapore Med J*, 45, 154-60.
- HONG, K. S., BANG, O. Y., KANG, D. W., YU, K. H., BAE, H. J., LEE, J. S., HEO, J. H., KWON, S. U., OH, C. W., LEE, B. C., KIM, J. S. & YOON, B. W. 2013. Stroke statistics in Korea: part I. Epidemiology and risk factors: a report from the Korean stroke society and clinical research center for stroke. *J Stroke*, 15, 2-20.
- HOSMER, D. W., LEMESHOW, S. & MAY, S. 2011. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, New York, USA: John Wiley & Sons.
- HOSMER, D. W., LEMESHOW, S. & STURDIVANT, R. X. 2013. *Applied Logistic Regression*, New York, USA.: John Wiley & Sons.
- HOWARD, G., CUSHMAN, M., HOWARD, V. J., KISSELA, B. M., KLEINDORFER, D. O., MOY, C. S., SWITZER, J. & WOO, D. 2013. Risk factors for intracerebral hemorrhage: the REasons for geographic and racial differences in stroke (REGARDS) study. *Stroke*, 44, 1282-7.
- HOWARD, G., PRINEAS, R., MOY, C., CUSHMAN, M., KELLUM, M., TEMPLE, E., GRAHAM, A. & HOWARD, V. 2006. Racial and geographic differences in awareness, treatment, and control of hypertension: the REasons for Geographic And Racial Differences in Stroke study. *Stroke*, 37, 1171-8.
- HOY, D. G., RAO, C., HOA, N. P., SUHARDI, S. & LWIN, A. M. 2013. Stroke mortality variations in South-East Asia: empirical evidence from the field. *Int J Stroke*, 8 Suppl A100, 21-7.
- HSU, C. H. & TAYLOR, J. M. 2009. Nonparametric comparison of two survival functions with dependent censoring via nonparametric multiple imputation. *Stat Med*, 28, 462-75.

- HUHTAKANGAS, J., LOPPONEN, P., TETRI, S., JUVELA, S., SALOHEIMO, P., BODE, M. K. & HILLBOM, M. 2013. Predictors for recurrent primary intracerebral hemorrhage: a retrospective population-based study. *Stroke*, 44, 585-90.
- HUNT, B. R., DEOT, D. & WHITMAN, S. 2014. Stroke mortality rates vary in local communities in a metropolitan area: racial and spatial disparities and correlates. *Stroke*, 45, 2059-65.
- HYUN, K. K., HUXLEY, R. R., ARIMA, H., WOO, J., LAM, T. H., UESHIMA, H., FANG, X., PETERS, S. A., JEE, S. H., GILES, G. G., BARZI, F. & WOODWARD, M. 2013. A comparative analysis of risk factors and stroke risk for Asian and non-Asian men: the Asia Pacific Cohort Studies Collaboration. *Int J Stroke*, 8, 606-11.
- INGALL, T., ASPLUND, K., MÄHÖNEN, M. & BONITA, R. 2000. A multinational comparison of subarachnoid hemorrhage epidemiology in the WHO MONICA stroke study. *Stroke*, 31, 1054-1061.
- JACKSON, D., WHITE, I. R., SEAMAN, S., EVANS, H., BAISLEY, K. & CARPENTER, J. 2014. Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation. *Stat Med*, 33, 4681-94.
- JANSSEN, P. M., VISSER, N. A., DORHOUT MEES, S. M., KLIJN, C. J., ALGRA, A. & RINKEL, G. J. 2010. Comparison of telephone and face-to-face assessment of the modified Rankin Scale. *Cerebrovasc Dis*, 29, 137-9.
- JAYA, F., WIN, M. N., ABDULLAH, M. R. & ABDULLAH, J. M. 2002. Stroke patterns in Northeast Malaysia: a hospital-based prospective study. *Neuroepidemiology*, 21, 28-35.
- JEWELL, N. P. 2004. *Statistics for Epidemiology*, Boca Raton, FL: CRC Press.
- JIA, Q., LIU, L. & WANG, Y. 2011. Risk factors and prevention of stroke in the Chinese population. *J Stroke Cerebrovasc Dis*, 20, 395-400.
- JOFFE, M. M. 2003. Invited commentary: screening as a nuisance variable in cancer epidemiology: methodological considerations. *Am J Epidemiol*, 157, 962-4.
- JOHNSTON, S. C., MENDIS, S. & MATHERS, C. D. 2009. Global variation in stroke burden and mortality: Estimates from monitoring, surveillance, and modelling. *Lancet Neurol*, 8, 345-54.
- KALANTRI, A. & KALANTRI, S. 2010. Distinguishing hemorrhagic stroke from ischemic stroke. *JAMA*, 304, 1327-8; author reply 1328.

- KAMARI, Z. 2009. Hospital Universiti Sains Malaysia (HUSM): 25 Years Of Excellent Service. *Mal J Med Sci*, 16, 16-24.
- KATZ, M. H. 2011. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers*, Cambridge, UK: Cambridge University Press.
- KAUL, S., BANDARU, V. C., SUVARNA, A. & BODDU, D. B. 2009. Stroke burden and risk factors in developing countries with special reference to India. *J Indian Med Assoc*, 107, 358, 367-70.
- KAZUI, S., NARITOMI, H., YAMAMOTO, H., SAWADA, T. & YAMAGUCHI, T. 1996. Enlargement of spontaneous intracerebral hemorrhage incidence and time course. *Stroke*, 27, 1783-7.
- KEIR, S. L., WARDLAW, J. M. & WARLOW, C. P. 2002. Stroke epidemiology studies have underestimated the frequency of intracerebral haemorrhage. A systematic review of imaging in epidemiological studies. *J Neurol*, 249, 1226-31.
- KHAN, M., AHMED, B., AHMED, M., NAJEEB, M., RAZA, E., KHAN, F., MOIN, A., SHUJAAT, D., ARSHAD, A. & KAMAL, A. K. 2012. Functional, cognitive and psychological outcomes, and recurrent vascular events in Pakistani stroke survivors: a cross sectional study. *BMC Res Notes*, 5, doi: 10.1186/1756-0500-5-89.
- KHOO, K. F., TAN, H. J. & RAYMOND, A. A. 2013. Prevalence of depression in stroke patients with vascular dementia in Universiti Kebangsaan Malaysia Medical Center. *Med J Malaysia*, 68, 105-10.
- KIM, A. S. & JOHNSTON, S. C. 2011. Global variation in the relative burden of stroke and ischemic heart disease. *Circulation*, 124, 314-23.
- KIM, J. S. 2014. Stroke in Asia: A global disaster. *Int J Stroke*, 9, 856-7.
- KIM, K. H. 2009. Predictors of 30-day mortality and 90-day functional recovery after primary intracerebral hemorrhage : hospital based multivariate analysis in 585 patients. *J Korean Neurosurg Soc*, 45, 341-9.
- KITA, Y., TURIN, T. C., ICHIKAWA, M., SUGIHARA, H., MORITA, Y., TOMIOKA, N., RUMANA, N., OKAYAMA, A., NAKAMURA, Y., ABBOTT, R. D. & UESHIMA, H. 2009. Trend of stroke incidence in a Japanese population: Takashima Stroke Registry, 1990-2001. *Int J Stroke*, 4, 241-9.
- KLEINBAUM, D. G. & KLEIN, M. 2012. *Survival Analysis: A Self-Learning Text*, New York: Springer.

