CHAPTER V

ANALYSIS

After finishing the data collection process, all data was recorded in SPSS files and the raw data was rechecked by the 2 persons who processed the data entry. This chapter will deal with the data analysis.

Statistical Tests.

The univariate comparison of possible prognostic factors among the survival and dead groups was undertaken first. The chi-square test was used for the analysis of categorical outcomes, or the Fisher's exact test was used if the smallest expected value in the table was less than 5 or a nonparametric test was chosen if data was not normally distributed. A two-tailed *t*-test was used to analyze continuous data such as age and lesion size(cm^3). The exact P value of all variables were shown.

Stratified analysis was performed to compare the effects among various age groups.

Multiple logistic regression analysis were used (Lee, 1981) because the outcome is a dichotomous variable (survival or dead) and there were more than one predictor and the predictors were categorical and continuous data. The statistical program SPSS and STATA program were used for the analysis and for the performance of the ROC curve for the model. In multiple logistic regressions all variables are coded numerically. A discrete variable is coded 0/1 according to the presence or absence of that

variable. Continuous variables (for example age) are coded in units.

A logistic regression coefficient was calculated for each variables which, after allowance for differences in units of measurement reflect that variable's value in separating the groups. Comparison of the regression coefficient value of variables were made by comparing the size of their regression coefficients.

In the stepwise procedure employed by the SPSS program, each variable was added, starting with the most discriminating variable. The change in the separation of the prognostic groups and its statistical significance was measured with the addition of each variables.

The logistic regression model specifies that the probability of disease depended on a set of variables x_1, x_2,....., x_m in the following way:

$$pj = p(d=1|x)$$

$$= 1/\{1+\exp[-(\alpha+ \beta_1 x_{1j}+ \ldots+ \beta_i x_{ij} + \ldots \beta_m x_{mj})]\}$$

p = probability of the outcome in j-th subject ($j = 1,2,\ldots n$)

d = 0 denoted survival, and d = 1 denoted death

X_i = prognostic factor $X_{ij}(i = 1,2,\ldots m)$

α = logistic intercept

β_i = logistic regression coefficient

Hypothesis : $H_O : \beta = 0$, $H_A : \beta \neq 0$ at the conventional 5 percent level.

Calculate 1. Estimated odd ratio is defined as the odds of death for a subject with value of the $\boldsymbol{x}$ variables equal to $X_i = x_i + d_i$ relative to the odds of recovery with $X_i = x_i (i = 1,2,...m)$

2. Maximum likelihood estimation to estimate the actual magnitudes of the parameters or the probability of events under the logistic model.

3. Likelihood ratio test was used to assess if an apparent association ($\beta \neq 0$) is statistically significant. (χ^2, df = number of variables, P value)

4. Goodness of fit of the model, to assessing the adequacy of a logistic model by Hosmer Lemeshow χ^2.

5. Sensitivity, specificity and predictive accuracy of model was calculated with ROC curves.

SUMMARY OF BASELINE RISK FACTORS IN SURVIVAL AND NON-SURVIVAL (FIRST 24 HOURS AFTER ADMISSION)

Risk factor	Over all (%)	Survivor	Death	*p* value
Previous Stroke				
History of MI				
History of hypertension				
Diabetes mellitus				

TABLE 5.1

SUMMARY OF BASELINE CHARACTERISTICS OF SURVIVAL AND NON-SURVIVAL GROUPS(FIRST 24 HOURS AFTER ADMISSION)

	Survivors	Deaths	*p* value
1. Age in years (mean ± 2 S.E.)			
2. Sex			
3. Time from onset to treatment			
4. Glasgow Coma Scale at admission			
GCS 3-7			
GCS 8-11			
GCS 12-15			
5. Side of lesion (left)			
6. Lesion size (cm^3)(mean ± 2 S.E.)			
7. Site of lesion			
Basal ganglia			
Thalamus			
Lobar			
8. Intraventricular hemorrhage			
9. Fasting venous plasma glucose			
10. Systolic blood pressure			
11. Diastolic blood pressure			

TABLE 5.2

SUMMARY OF MAXIMUM LIKELIHOOD OF FIT OF LOGISTIC REGRESSION MODEL IN PATIENTS WITH SPONTANEOUS SUPRATENTORIAL INTRACEREBRAL HEMORRHAGE.

Variables	Logistic coefficient	Standard error	Wald χ^2	df	P-value
Variable 1					
Variable 2					
Variable 3					
Variable 4					
Variable 5					
Constant					

TABLE 5.3

SUMMARY OF SENSITIVITY, SPECIFICITY AND PREDICTIVE ACCURACY OF PROGNOSTIC FACTOR DETERMINED BY THE LOGISTIC REGRESSION MODEL

	Outcome		
Predictive outcome (for various cut-points)	Non-survival	Survival	Total
Non-survival	A	B	A+B
Survival	C	D	C+D
Total	A+C	B+D	**N**

TABLE 5.4