THE IDENTIFICATION OF PROGNOSTIC FACTORS IN PATIENTS SUFFERING FROM THROMBOEMBOLIC STROKE

THE IDENTIFICATION OF PROGNOSTIC FACTORS IN PATIENTS SUFFERING FROM THROMBOEMBOLIC STROKE

by

ERIC KWAME DUKU, B.Sc., Grad. Dip.

A Project

in Partial Fulfilment of the Requirements

For the Degree of

Master of Science

McMaster University

July, 1986

MASTER OF SCIENCE (1986) MCMASTER UNIVERSITY,

(STATISTICS) HAMILTON, ONTARIO.

TITLE : The identification of prognostic factors in patients suffering from thromboembolic stroke.

AUTHOR : Eric Kwame Duku,
B.Sc.(U.S.T)Ghana,
Grad. Dip.(Pop. Studies),Ghana.

SUPERVISOR : Prof. R. S. Roberts,
Dept. of Clinical Epidemiology and Biostatistics.

NUMBER OF PAGES : iii, 77

ACKNOWLEDGEMENT

A lot of assisstance and help have been given to me throughout the course of this project. First and foremost, I would like to thank my supervisor Prof. R. S. Roberts for his patience and knowledge in directing and encouraging me and the Department of Clinical Epidemiology and Biostatics for allowing me the use of the data.

To all my colleagues - N. Shrowthi, W. S. Lai and E. M. Ng - they are acknowledged for their encouragement and aid given me in many ways. I would also like to thank J. Kwajah and S. Hiscock for all their help during this project.

Finally, I would like to thank my Dad, Mr. Kofi Duku without whom all this would not have been possible.

ABSTRACT

In this project stroke data were analyzed with the use of survival techniques and incomplete principal component cox analysis. The data set resulted from a multi-centre randomised controlled trial coordinated by investigators from the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton with 438 patients. It was found that among stroke survivors, congestive heart failure along with other cardiac impairments pose the major risks. Other factors found to be important were patient age, previoius TIAs, presence of ulcers, diabetes and sex.

TABLE OF CONTENTS

CHAPTER ONE

1. INTRODUCTION

1.1 Project Overview

The primary objective of this project has been the application of biostatistical modelling techniques suitable for censored failure-time data in order to learn about the prognosis of patients who have suffered a stroke.

The data set which has been used, resulted from a multi-centre randomised controlled trial coordinated by investigators in the Department of Clinical Epidemiology and Biostatistics at McMaster University, Hamilton. The main purpose of the trial was to evaluate the efficacy of the platelet-active drug, Suloctidil, in reducing the risk of subsequent thrombo-embolic events (i.e. strokes, heart attacks) in patients who had already had at least one thrombo-embolic stroke. While the drug was found to be ineffective in this regard, the data set generated on the 438 patients followed for an average of 20 months offers an excellent opportunity to investigate the natural history of this disease.

It is well-known that patients who have had one thrombo-embolic event, in this case stroke, are at higher risk for subsequent events. While providing good quality estimates of this overall risk, the chief clinical interest in this data has been the identification of important "prognostic factors", that is, patient or disease characteristics which appear to influence subsequent risk. A wealth of potential factors were recorded in this study covering demographic characteristics, clinical

history, and the nature and severity of the presenting stroke. One of the most challenging practical difficulties has been in the data reduction area, the combining together of similar data items into composite variables, and the recognition of the small number of important features from the mass of data available.

From a statistical viewpoint, the project has involved the classical problem of censored failure-time data. Patients were recruited sequentially over a two-year period, each patient was followed for outcome events from the point of recruitment (i.e. date of randomization) to a common point six months after the last patient was randomized. The duration of follow-up therefore varied by subject and, even at the end of the study, relatively few had had a subsequent event. The techniques applied to this data centre on the estimation of the survival curve (Kaplan-Meier approach), the comparison of two or more survival curves (Mantel-Haenszel test), and most importantly, modelling the nature of the hazard in terms of prognostic factors (Cox's proportional hazard model). The statistical objectives of this work have thus been to gain experience with the theory and practical application of these techniques on a "real life" data set of some complexity.