- KNAUS, W. A., ZIMMERMAN, J. E., WAGNER, D. P., DRAPER, E. A. & LAWRENCE, D. E. 1981. APACHE-Acute Physiology and Chronic Health Evaluation: A physiologically based classification system. *Crit Care Med*, 9, 591-7.
- KNOFLACH, M., MATOSEVIC, B., RUCKER, M., FURTNER, M., MAIR, A., WILLE, G., ZANGERLE, A., WERNER, P., FERRARI, J., SCHMIDAUER, C., SEYFANG, L., KIECHL, S., WILLEIT, J. & AUSTRIAN STROKE UNIT REGISTRY, C. 2012. Functional recovery after ischemic stroke - a matter of age: data from the Austrian Stroke Unit Registry. *Neurology*, 78, 279-85.
- KNOL, M. J. & VANDERWEELE, T. J. 2012. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol*, 41, 514-20.
- KOKOTAILO, R. A. & HILL, M. D. 2005. Coding of stroke and stroke risk factors using International Classification of Diseases, Revisions 9 and 10. *Stroke*, 36, 1776-81.
- KONG, K. H. & LEE, J. 2014. Temporal recovery of activities of daily living in the first year after ischemic stroke: a prospective study of patients admitted to a rehabilitation unit. *NeuroRehabilitation*, 35, 221-6.
- KORNER-BITENSKY, N. & WOOD-DAUPHINEE, S. 1995. Barthel Index information elicited over the telephone. Is it reliable? *Am J Phys Med Rehabil*, 74, 9-18.
- KOTON, S., TANNE, D., GREEN, M. S. & BORNSTEIN, N. M. 2009. Mortality and predictors of death 1 month and 3 years after first-ever ischemic stroke: data from the first national acute stroke Israeli survey (NASIS 2004). *Neuroepidemiology*, 34, 90-6.
- KRISHNAMURTHI, R. V., FEIGIN, V. L., FOROUZANFAR, M. H., MENSAH, G. A., CONNOR, M., BENNETT, D. A., MORAN, A. E., SACCO, R. L., ANDERSON, L. M., TRUELSEN, T., O'DONNELL, M., VENKETASUBRAMANIAN, N., BARKER-COLLO, S., LAWES, C. M., WANG, W., SHINOHARA, Y., WITT, E., EZZATI, M., NAGHAVI, M., MURRAY, C., GLOBAL BURDEN OF DISEASES, I. R. F. S. & GROUP, G. B. D. S. E. 2013. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet Glob Health*, 1, e259-81.
- KRISHNAMURTHI, R. V., MORAN, A. E., FOROUZANFAR, M. H., BENNETT, D. A., MENSAH, G. A., LAWES, C. M., BARKER-COLLO, S., CONNOR, M., ROTH, G. A., SACCO, R., EZZATI, M., NAGHAVI, M., MURRAY, C. J., FEIGIN, V. L., GLOBAL BURDEN OF DISEASES, I. & RISK FACTORS STUDY STROKE EXPERT, G. 2014. The global burden of hemorrhagic stroke: a summary of findings from the GBD 2010 study. *Glob Heart*, 9, 101-6.
- KUBO, M., KIYOHARA, Y., KATO, I., TANIZAKI, Y., ARIMA, H., TANAKA, K., NAKAMURA, H., OKUBO, K. & IIDA, M. 2003. Trends in the incidence, mortality, and survival rate of cardiovascular disease in a Japanese community: the Hisayama study. *Stroke*, 34, 2349-54.

- KUMAR, A., PRASAD, M., KATHURIA, P., NAIR, P., PANDIT, A. K., SAHU, J. K. & PRASAD, K. 2015. Low socioeconomic status is an independent risk factor for ischemic stroke: a case-control study in North Indian population. *Neuroepidemiology*, 44, 138-43.
- KURTH, T., GAZIANO, J. M., REXRODE, K. M., KASE, C. S., COOK, N. R., MANSON, J. E. & BURING, J. E. 2005. Prospective study of body mass index and risk of stroke in apparently healthy women. *Circulation*, 111, 1992-8.
- KWAKKEL, G. & KOLLEN, B. J. 2013. Predicting activities after stroke: what is clinically relevant? *Int J Stroke*, 8, 25-32.
- KWAKKEL, G., VEERBEEK, J. M., HARMEILING-VAN DER WEL, B. C., VAN WEGEN, E., KOLLEN, B. J. & EARLY PREDICTION OF FUNCTIONAL OUTCOME AFTER STROKE, I. 2010. Diagnostic accuracy of the Barthel Index for measuring activities of daily living outcome after ischemic hemispheric stroke: does early poststroke timing of assessment matter? *Stroke*, 42, 342-6.
- KWON, H. S., KIM, C., LEE, S. H., JUNG, K. H., KIM, Y. D., KWON, H. M., HEO, S. H., CHANG, D. I., KIM, B. J., KIM, J. M., KIM, H. Y. & KIM, Y. S. 2016. Protocol of the stroke in Korean Young Adults Study: A multicenter case-control study and prospective cohort study. *J Stroke Cerebrovasc Dis*, 25, 1503-8.
- LAURITSEN, J. M. 2000. EpiData Data Entry, Data Management and Basic Statistical Analysis System. Odense, Denmark: The EpiData Association.
- LAVADOS, P. M., SACKS, C., PRINA, L., ESCOBAR, A., TOSSI, C., ARAYA, F., FEUERHAKE, W., GALVEZ, M., SALINAS, R. & ALVAREZ, G. 2005. Incidence, 30-day case-fatality rate, and prognosis of stroke in Iquique, Chile: a 2-year community-based prospective study (PISCIS project). *Lancet*, 365, 2206-15.
- LEE, J., HENG, D., CHIA, K. S., CHEW, S. K., TAN, B. Y. & HUGHES, K. 2001. Risk factors and incident coronary heart disease in Chinese, Malay and Asian Indian males: the Singapore Cardiovascular Cohort Study. *Int J Epidemiol*, 30, 983-8.
- LEE, L. T., ALEXANDROV, A. W., HOWARD, V. J., KABAGAMBE, E. K., HESS, M. A., MCLAIN, R. M., SAFFORD, M. M. & HOWARD, G. 2014. Race, regionality and pre-diabetes in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study. *Prev Med*, 63, 43-7.
- LEE, S., SHAFE, A. C. E. & COWIE, M. R. 2011. UK stroke incidence, mortality and cardiovascular risk management 1999-2008: time-trend analysis from the General Practice Research Database. *BMJ Open*, 1, doi:10.1136/bmjopen-2011-000269.

- LI, F. & CHEN, Q. X. 2014. Risk factors for mental disorders in patients with hypertensive intracerebral hemorrhage following neurosurgical treatment. *J Neurol Sci*, 341, 128-32.
- LI, L., ZHANG, M. & HOLMAN, D. 2011. Population versus hospital controls for case-control studies on cancers in Chinese hospitals. *BMC Med Res Methodol*, 11, 167.
- LI, S., ZHAO, X., WANG, C., LIU, L., LIU, G., WANG, Y., WANG, C., JING, J. & WANG, Y.-J. 2013. Risk factors for poor outcome and mortality at 3 months after the ischemic stroke in patients with atrial fibrillation. *J Stroke Cerebrovasc Dis*, 22, e419-25.
- LIN, H.-C., LIN, Y.-J., LIU, T.-C., CHEN, C.-S. & CHIU, W.-T. 2007. Urbanization and stroke prevalence in Taiwan: Analysis of a nationwide survey. *J Urban Health*, 84, 604-14.
- LINDLEY, R. I. 2008. *Stroke*: Oxford University Press.
- LINDMARK, A., GLADER, E.-L., ASPLUND, K., NORRVING, B., ERIKSSON, M. & COLLABORATION, R.-S. 2013. Socioeconomic disparities in stroke case fatality - Observations from Riks-Stroke, the Swedish stroke register. *Int J Stroke*, 9, 429-36.
- LISABETH, L. & BUSHNELL, C. 2012. Stroke risk in women: the role of menopause and hormone therapy. *Lancet Neurol*, 11, 82-91.
- LISABETH, L. D., SANCHEZ, B. N., BAEK, J., SKOLARUS, L. E., SMITH, M. A., GARCIA, N., BROWN, D. L. & MORGENSTERN, L. B. 2014. Neurological, functional, and cognitive stroke outcomes in Mexican Americans. *Stroke*, 45, 1096-101.
- LIU, F. & MCCULLOUGH, L. D. 2012. Interactions between age, sex, and hormones in experimental ischemic stroke. *Neurochem Int*, 61, 1255-65.
- LIU, L., IKEDA, K. & YAMORI, Y. 2001. Changes in stroke mortality rates for 1950 to 1997: a great slowdown of decline trend in Japan. *Stroke*, 32, 1745-9.
- LIU, N., CADILHAC, D. A., ANDREW, N. E., ZENG, L., LI, Z., LI, J., LI, Y., YU, X., MI, B., LI, Z., XU, H., CHEN, Y., WANG, J., YAO, W., LI, K., YAN, F. & WANG, J. 2014. Randomized controlled trial of early rehabilitation after intracerebral hemorrhage stroke: difference in outcomes within 6 months of stroke. *Stroke*, 45, 3502-7.
- LONG, J. S. T. & FREESE, J. 2006. *Regression Models for Categorical Dependent Variables using Stata*, Texas, USA: StataCorp.
- LOO, K. W. & GAN, S. H. 2012. Burden of stroke in Malaysia. *Int J Stroke*, 7, 165-7.

