# Summary of Survival Analysis Assigment

Whitney (Huei-Chen) Chiu

## 1 Objective

The purpose of this article is to analyze patients in the North American Symptomatic Carotid Endarterectomy Trial (NASCET) with the aim of understanding the relationship between the main predictor of a history of cardiovascular disease (CVD) and the outcome of time to stroke after randomization and adjusting for confounders including antihypertensive medication, antilipidemic medication, anti-thrombotic medication, diabetes, severe stenosis, current smoking status, the type qualifying event, and sex.

## 2 Methods

### 2.1 Explanatory and Descriptive analysis

The dataset used in this study is a collection of 1449 patients who have a qualifying event of either a transient ischemic heart attack (TIA) or a stroke from the North American Symptomatic Carotid Endarterectomy Trial (NASCET). The data used for analysis was complete without any missing values present. No data cleaning was performed.

Data were entered and analyzed using R Studio. R Core Team (2018). (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.) All statistical tests were two-tailed with a statistical significance level of 5% adopted.

### 2.2 Modelling Approach

The main response of interest was time to stroke and the main predictor of interest was a history of cardiovascular disease. Sex, age, and qualifying event was tested as a potential effect modifier. Other six variables including: antihypertensive medication, antithrombotic medication, antilipidemic medication, diabetes, severe stenosis, current smoking status were assessed for confounding. A Cox proportional hazard model was used to estimate the association between the main predictor and the outcome. The nonparametric Kaplan-Meier estimate for all patients was investigated. The interaction term was tested using the likelihood ratio test. The $R^2$ produced in the regression results which is the squared multiple correlation coefficient explains the proportion of the variability in the response that is fitted by our model. A large $R^2$ would indicate a favorable predictive capacity of the model.

Collinearity between the independent variables in our fitted model was assessed, considering variance inflation factor (VIF) values larger than 10 evident. Collinearity would be dealt with if any variance inflation factors were larger than 10 since collinearity could cause false estimated beta coefficients.

### 2.3 Diagnostics and validation

Schoenfeld residuals were assessed after fitting our model into our data. We tested the null hypothesis which was the proportional hazard assumption being satisfied. In testing the proportional hazard assumption, we are interested in whether or not Beta i(t) is constant overtime (ie. slope is 0). Therefore, we plotted Beta(t) for all the variables by time to stroke in days see if our proportional hazard model was appropriate. Martingale residuals were assessed to examine the linearity of age. Influential observations were also accessed to see if there were outlying data points influencing our model. The ordinary bootstrap method with B=100 was used to assess overfitting in our model and produce a corrected index of discrimination. If there is overfitting present in our model, the Somer's Dxy index and G index would both decrease in the testing data compared to the training data.

## 3 Main results

A history of cardiovascular disease was significant in predicting the outcome of time to stroke, with a beta coefficient of 0.291 and a p-value of 0.016, accounting for all the other variables. Those who have a history of cardiovascular (CVD) diseases have a 34% increase in experiencing a shorter time to stroke (hazard ratio 1.34), with a confidence interval of 6% to 70%. Among all the variables included, six variables that appeared to be significant in predicting the outcome were the history of CVD, sex, qualifying event, severe stenosis, antihypertensive medication, and anti-thrombotic medication. Effect sizes were presented in Table 1.

The estimated time to stroke predicted by the history of CVD is showed in Figure 1.

## 4 Discussion

Given the finding of this report, those who have a history of cardiovascular disease have a 34% increased hazard in experiencing a shorter time to stroke and allocating treatment resources appropriately to the high-risk patients could possibly prolong their survival time.

Table 1: Effect size, 95% CI, p-values of all independent vairiables included in our model

| Predictors | Effect Sizes | 95% CI | p-value |
|---|---|---|---|
| Age | 1.26 | (1.06, 1.48) | 0.008 |
| Sex (Female/Male) | 1.34 | (0.57, 0.97) | 0.030 |
| Qualifying event (Stroke/TIA) | 1.39 | (1.10, 1.76) | 0.006 |
| CVD (Yes/No) | 1.34 | (1.05, 1.69) | 0.0156 |
| Diabetes (Yes/No) | 1.13 | (0.85, 1.48) | 0.399 |
| Current smoking (Yes/No) | 1.16 | (0.89, 1.52) | 0.262 |
| Severe stenosis (70-99/<70) | 1.48 | (1.15, 1.91) | 0.002 |
| Antihypertensive medication (No/Yes) | 0.75 | (0.58, 0.95) | 0.021 |
| Anti-lipidemic medication (Yes/No) | 0.64 | (0.43, 0.94) | 0.025 |
| Anti-thrombotic medication (No/Yes) | 0.69 | (0.26, 1.84) | 0.454 |

Figure 1: Kaplan–Meier survivor function with Greenwood 95% Confidence interval