1.2 Arteriosclerosis

The central biological mechanism at play in this work is an aging process called arteriosclerosis. Oxygenated blood is pumped by the heart to all parts of the human body via the arteries. As we get older, the walls of these vessels become thickened and less elastic, through the accumulation of fatty deposits. Although present to a certain extent in all people,this process seems to be enhanced by fatty diets, smoking, high blood pressure, and possibly, lack of exercise. Arterial thickening progresses to a

roughening of the interior surface of the artery (an arteriosclerotic plague) which in turn promotes the development of blood clots (thrombus), which are prone to break away. A dislodged blood clot, called an embolus, will be carried by the blood stream down smaller and smaller calibre vessels until they get stuck. Unless dissolved (i.e. lysed) quickly by naturally produced enzymes, the tissue downstream of this point may be starved of oxygenated blood, thus producing ischemia and subsequently tissue death - infarction.

Depending on the site of the arteriosclerosis, the embolus tends to wind up in the brain (cerebral infarction leading to stroke), in the heart muscle circulation (myocardial infarction or heart attack), or less commonly in the peripheral circulation which may lead to gangrene and limb amputation. Although not the focus of this analysis, the drug Suloctidil affects blood platelets making them less likely to initiate the clotting process at an arteriosclerotic lesion. This activity, demonstrable in laboratory tests did not provide any clinical benefit in stroke patients as judged from the results of this trial. Aspirin, another platelet-active drug, has however, been shown to be efficacious in patients at risk from myocardial infarction (MI) and in those having transient ischemic attacks (TIA) in the cerebral circulation.

1.3 Stroke

Strokes (or cerebrovascular accidents), are the third most common cause of death in North America and it has been estimated that 40% of stroke victims die within a month of having a stroke, while survivors remain at a very high risk of cerebral or myocardial infarction or death (N.I.N.C.D.S., 1979).

Several prospective studies of stroke have been recently reported, describing the

natural history of stroke (e.g. Gent et al., 1985; Sacco et al, 1982; Whisnant et al., 1973). These studies have improved the current knowledge of factors which influence the incidence of stroke and prognosis following a stroke. Thus it has appeared that prevention of the recurrence (i.e. secondary prevention) of cerebrovascular accidents or mortality from cerebrovascular accidents offers the greatest promise. The primary prevention of stroke, for example, efficacious treatment of arterial hypertension, also offers some hope of benefit (Levy, 1979).

1.4 The Suloctidil Trial

The majority of strokes, are known to be thrombo-embolic and are sometimes preceded by transient ischemic attacks (Sacco et al., 1982). However, the major problem, the prevention and treatment in high risk survivors, remains unanswered (Gent at al., 1985). Hence a multi-centre placebo-controlled randomised trial to assess the benefits of Suloctidil (200 mg t.i.d.) in reducing the incidence of further stroke was conducted at four centres - Hamilton, Montreal, Toronto and London - consisting of 438 patients who had suffered a thrombo-embolic stroke no less than 4 weeks or more than 4 months before the trial. Average follow-up was 20 months.

Primary analysis of the efficacy of Suloctidil was based on the incidence of the first event of stroke, myocardial infarction, or cardiovascular death, but excluding events that occurred more than 28 days after complete withdrawal from study medication for any particular reason. Thus primary analysis included 38 events in the Suloctidil group and 47 in the placebo group (see Appendix 1(b)). The absolute differences in outcomes were found to be similar but little evidence of the benefit of Suloctidil in preventing the recurrence of stroke was found (Gent et al., 1985).

1.5 Study Objectives

Since there was no real evidence of benefit of Suloctidil this study aims at

a. identifying the important prognostic factors of stroke survival; and

b. assessing the relative importance of these important factors.

These factors can then be compared among those surviving a stroke and as a result candidates for recurrence can be identified and possible preventive measures recommended. These objectives can be achieved, as said earlier, using the Kaplan-Meier estimation procedure, the Mantel-Haenszel test, and Cox's proportional hazards model. Another technique to be used is the Incomplete Principal Components Cox Regression which is discussed in detail in the next chapter.

1.6 Outline Of Study

The study is basically divided into five chapters. Chapter one, the introductory chapter, deals with the background and overview of the study, objectives of the study, and an outline of the study. In the second chapter, an attempt is made to discuss some of the basic concepts and techniques used in censored failure-time data analysis.