- LOPEZ, A. D., MATHERS, C. D., EZZATI, M., JAMISON, D. T. & MURRAY, C. J. L. 2006. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *The Lancet*, 367, 1747-1757.
- LUVIZUTTO, G. J., MONTEIRO, T. A., BRAGA, G., PONTES-NETO, O. M., DE LIMA RESENDE, L. A. & BAZAN, R. 2012. Validation of the scandinavian stroke scale in a multicultural population in Brazil. *Cerebrovasc Dis Extra*, 2, 121-6.
- MACKAY, M., LEGG, K. & NEARING, S. 2007. Neurological assessment of the stroke patient: the Canadian Neurological Scale. *Axone*, 28, 9-10.
- MAHESWARAN, R., HAINING, R. P., BRINDLEY, P., LAW, J., PEARSON, T., FRYERS, P. R., WISE, S. & CAMPBELL, M. J. 2005. Outdoor air pollution and stroke in Sheffield, United Kingdom: a small-area level geographical study. *Stroke*, 36, 239-43.
- MAHONEY, F. I. & BARTHEL, D. W. 1965. Funtional evaluation : The Barthel Index. *Md State Med J*, 14, 61-5.
- MARTINS, S. C., FRIEDRICH, M. A., BRONDANI, R., DE ALMEIDA, A. G., DE ARAUJO, M. D., CHAVES, M. L., BERGER, J. R. & MASSARO, A. R. 2011. Thrombolytic therapy for acute stroke in the elderly: an emergent condition in developing countries. *J Stroke Cerebrovasc Dis*, 20, 459-64.
- MATHERS, C. D., LOPEZ, A. D. & MURRAY, C. J. L. 2006. *Global Burden of Disease and Risk Factors: The Burden of Disease and Mortality by Condition: Data, Methods, and Results for 2001*, New York, USA: The World Bank and Oxford University Press.
- MCCORMICK, N., BHOLE, V., LACAILE, D. & AVINA-ZUBIETA, J. A. 2015. Validity of diagnostic codes for acute stroke in administrative databases: A systematic review. *PLoS One*, 10, e0135834.
- MCNAUGHTON, H., WEATHERALL, M., TAYLOR, W. & MCPHERSON, K. 2001. Factors influencing rate of Barthel Index change in hospital following stroke. *Clin Rehabil*, 15, 422-7.
- MENSAH, G. A. 2008. Epidemiology of stroke and high blood pressure in Africa. *Heart*, 94, 697-705.
- MERICAN, M. I., ROHAIZAT, Y. & HANIZA, S. 2004. Developing the Malaysian health system to meet the challenges of the future. *Med J Malaysia*, 59, 84-93.
- MINISTRY OF HEALTH 2011. Annual Report: Ministry of Health 2011. Kuala Lumpur, Malaysia: Ministry of Health, Malaysia.

- MINISTRY OF HEALTH 2012. Health indicators 2012. Kuala Lumpur, Malaysia.
- MINISTRY OF HEALTH MALAYSIA. 2013a. *Caj Pesakit Luar (Outpatient Charge)* [Online]. Ministry of Health, Malaysia. Available: <http://www.moh.gov.my/index.php/pages/view/160> [Accessed 13 Dec 2016].
- MINISTRY OF HEALTH MALAYSIA. 2013b. *Caj Rawatan (Treatment Charge)* [Online]. Ministry of Health, Malaysia. Available: <http://www.moh.gov.my/index.php/pages/view/162> [Accessed 13 Dec 2016].
- MINISTRY OF HEALTH MALAYSIA. 2015. *Health Facts 2015* [Online]. Health Informatics Centre. Available: http://vlib.moh.gov.my/cms/documentstorage/com.tms.cms.document.Document_2e8efe25-a0188549-d5315d00-e9ca9560/KKM_HEALTH_FACTS_2015.pdf [Accessed 15 Dec 2016].
- MODIG, K., ANDERSSON, T., DREFAHL, S. & AHLBOM, A. 2013a. Age-specific trends in morbidity, mortality and case-fatality from cardiovascular disease, myocardial infarction and stroke in advanced age: evaluation in the Swedish population. *PLoS One*, 8, e64928.
- MODIG, K., DREFAHL, S. & AHLBOM, A. 2013b. Limitless longevity: Comment on the Contribution of rectangularization to the secular increase of life expectancy. *Int J Epidemiol*, 42, 914-6.
- MOHD ZULKIFLY, M. F., GHAZALI, S. E., CHE DIN, N. & SUBRAMANIAM, P. 2016. The influence of demographic, clinical, psychological and functional determinants on post-stroke cognitive impairment at day care stroke center, Malaysia. *Mal J Med Sci*, 23, 53-64.
- MUKHERJEE, D. & PATIL, C. G. 2011. Epidemiology and the global burden of stroke. *World Neurosurg*, 76, S85-90.
- NAKAO, S., TAKATA, S., UEMURA, H., KASHIHARA, M., OSAWA, T., KOMATSU, K., MASUDA, Y., OKAHISA, T., NISHIKAWA, K., KONDO, S., YAMADA, M., TAKAHARA, R., OGATA, Y., NAKAMURA, Y., NAGAHIRO, S., KAJI, R. & YASUI, N. 2010. Relationship between Barthel Index scores during the acute phase of rehabilitation and subsequent ADL in stroke patients. *J Med Invest*, 57, 81-8.
- NAKIBUUKA, J., SAJATOVIC, M., NANKABIRWA, J., FURLAN, A. J., KAYIMA, J., DDUMBA, E., KATABIRA, E. & BYAKIKA-TUSIIME, J. 2015. Stroke risk factors differ between rural and urban communities: population survey in Central Uganda. *Neuroepidemiology*, 44, 156-65.

- NASHEF, S. A., ROQUES, F., MICHEL, P., GAUDUCHEAU, E., LEMESHOW, S. & SALAMON, R. 1999. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg*, 16, 9-13.
- NEDELTCHEV, K., RENZ, N., KARAMESHEV, A., HAEFELI, T., BREKENFELD, C., MEIER, N., REMONDA, L., SCHROTH, G., ARNOLD, M. & MATTLE, H. P. 2010. Predictors of early mortality after acute ischaemic stroke. *Swiss Med Wkly*, 140, 254-9.
- NEELAMEGAM, M., LOOI, I., CHEAH, W. K., NARAYANAN, P., HAMID, A. M. & ONG, L. M. 2013. Stroke incidence in the South West District of the Penang Island, Malaysia: PEARLS: Penang Acute Stroke Research Longitudinal Study. *Prev Med*, 57 Suppl, S77-9.
- NETER, J. 1996. *Applied Linear Statistical Methods*, Boston, USA: McGraw-Hill.
- NEUPANE, B., WALTER, S. D., KRUEGER, P. & LOEB, M. 2010. Community controls were preferred to hospital controls in a case-control study where the cases are derived from the hospital. *J Clin Epidemiol*, 63, 926-31.
- NIEWADA, M., KOBAYASHI, A., SANDERCOCK, P. A. G., KAMIŃSKI, B., CZŃONKOWSKA, A. & GROUP, I. S. T. C. 2005. Influence of gender on baseline features and clinical outcomes among 17,370 patients with confirmed ischaemic stroke in the international stroke trial. *Neuroepidemiology*, 24, 123-8.
- NILANONT, Y., PHATTHARAYUTTAWAT, S., CHIEWIT, P., CHOTIKANUCHIT, S., LIMSRIWILAI, J., CHALERNPONG, L., YAMKAEW, N., LIRATHPONG, N., ANPRASERTPORN, P., KOMOLTRI, C., PRAYOONWIWAT, N. & POUNGVARIN, N. 2010. Establishment of the Thai version of National Institute of Health Stroke Scale (NIHSS) and a validation study. *J Med Assoc Thai*, 93 Suppl 1, S171-8.
- NOR AZLIN, M. N., SYED ALJUNID, S. J., NOOR AZAHZ, A., AMRIZAL, M. N. & SAPERI, S. 2012. Direct medical cost of stroke: findings from a tertiary hospital in Malaysia. *Med J Malaysia*, 67, 473-7.
- NORAZIANI K, NURUL AIN A, AZHIM M Z, SARA REKAB ESLAMI, BILAL DRAK, SHARIFAH EZAT W P & SITI NURUL AKMA A 2013. An Overview of Electronic Medical Record Implementation in Healthcare System: Lesson to Learn. *World App Sci J*, 25, 9.
- NORRVING, B. & KISSELA, B. 2013. The global burden of stroke and need for a continuum of care. *Neurology*, 80, S5-12.
- O'DONNELL, M., XAVIER, D., DIENER, C., SACCO, R., LISHENG, L., ZHANG, H., PIAS, P., TRUELSEN, T., CHIN, S. L., RANGARAJAN, S., DEVILLIERS, L., DAMASCENO, A., MONDO, C., LANAS, F., AVEZUM, A., DIAZ, R., VARIGOS, J., HANKEY, G., TEAL, P., KAPRAL, M., RYGLEWICZ, D., CZLONKOWSKA, A., SKOWRONSKA, M., LOPEZ-JARAMILLO, P., DANS,

- T., LANGHORNE, P., YUSUF, S. & INVESTIGATORS, I. N. T. E. R. S. T. R. O. K. E. 2010a. Rationale and design of INTERSTROKE: a global case-control study of risk factors for stroke. *Neuroepidemiology*, 35, 36-44.
- O'DONNELL, M. J., XAVIER, D., LIU, L., ZHANG, H., CHIN, S. L., RAO-MELACINI, P., RANGARAJAN, S., ISLAM, S., PAIS, P., MCQUEEN, M. J., MONDO, C., DAMASCENO, A., LOPEZ-JARAMILLO, P., HANKEY, G. J., DANS, A. L., YUSOFF, K., TRUELSEN, T., DIENER, H.-C., SACCO, R. L., RYGLEWICZ, D., CZLONKOWSKA, A., WEIMAR, C., WANG, X., YUSUF, S. & INVESTIGATORS, I. N. T. E. R. S. T. R. O. K. E. 2010b. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet*, 376, 112-23.
- O'MALLEY, K. J., COOK, K. F., PRICE, M. D., WILDES, K. R., HURDLE, J. F. & ASHTON, C. M. 2005. Measuring diagnoses: ICD code accuracy. *Health Serv Res*, 40, 1620-39.
- OECD/WORLD HEALTH ORGANIZATION. 2014. Health at a Glance: Asia/Pacific 2014: Measuring Progress towards Universal Health Coverage. Available: http://dx.doi.org/10.1787/health_glance_ap-2014-en.
- OGBU, U. C., WESTERT, G. P., SLOBBE, L. C. J., STRONKS, K. & ARAH, O. A. 2011. A multifaceted look at time of admission and its impact on case-fatality among a cohort of ischaemic stroke patients. *Journal of Neurology, Neurosurgery & Psychiatry*, 82, 8-13.
- OHIRA, T., SHAHAR, E., CHAMBLESS, L. E., ROSAMOND, W. D., MOSLEY, J. T. H. & FOLSOM, A. R. 2006. Risk factors for ischemic stroke subtypes: the Atherosclerosis Risk in Communities study. *Stroke*, 37, 2493-8.
- OKUMURA, K., OHYA, Y., MAEHARA, A., WAKUGAMI, K., ISEKI, K. & TAKISHITA, S. 2005. Effects of blood pressure levels on case fatality after acute stroke. *J Hypertens*, 23, 1217-23.
- OLSEN, T. S. & ANDERSEN, K. K. 2010. Female survival advantage relates to male inferiority rather than female superiority: A hypothesis based on the impact of age and stroke severity on 1-week to 1-year case fatality in 40,155 men and women. *Gen Med*, 7, 284-95.
- OLSEN, T. S., ANDERSEN, Z. J. & ANDERSEN, K. K. 2011. Age trajectories of stroke case fatality: leveling off at the highest ages. *Epidemiology*, 22, 432-6.
- OLSON, K. L., WOOD, M. D., DELATE, T., LASH, L. J., RASMUSSEN, J., DENHAM, A. M. & MERENICH, J. A. 2014. Positive predictive values of ICD-9 codes to identify patients with stroke or TIA. *Am J Manag Care*, 20, e27-34.
- OMRAN, A. R. 1971. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Mem Fund Q*, 49, 509-38.

- OMRAN, A. R. 1983. The epidemiologic transition theory. A preliminary update. *J Trop Pediatr*, 29, 305-16.
- OMRAN, A. R. 2005. The epidemiologic transition: a theory of the epidemiology of population change. 1971. *Milbank Q*, 83, 731-57.
- OVBIAEGELE, B. & NGUYEN-HUYNH, M. N. 2011. Stroke epidemiology: advancing our understanding of disease mechanism and therapy. *Neurotherapeutics*, 8, 319-29.
- PACI, M., NANNETTI, L., D'IPPOLITO, P. & LOMBARDI, B. 2011. Outcomes from ischemic stroke subtypes classified by the Oxfordshire Community Stroke Project: a systematic review. *Eur J Phys Rehabil Med*, 47, 19-23.
- PAGALAVAN, L. 2011. *For Future Doctors: Malaysian Healthcare System for the Dummies Part 3* [Online]. Available: <https://pagalavan.com/2011/08/10/for-future-doctors-malaysian-healthcare-system-for-the-dummies-part-3/> [Accessed 15 Dec 2016].
- PARK, T. H., SAPOSNIK, G., BAE, H. J., LEE, S. J., LEE, K. B., LEE, J., PARK, J. M., CHOI, J. C., KIM, D. E., CHO, Y. J., KIM, J. T., CHA, J. K., LEE, J., YU, K. H., LEE, B. C. & YOON, B. W. 2013. The iScore predicts functional outcome in Korean patients with ischemic stroke. *Stroke*, 44, 1440-2.
- PEDIGO, A., ALDRICH, T. & ODOI, A. 2011. Neighborhood disparities in stroke and myocardial infarction mortality: a GIS and spatial scan statistics approach. *BMC Public Health*, 11, doi: 10.1186/1471-2458-11-644.
- PETREA, R. E., BEISER, A. S., SESHADRI, S., KELLY-HAYES, M., KASE, C. S. & WOLF, P. A. 2009. Gender differences in stroke incidence and poststroke disability in the Framingham heart study. *Stroke*, 40, 1032-7.
- PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. & R DEVELOPMENT CORE TEAM. 2015. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-120.
- POUNGVARIN, N. 1998. Stroke in the developing world. *Lancet*, 352 Suppl 3, SIII19-SIII22.
- PROSPECTIVE STUDIES, C., LEWINGTON, S., WHITLOCK, G., CLARKE, R., SHERLIKER, P., EMBERSON, J., HALSEY, J., QIZILBASH, N., PETO, R. & COLLINS, R. 2007. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet*, 370, 1829-39.
- PROSPECTIVE STUDIES, C., WHITLOCK, G., LEWINGTON, S., SHERLIKER, P., CLARKE, R., EMBERSON, J., HALSEY, J., QIZILBASH, N., COLLINS, R. & PETO, R. 2009. Body-mass

index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet*, 373, 1083-96.