The data set along with the design of the Suloctidil trial, a brief review of relevant literature on the natural history of stroke prognosis are discussed in the third chapter. The fourth chapter discusses the analyses carried out and in the fifth and final chapter there is a summary of the findings and possible recommendations.

CHAPTER TWO

2. SURVIVAL DATA ANALYSIS

Survival data analysis comprises a variety of statistical techniques for analyzing positive-valued random variables. The value of the random variable being time to failure of a physical component (eg. in physics or engineering) or time to an event or death of a biological unit (eg. a patient in a medical study).

What distinguishes survival data analysis from other fields of statistics is censoring. Generally speaking, a censored observation contains only partial information about the random variable of interest. For example, in medical applications with clinical trials, where patients enter the study at differing times and each is treated with one of the several possible therapies, censoring occurs due mainly to sequential intake of patients. Thus if follow-up stops at a fixed point in time, censoring occurs early for patients who entered last. This chapter therefore discusses some of the basic concepts and techniques of survival data analysis which have some bearing on the present study.

2.1 Estimating The Survival Curve

The graphical representation of the total survival experience during the period of observation for a group of patients is called the survival curve. In the Kaplan and Meier (1958) approach to estimating the survival curve, ordered observations are used instead of grouped data used in the life-table method. This method has the

advantage of yielding results that are not dependent on the choice of the time intervals, and has been used with small samples where it is difficult to decide on an appropriate parametric distribution (Gross and Clark, 1975).

Kaplan and Meier(1958) introduced the product-limit estimate $\hat{S}(t)$ of the survival function, S(t) - the probability that an individual survives beyond time t. Their procedure assumes that the exact time of entry into the trial is known and that the exact time of death or loss to follow-up is also known. To define the product-limit estimate, let us assume we have observations on n individuals. Order the n observed survival times such that

$$t(1) < t(2) <...< t(n).$$

If a death and loss occur at the same time,Kaplan and Meier suggest treating the death as if it occurred slightly before the loss-to-followup.The survival curve is then estimated as

$$\hat{S}(t) = \prod_{i:t(i)<t} (n - i + 1 - d(i))/(n - i + 1)$$

where

$$d(i) = \begin{cases} 1 & \text{if t(i) represents a death.} \\ 0 & \text{if t(i) is a censored observation.} \end{cases}$$

Each term of $\hat{S}(t)$ in the product can be thought of as an estimate of the conditional probability of surviving past time t(i), given survival till just prior to t(i). Thus $\hat{S}(t)$ is a step function that changes at each distinct time of death.If no patient losses occur, $\hat{S}(t)$ reduces to the ordinary binomial estimate of the probability of survival at time t

(that is, s/n where s is the number of patients event-free after time t).

An approximate estimate of the variance of $\hat{S}(t)$ is given by

$$V\{\hat{S}(t)\} = \{\hat{S}(t)\}^2 \sum_{i:t(i)<t} d(i)/((n - i)(n - i + 1)).$$

This is also known as Greenwood's (1926) formula. Computer programs are available for this technique, BMDP1L (Dixon et al, 1983).

2.2 Comparison Of Two Survival Curves

For two different groups, such as the placebo (no treatment) group and the suloctidil (treatment) group in the Suloctidil trial (Gent et al,1985), the problem is to determine whether the two survival curves, say $S_P(t)$ and $S_T(t)$ are different based on their estimates $\hat{S_P}(t)$ and $\hat{S_T}(t)$. Two techniques for doing so are discussed next.

2.2.1 Point Comparison

For a specific point in time, say t, for which survival estimates have been computed using the Kaplan and Meier technique, one can compare the survival estimates $\hat{S_P}(t)$ and $\hat{S_T}(t)$ using the statistic

$$Z(t) = (\hat{S_P}(t) - \hat{S_T}(t)) / \sqrt{ (V[\hat{S_P}(t)] + V[\hat{S_T}(t)])}$$

where $V[\hat{S_P}(t)]$ and $V[\hat{S_T}(t)]$ are the Greenwoods' estimates of variance defined earlier.