PULLICINO, P. M., MCCLURE, L. A., WADLEY, V. G., AHMED, A., HOWARD, V. J., HOWARD, G. & SAFFORD, M. M. 2009. Blood pressure and stroke in heart failure in the REasons for Geographic And Racial Differences in Stroke (REGARDS) study. *Stroke*, 40, 3706-10.

QUINN, T. J., LANGHORNE, P. & STOTT, D. J. 2011. Barthel index for stroke trials: development, properties, and application. *Stroke*, 42, 1146-51.

QURESHI, A. I., TUHRIM, S., BRODERICK, J. P., BATJER, H. H., HONDO, H. & HANLEY, D. F. 2001. Spontaneous intracerebral hemorrhage. *New England Journal of Medicine*, 344, 1450-60.

R DEVELOPMENT CORE TEAM 2013. R: A Language and Environment for Statistical Computing 3.1.1. Vienna, Austria: R Foundation for Statistical Computing.

RABE-HESKETH, S. & SKRONDAL, A. 2012. *Multilevel and Longitudinal Modeling Using Stata*, Texas, US: StataCorp.

RACHPUKDEE, S., HOWTEERAKUL, N., SUWANNAPONG, N. & TANG-AROONSIN, S. 2013. Quality of life of stroke survivors: a 3-month follow-up study. *J Stroke Cerebrovasc Dis*, 22, e70-8.

RAMEEZAN, B. A. & ZALIHA, O. 2005. Functional status of acute stroke patients in University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia. *Med J Malaysia*, 60, 548-59.

RANGANATHAN, P. & PRAMESH, C. S. 2012. Censoring in survival analysis: Potential for bias. *Perspect Clin Res*, 3, 40.

REEVES, M. J., BUSHNELL, C. D., HOWARD, G., GARGANO, J. W., DUNCAN, P. W., LYNCH, G., KHATIWODA, A. & LISABETH, L. 2008. Sex differences in stroke: epidemiology, clinical presentation, medical care, and outcomes. *Lancet Neurol*, 7, 915-26.

RENNA, R., PILATO, F., PROFICE, P., DELLA MARCA, G., BROCCOLINI, A., MOROSETTI, R., FRISULLO, G., ROSSI, E., DE STEFANO, V. & DI LAZZARO, V. 2014. Risk factor and etiology analysis of ischemic stroke in young adult patients. *J Stroke Cerebrovasc Dis*, 23, e221-7.

RESCHE-RIGON, M., AZOULAY, E. & CHEVRET, S. 2006. Evaluating mortality in intensive care units: contribution of competing risks analyses. *Crit Care*, 10, R5.

- ROBERTS, S. E. & GOLDACRE, M. J. 2003. Case fatality rates after admission to hospital with stroke: linked database study. *BMJ*, 326, 193-4.
- ROGER, V. L., GO, A. S., LLOYD-JONES, D. M., BENJAMIN, E. J., BERRY, J. D., BORDEN, W. B., BRAVATA, D. M., DAI, S., FORD, E. S., FOX, C. S., FULLERTON, H. J., GILLESPIE, C., HAILPERN, S. M., HEIT, J. A., HOWARD, V. J., KISSELA, B. M., KITTNER, S. J., LACKLAND, D. T., LICHTMAN, J. H., LISABETH, L. D., MAKUC, D. M., MARCUS, G. M., MARELLI, A., MATCHAR, D. B., MOY, C. S., MOZAFFARIAN, D., MUSSOLINO, M. E., NICHOL, G., PAYNTER, N. P., SOLIMAN, E. Z., SORLIE, P. D., SOTOODEHNIA, N., TURAN, T. N., VIRANI, S. S., WONG, N. D., WOO, D., TURNER, M. B., COMMITTEE, A. H. A. S. & SUBCOMMITTEE, S. S. 2012. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation*, 125, e2-e220.
- ROJAS, J. I., ZURRÚ, M. C., ROMANO, M., PATRUCCO, L. & CRISTIANO, E. 2007. Acute ischemic stroke and transient ischemic attack in the very old--risk factor profile and stroke subtype between patients older than 80 years and patients aged less than 80 years. *Eur J Neurol*, 14, 895-9.
- ROMERO, J. E. R., MORRIS, J. & PIKULA, A. 2008. Stroke prevention: modifying risk factors. *Ther Adv Cardiovasc Dis*, 2, 287-303.
- ROOS, Y. B., DE HAAN, R. J., BEENEN, L. F., GROEN, R. J., ALBRECHT, K. W. & VERMEULEN, M. 2000. Complications and outcome in patients with aneurysmal subarachnoid haemorrhage: a prospective hospital based cohort study in the Netherlands. *J Neurol Neurosurg Psychiatry*, 68, 337-41.
- ROQUES, F., NASHEF, S. A., MICHEL, P., GAUDUCHEAU, E., DE VICENTIIS, C., BAUDET, E., CORTINA, J., DAVID, M., FAICHNEY, A., GABRIELLE, F., GAMS, E., HARJULA, A., JONES, M. T., PINTOR, P. P., SALAMON, R. & THULIN, L. 1999. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *Eur J Cardiothorac Surg*, 15, 816-22; discussion 822-3.
- ROSAMOND, W., FLEGAL, K., FRIDAY, G., FURIE, K., GO, A., GREENLUND, K., HAASE, N., HO, M., HOWARD, V., KISSELA, B., KISSELA, B., KITTNER, S., LLOYD-JONES, D., MCDERMOTT, M., MEIGS, J., MOY, C., NICHOL, G., O'DONNELL, C. J., ROGER, V., RUMSFELD, J., SORLIE, P., STEINBERGER, J., THOM, T., WASSERTHIEL-SMOLLER, S., HONG, Y., COMMITTEE, A. H. A. S. & SUBCOMMITTEE, S. S. 2007. Heart disease and stroke statistics--2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*, 115, e69-171.
- ROSAMOND, W., FLEGAL, K., FURIE, K., GO, A., GREENLUND, K., HAASE, N., HAILPERN, S. M., HO, M., HOWARD, V., KISSELA, B., KISSELA, B., KITTNER, S., LLOYD-JONES, D., MCDERMOTT, M., MEIGS, J., MOY, C., NICHOL, G., O'DONNELL, C., ROGER, V., SORLIE, P., STEINBERGER, J., THOM, T., WILSON, M., HONG, Y., COMMITTEE, A. H. A. S. & SUBCOMMITTEE, S. S. 2008. Heart disease and stroke statistics--2008 update: a report

from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*, 117, e25-146.

ROSENBERG, P. S., KATKI, H., SWANSON, C. A., BROWN, L. M., WACHOLDER, S. & HOOVER, R. N. 2003. Quantifying epidemiologic risk factors using non-parametric regression: model selection remains the greatest challenge. *Stat Med*, 22, 3369-81.

ROYSTON, P. & ALTMAN, D. G. 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *App Stat*, 429-67.

ROYSTON, P. & ALTMAN, D. G. 2010. Visualizing and assessing discrimination in the logistic regression model. *Stat Med*, 29, 2508-20.

ROYSTON, P., ALTMAN, D. G. & SAUERBREI, W. 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*, 25, 127-41.

ROYSTON, P. & SAUERBREI, W. 2004. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Stat Med*, 23, 2509-25.

ROYSTON, P. & SAUERBREI, W. 2005. Building multivariable regression models with continuous covariates in clinical epidemiology--with an emphasis on fractional polynomials. *Methods Inf Med*, 44, 561-71.

ROYSTON, P. & SAUERBREI, W. 2013. Interaction of treatment with a continuous variable: simulation study of significance level for several methods of analysis. *Stat Med*, 32, 3788-803.

RUNCHEY, S. & MCGEE, S. 2010. Does this patient have a hemorrhagic stroke? Clinical findings distinguishing hemorrhagic stroke from ischemic stroke. *JAMA*, 303, 2280-6.

SACCO, R. L., BENJAMIN, E. J., BRODERICK, J. P., DYKEN, M., EASTON, J. D., FEINBERG, W. M., GOLDSTEIN, L. B., GORELICK, P. B., HOWARD, G., KITTNER, S. J., MANOLIO, T. A., WHISNANT, J. P. & WOLF, P. A. 1997. American Heart Association Prevention Conference. IV. Prevention and Rehabilitation of Stroke. Risk factors. *Stroke*, 28, 1507-17.

SACCO, R. L., ROBERTS, J. K. & JACOBS, B. S. 1998. Homocysteine as a risk factor for ischemic stroke: an epidemiological story in evolution. *Neuroepidemiology*, 17, 167-73.

SAFURAH, J., KAMALIAH, M. N., KHAIRIYAH, A. M. & NOUR HANAH, O. 2012. *Malaysia Health System Review*, Geneva, Switzerland: World Health Organization.

- SAHATHEVAN, R., MOHD ALI, K., ELLERY, F., MOHAMAD, N. F., HAMDAN, N., MOHD IBRAHIM, N., CHURILOV, L. & CUMMING, T. B. 2014. A Bahasa Malaysia version of the Montreal Cognitive Assessment: validation in stroke. *Int Psychogeriatr*, 26, 781-6.
- SANDERMAN, R., COYNE, J. C. & RANCHOR, A. V. 2006. Age: nuisance variable to be eliminated with statistical control or important concern? *Patient Educ Couns*, 61, 315-6.
- SANGHA, H., LIPSON, D., FOLEY, N., SALTER, K., BHOGAL, S., POHANI, G. & TEASELL, R. W. 2005. A comparison of the Barthel Index and the Functional Independence Measure as outcome measures in stroke rehabilitation: patterns of disability scale usage in clinical trials. *Int J Rehabil Res*, 28, 135-9.
- SAPOSNIK, G., HASSAN, K. A., SELCHEN, D., FANG, J., KAPRAL, M. K., SMITH, E. E. & INVESTIGATORS OF THE REGISTRY OF THE CANADIAN STROKE NETWORK FOR THE STROKE OUTCOME RESEARCH CANADA WORKING, G. 2011. Stroke unit care: does ischemic stroke subtype matter? *Int J Stroke*, 6, 244-50.
- SAPOSNIK, G., HILL, M. D., O'DONNELL, M., FANG, J., HACHINSKI, V., KAPRAL, M. K. & STROKE OUTCOME RESEARCH CANADA WORKING GROUP. 2008. Variables associated with 7-day, 30-day, and 1-year fatality after ischemic stroke. *Stroke*, 39, 2318-24.
- SARTI, C., RASTENYTE, D., CEPAITIS, Z. & TUOMILEHTO, J. 2000. International trends in mortality from stroke, 1968 to 1994. *Stroke*, 31, 1588-601.
- SARTI, C., STEGMAYR, B., TOLONEN, H., MAHONEN, M., TUOMILEHTO, J., ASPLUND, K. & PROJECT, W. M. 2003. Are changes in mortality from stroke caused by changes in stroke event rates or case fatality? Results from the WHO MONICA Project. *Stroke*, 34, 1833-40.
- SAUERBREI, W., ROYSTON, P. & BINDER, H. 2007. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*, 26, 5512-28.
- SCARBOROUGH, P., MORGAN, R. D., WEBSTER, P. & RAYNER, M. 2011. Differences in coronary heart disease, stroke and cancer mortality rates between England, Wales, Scotland and Northern Ireland: the role of diet and nutrition. *BMJ Open*, 1, e000263.
- SCHALLER, B. J. 2007. Influence of age on stroke and preconditioning-induced ischemic tolerance in the brain. *Exp Neurol*, 205, 9-19.
- SEALY-JEFFERSON, S., WING, J. J., SANCHEZ, B. N., BROWN, D. L., MEURER, W. J., SMITH, M. A., MORGENSTERN, L. B. & LISABETH, L. D. 2012. Age- and ethnic-specific sex differences in stroke risk. *Gen Med*, 9, 121-8.

- SEKI, M., HASE, K., TAKAHASHI, H. & LIU, M. 2014. Comparison of three instruments to assess changes of motor impairment in acute hemispheric stroke: the Stroke Impairment Assessment Set (SIAS), the National Institute of Health Stroke Scale (NIHSS) and the Canadian Neurological Scale (CNS). *Disabil Rehabil*, 36, 1549-54.
- SHAH, C. H., ISMAIL, I. M. & MOHSIN, S. S. 2008. Ambulance response time and emergency medical dispatcher program: a study in Kelantan, Malaysia. *Southeast Asian J Trop Med Public Health*, 39, 1150-4.
- SHAH, S., COOPER, B. & MAAS, F. 1992. The Barthel Index and ADL evaluation in stroke rehabilitation in Australia, Japan, the UK and the USA. *Aust Occup Ther J*, 39, 5-13.
- SHAH, S., VANCLAY, F. & COOPER, B. 1989. Predicting discharge status at commencement of stroke rehabilitation. *Stroke*, 20, 766-9.
- SHARMA, V. K., TSIVGOULIS, G., TEOH, H. L., ONG, B. K. & CHAN, B. P. 2012. Stroke risk factors and outcomes among various Asian ethnic groups in Singapore. *J Stroke Cerebrovasc Dis*, 21, 299-304.
- SHEIKH, K. & BULLOCK, C. M. 2007. Effect of measurement on sex difference in stroke mortality. *Stroke*, 38, 1085-7.
- SHIBER, J. R., FONTANE, E. & ADEWALE, A. 2010. Stroke registry: hemorrhagic vs ischemic strokes. *Am J Emerg Med*, 28, 331-3.
- SHIGEMATSU, K., NAKANO, H., WATANABE, Y., SEKIMOTO, T., SHIMIZU, K., NISHIZAWA, A., MAKINO, M., OKUMURA, A., BANDO, K. & KITAGAWA, Y. 2013. Characteristics, risk factors and mortality of stroke patients in Kyoto, Japan. *BMJ Open*, 3, e002181, doi:10.1136/bmjopen-2012-002181.
- SHIH, W. 2002. Problems in dealing with missing data and informative censoring in clinical trials. *Curr Control Trials Cardiovasc Med*, 3, 4.
- SIA, S. F., TAN, K. S. & WARAN, V. 2007. Primary intracerebral haemorrhage in Malaysia: in-hospital mortality and outcome in patients from a hospital based registry. *Med J Malaysia*, 62, 308-12.
- SIANNIS, F., COPAS, J. & LU, G. 2005. Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6, 77-91.
- SINGER, J. D. & WILLET, J. B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event*, New York: Oxford University Press.

- SINGH, R. B., SUH, I. L., SINGH, V. P., CHAITHIRAPHAN, S., LAOTHAVORN, P., SY, R. G., BABILONIA, N. A., RAHMAN, A. R., SHEIKH, S., TOMLINSON, B. & SARRAF-ZADIGAN, N. 2000. Hypertension and stroke in Asia: prevalence, control and strategies in developing countries for prevention. *J Hum Hypertens*, 14, 749-63.
- SOHRABJI, F., BAKE, S. & LEWIS, D. K. 2013. Age-related changes in brain support cells: Implications for stroke severity. *Neurochem Int*, 63, 291-301.
- SONG, Y. M., SUNG, J., DAVEY SMITH, G. & EBRAHIM, S. 2004. Body mass index and ischemic and hemorrhagic stroke: a prospective study in Korean men. *Stroke*, 35, 831-6.
- SRIDHARAN, S. E., UNNIKRISHNAN, J. P., SUKUMARAN, S., SYLAJA, P. N., NAYAK, S. D., SARMA, P. S. & RADHAKRISHNAN, K. 2009. Incidence, types, risk factors, and outcome of stroke in a developing country: the Trivandrum Stroke Registry. *Stroke*, 40, 1212-8.
- STANTON, M. C. & DIGGLE, P. J. 2013. Geostatistical analysis of binomial data: generalised linear or transformed Gaussian modelling? *Environmetrics*, 24, 158-71.
- STATACORP. 2010. Stata Statistical Software: Release 11. Texas, USA: StataCorp.
- STEGMAYR, B., VINOGRADOVA, T., MALYUTINA, S., PELTONEN, M., NIKITIN, Y. & ASPLUND, K. 2000. Widening gap of stroke between East and West eight-year trends in occurrence and risk factors in Russia and Sweden. *Stroke*, 31, 2-8.
- STORHAUG, H. M., NORVIK, J. V., TOFT, I., ERIKSEN, B. O., LOCHEN, M. L., ZYKOVA, S., SOLBU, M., WHITE, S., CHADBAN, S. & JENSSEN, T. 2013. Uric acid is a risk factor for ischemic stroke and all-cause mortality in the general population: a gender specific analysis from The Tromso Study. *BMC Cardiovasc Disord*, 13, 115, doi: 10.1186/1471-2261-13-115.
- STROKE UNIT TRIALISTS, C. 2007. Organised inpatient (stroke unit) care for stroke. *Cochrane Database Syst Rev*, CD000197.
- STRONG, K., MATHERS, C. & BONITA, R. 2007. Preventing stroke: saving lives around the world. *Lancet Neurol*, 6, 182-7.
- STURGEON, J. D., FOLSOM, A. R., LONGSTRETH, W. T., JR., SHAHAR, E., ROSAMOND, W. D. & CUSHMAN, M. 2007. Risk factors for intracerebral hemorrhage in a pooled prospective study. *Stroke*, 38, 2718-25.
- SUBHA, P. P., PILLAI GEETHAKUMARI, S. M., ATHIRA, M. & NUJUM, Z. T. 2015. Pattern and risk factors of stroke in the young among stroke patients admitted in medical college hospital, Thiruvananthapuram. *Ann Indian Acad Neurol*, 18, 20-3.

- SUDIRMAN, N., SHAHARUDDIN, S., LONG, C. M., HASHIM, R., ZULKIFLY, H. H., KASIM, S. S. & LIM, C. W. 2015. Relationship of time in therapeutic range (Ttr) with bleeding and stroke incidences of warfarinised patients with pulmonary embolism in a tertiary hospital in Malaysia. *Value Health*, 18, A378, doi:10.1016/j.jval.2015.09.794.
- SUDLOW, C. L. & WARLOW, C. P. 1996. Comparing stroke incidence worldwide: what makes studies comparable? *Stroke*, 27, 550-8.
- SUDLOW, C. L. & WARLOW, C. P. 1997. Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration. International Stroke Incidence Collaboration. *Stroke*, 28, 491-9.
- SUZUKI, K., IZUMI, M., SAKAMOTO, T. & HAYASHI, M. 2011. Blood pressure and total cholesterol level are critical risks especially for hemorrhagic stroke in Akita, Japan. *Cerebrovasc Dis*, 31, 100-6.
- SZKLO, M. & NIETO, F. J. 2007. *Epidemiology - Beyond the Basics*, Sudbury, MA 01776: Jones and Bartlett Publishers Inc.
- TEASDALE, G. & JENNETT, B. 1974. Assessment of coma and impaired consciousness. A practical scale. *Lancet*, 2, 81-4.
- TEGOS, T. J., KALODIKI, E., DASKALOPOULOU, S. S. & NICOLAIDES, A. N. 2000. Stroke: epidemiology, clinical picture, and risk factors--Part I of III. *Angiology*, 51, 793-808.
- TESTAI, F. D. & AIYAGARI, V. 2008. Acute hemorrhagic stroke pathophysiology and medical interventions: blood pressure control, management of anticoagulant-associated brain hemorrhage and general management principles. *Neurol Clin*, 26, 963-85, viii-ix.
- THAMMAROJ, J., SUBRAMANIAMATHI, V. & BHATTACHARYA, J. J. 2005. Stroke in Asia. *Neuroimaging clinics*, 15, 273-82.
- THE WORLD BANK. 2016. *Urban population (% of total) for Malaysia* [Online]. Available: <http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=MY>.
- THORVALDSEN, P., ASPLUND, K., KUULASMAA, K., RAJAKANGAS, A. M. & SCHROLL, M. 1995. Stroke incidence, case fatality, and mortality in the WHO MONICA project. World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease. *Stroke*, 26, 361-7.
- THORVALDSEN, P., KUULASMAA, K., RAJAKANGAS, A. M., RASTENYTE, D., SARTI, C. & WILHELMSSEN, L. 1997. Stroke trends in the WHO MONICA project. *Stroke*, 28, 500-6.

- TIBA EK, S. & DEHLENDORFF, C. 2011. Is Barthel Index a relevant measure for measuring prevalence of urinary incontinence in stroke patients? *Neurol Urodyn*, 31, 44-9.
- TORPY, J. M., BURKE, A. E. & GLASS, R. M. 2010. JAMA patient page. Hemorrhagic stroke. *JAMA*, 303, 2312.
- TOSTESON, A. N., WEINSTEIN, M. C., WILLIAMS, L. W. & GOLDMAN, L. 1990. Long-term impact of smoking cessation on the incidence of coronary heart disease. *Am J Public Health*, 80, 1481-6.
- TOWFIGHI, A., MARKOVIC, D. & OVBIAGELE, B. 2011. Persistent sex disparity in midlife stroke prevalence in the United States. *Cerebrovasc Dis*, 31, 322-8.
- TOWFIGHI, A., SAVER, J. L., ENGELHARDT, R. & OVBIAGELE, B. 2007. A midlife stroke surge among women in the United States. *Neurology*, 69, 1898-904.
- TREVOR, T. J. C. & FEDI, M. 2013. Acute management of stroke - II: haemorrhagic stroke. *Anaesthesia & Intensive Care Medicine*, 14, 366-70
- TRUELSEN, T., HEUSCHMANN, P. U., BONITA, R., ARJUNDAS, G., DALAL, P., DAMASCENO, A., NAGARAJA, D., OGUNNIYI, A., OVEISGHARAN, S., RADHAKRISHNAN, K., SKVORTSOYA, V. I. & STAKHOVSKAYA, V. 2007. Standard method for developing stroke registers in low-income and middle-income countries: experiences from a feasibility study of a stepwise approach to stroke surveillance (STEPS Stroke). *Lancet Neurol*, 6, 134-9.
- TRUELSEN, T., PIECHOWSKI-JÓ?WIAK, B., BONITA, R., MATHERS, C., BOGOUSSLAVSKY, J. & BOYSEN, G. 2006. Stroke incidence and prevalence in Europe: a review of available data. *Eur J Neurol*, 13, 581-98.
- TVEITEN, A., LJOSTAD, U., MYGLAND, A. & NAESS, H. 2014. Functioning of long-term survivors of first-ever intracerebral hemorrhage. *Acta Neurol Scand*, 129, 269-75.
- TZIOMALOS, K., ATHYROS, V. G., KARAGIANNIS, A. & MIKHAILIDIS, D. P. 2009. Dyslipidemia as a risk factor for ischemic stroke. *Curr Top Med Chem*, 9, 1291-7.
- UESHIMA, H., SEKIKAWA, A., MIURA, K., TURIN, T. C., TAKASHIMA, N., KITA, Y., WATANABE, M., KADOTA, A., OKUDA, N., KADOWAKI, T., NAKAMURA, Y. & OKAMURA, T. 2008. Cardiovascular disease and risk factors in Asia: a selected review. *Circulation*, 118, 2702-9.

- UYTTENBOOGAART, M., LUIJCKX, G. J., VROOMEN, P. C., STEWART, R. E. & DE KEYSER, J. 2007. Measuring disability in stroke: relationship between the modified Rankin scale and the Barthel index. *J Neurol*, 254, 1113-7.
- VAN ALMENKERK, S., SMALBRUGGE, M., DEPLA, M. F., EEFSTING, J. A. & HERTOOGH, C. M. 2013. What predicts a poor outcome in older stroke survivors? A systematic review of the literature. *Disabil Rehabil*, 35, 1774-82.
- VAN GIJN, J. & RINKEL, G. J. 2001. Subarachnoid haemorrhage: diagnosis, causes and management. *Brain*, 124, 249-278.
- VAN PEPPEN, R. P., KWAKKEL, G., WOOD-DAUPHINEE, S., HENDRIKS, H. J., VAN DER WEES, P. J. & DEKKER, J. 2004. The impact of physical therapy on functional outcomes after stroke: what's the evidence? *Clin Rehabil*, 18, 833-62.
- VANDERWEELE, T. J. 2009. On the distinction between interaction and effect modification. *Epidemiology*, 20, 863-71.
- VENKETASUBRAMANIAN, N. 1998. The epidemiology of stroke in Asean countries - a review. *Neurology Asia*, 3, 5.
- VENKETASUBRAMANIAN, N., CHANG, H. M., CHAN, B. P., YOUNG, S. H., KONG, K. H., TANG, K. F., ANG, Y. H., AHMAD, A., CHOW, K. Y. & SINGAPORE STROKE, R. 2015. Countrywide stroke incidence, subtypes, management and outcome in a multiethnic Asian population: the Singapore Stroke Registry--methodology. *Int J Stroke*, 10, 767-9.
- VENKETASUBRAMANIAN, N. & CHEN, C. L. H. 2008. Burden of stroke in Singapore. *Int J Stroke*, 3, 51-4.
- VERPOORTE, R. 2012. Primary data are the basis of all science! *Journal of Ethnopharmacology*, 139, 683-4.
- VOLPATO, S., ZULIANI, G., GURALNIK, J. M., PALMIERI, E. & FELLIN, R. 2001. The inverse association between age and cholesterol level among older patients: the role of poor health status. *Gerontology*, 47, 36-45.
- VON VOGELSANG, A. C., FORSBERG, C., SVENSSON, M. & WENGSTROM, Y. 2015. Patients experience high levels of anxiety 2 years following aneurysmal subarachnoid hemorrhage. *World Neurosurg*, 83, 1090-7.
- WAGSTAFF, A. J., OVERVAD, T. F., LIP, G. Y. & LANE, D. A. 2014. Is female sex a risk factor for stroke and thromboembolism in patients with atrial fibrillation? A systematic review and meta-analysis. *QJM*, 107, 955-67.

- WAHAB, K. W., OKUBADEJO, N. U., OJINI, F. I. & DANESI, M. A. 2008. Predictors of short-term intra-hospital case fatality following first-ever acute ischaemic stroke in Nigerians. *J Coll Physicians Surg Pak*, 18, 755-8.
- WAN-ARFAH, N., WAN MUHAMAD AMIR, W. A., MUZAIMI, M., MUSTAFA, M. & NAING, N. N. 2015. Determinants of mortality in first-ever stroke patients in the suburban Malaysia: A retrospective hospital-based study, 2005-2011. *Iran J Public Health*, 44, 1291-3.
- WARLOW, C. P. 1998. Epidemiology of stroke. *Lancet*, 352 Suppl 3, SIII1-4.
- WONG, E. S., WANG, B. C. M., GARRISON, L. P., ALFONSO-CRISTANCHO, R., FLUM, D. R., ARTERBURN, D. E. & SULLIVAN, S. D. 2011. Examining the BMI-mortality relationship using fractional polynomials. *BMC Med Res Methodol*, 11, 175, doi: 10.1186/1471-2288-11-175.
- WONG, K. S. 1999. Risk factors for early death in acute ischemic stroke and intracerebral hemorrhage: A prospective hospital-based study in Asia. Asian Acute Stroke Advisory Panel. *Stroke*, 30, 2326-30.
- WORLD HEALTH ORGANIZATION 2010. ICD-10 Training Tool User Guide. World Health Organization.
- WORLD HEALTH ORGANIZATION 2011a. Global Health and Aging. National Institute of Health (NIH).
- WORLD HEALTH ORGANIZATION 2011b. International Classification of Diseases (ICD).
- WORLD HEALTH ORGANIZATION. 2012. *International Classification of Diseases (ICD) Information Sheet* [Online]. Available: <http://www.who.int/classifications/icd/factsheet/en/> [Accessed 3 June 2012].
- WORLD HEALTH ORGANIZATION. 2014. STEPS Stroke Surveillance. Available: http://www.who.int/chp/steps/Section1_Introduction.pdf?ua=1.
- WORLD HEALTH ORGANIZATION. 2015. *International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)-2015-WHO Version for ;2015* [Online]. Available: <http://apps.who.int/classifications/icd10/browse/2015/en#/160-169> [Accessed 5 Sept 2012].
- XU, G., YIN, Q. & LIU, X. 2010. Narrowing the gap of stroke management between developed and developing countries: report of the fifth international stroke summit, Nanjing, China, July 17-19, 2009. *Cerebrovasc Dis*, 29, 101-2.

- YADAV, A. 2015. A comment on 'Low Socioeconomic Status Is an Independent Risk Factor for Ischemic Stroke: A Case-Control Study in North Indian Population'. *Neuroepidemiology*, 45, 70.
- YANG, B., ZHU, J., MIAO, Z., ZHOU, B., GE, W., ZHAO, H. & XU, X. 2015. Cystatin C is an independent risk factor and therapeutic target for acute ischemic stroke. *Neurotox Res*, 28, 1-7.
- YEN, S. T. & TAN, A. K. G. 2012. Who are eating and not eating fruits and vegetables in Malaysia? *Int J Public Health*, 57, 945-51.
- YOUSUF, R. M., FAUZI, A. R. M., JAMALLUDIN, A. R., HOW, S. H., AMRAN, M., SHAHRIN, T. C. A., MARZUKI, O. A. & SHAH, A. 2012. Predictors of in-hospital mortality in primary intracerebral haemorrhage in East coast of Peninsular Malaysia. *Neurology Asia*, 17, 93-99.
- YU, C. P., WHYNES, D. K. & SACH, T. H. 2008. Equity in health care financing: The case of Malaysia. *Int J Equity Health*, 7, 15.
- ZAN, A. 2016. *Why Break a Perfectly Decent Public Healthcare System?* [Online]. Available: <https://www.malaysiakini.com/columns/363145#ixzz4SoxJW7Bg> [Accessed 15 Dec 2016].
- ZELMANOVITZ, T., GROSS, J. L., OLIVEIRA, J. R., PAGGI, A., TATSCH, M. & AZEVEDO, M. J. 1997. The receiver operating characteristics curve in the evaluation of a random urine specimen as a screening test for diabetic nephropathy. *Diabetes Care*, 20, 516-9.
- ZHANG, X.-H., GUAN, T., MAO, J. & LIU, L. 2007. Disparity and its time trends in stroke mortality between urban and rural populations in China 1987 to 2001: changing patterns and their implications for public health policy. *Stroke*, 38, 3139-44.
- ZHOU, G., NIE, S., DAI, L., WANG, X. & FAN, W. 2013. Sex differences in stroke case fatality: a meta-analysis. *Acta Neurol Scand*, 128, 1-8.
- ZIA, E., PESSAH-RASMUSSEN, H., KHAN, F. A., NORRVING, B., JANZON, L., BERGLUND, G. & ENGSTROM, G. 2006. Risk factors for primary intracerebral hemorrhage: a population-based nested case-control study. *Cerebrovasc Dis*, 21, 18-25.