Now Z(t) has approximately a normal distribution with mean 0 and variance 1 under the null hypothesis that $S_P(t) = S_T(t)$ (Friedman et al., 1982). It should be

noted, however, that there are problems in interpretation of the results since one has to decide what point in time is the most important. Therefore, point comparison is not recommended unless a few points can be justified prior to data analysis and are specified in the study protocol (Friedman et al., 1982). Thus five-year survival is often quoted instead and used in cancer chemotherapy trials.

2.2.2 Total Curve Comparison

Due to the limitations of single point in time comparisons, statistical methods to assess overall survival experience were put forward by Gehan (1965) and Mantel (1966) and are commonly used in survival analysis. The Gehan statistic is more powerful for survival distributions of the form

$$S(t;\theta) = \exp(t+\theta) / (1 + \exp(t+\theta)).$$

However, in practice the distribution of the survival experience of the study population is not known. Thus the Mantel-Haenszel technique which is more powerful for tests of survival distributions of the form

$$S_T(t) = \{S_P(t)\}^{\theta} \qquad \text{where } \theta = 1,$$

is more often used (Friedman et al, 1982).

The Mantel-Haenszel technique makes use of the procedure described by Cochran (1954) and Mantel and Haenszel (1959) for combining a series of 2 x 2 tables.In this procedure , each time, say t(j), a death occurs in either group (i.e placebo or suloctidil), a 2 x 2 table is formed as follows :

Groups	Deaths at time t(j)	Survivors at time t(j)	At risk prior to t(j)
Treatment	a(j)	b(j)	a(j)+b(j)
Placebo	c(j)	d(j)	c(j)+d(j)
TOTAL	a(j)+c(j)	b(j)+d(j)	N(j)

where a(j) and c(j) represent the observed number of deaths at time t(j) in the treatment and placebo groups respectively and N(j) is the total number of patients at risk prior to time t(j). One important condition is that at least one of a(j) or c(j) must be non-zero.

The expected number of deaths in the treatment group can then be shown to be given by

$$E\{a(j)\} = [\{a(j) + c(j)\}\{a(j) + b(j)\}] / N(j)$$

and the variance of the expected number of deaths in the treatment group is given by

$$V\{a(j)\} = [\{a(j)+c(j)\}\{b(j)+d(j)\}\{a(j)+b(j)\}\{c(j)+d(j)\}] / N(j)^2\{N(j)-1\}$$

which are the mean and variance of a hypergeometric distribution given the fixed marginal totals above. The Mantel-Haenszel (MH) statistic is then

$$MH = \{\sum_{j=1}^{k} a(j) - E\{a(j)\}\}^2 / \sum_{j=1}^{k} V\{a(j)\}$$

where k is the number of distinct death times in the combined group. This MH statistic has approximately a chi-square distribution with 1 degree of freedom and

enables one to compare the observed deaths with each treatment group with the deaths to be expected if the treatments were equally effective.

When there are more than two groups or samples, the most commonly used statistic is the log-rank test (Savage, 1956) which is similar to that given by Mantel and Haenszel. The log-rank test provides a simple means of testing equality of several survival curves while accommodating heterogeneity in the populations to be compared and involves stratification on all auxiliary variables describing such heterogeneity and the corresponding variances over strata. An approximate chi-square statistic can then be formed from these summary statistics as before . This provides a viable means of initial analysis and presentation for many data sets. Kalbfleisch and Prentice (1980) provide a more detailed discussion of the log-rank test.

2.3 Hazard functions.

The hazard function $\lambda(t)$ is another useful concept for describing survival and is also termed the failure rate, the instantaneous death rate or the force of mortality. It is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \Pr(t \leq T < t+\Delta t | T \geq t) / \Delta t$$

where T is a non-negative random variable representing the survival times of individuals in some population. It specifies the instantaneous death rate at time t, given that the individual survived up to time t. $\lambda(t)$ is the death density function and is related to the unconditional death density function, f(t) as follows :

$$\lambda(t) = f(t) / S(t)$$

where S(t) is the probability of an individual surviving till time t (Lawless,1982). By definition, f(t) = - S'(t),

$$S(t) = \exp\left[- \int_0^t \lambda(x)dx \right]$$

and

$$f(t) = \lambda(t) \times \left(\exp\left[- \int_0^t \lambda(x)dx\right]\right).$$

The hazard function is particularly useful with survival distributions since it describes the way in which the instantaneous probability of death for an individual varies with time. In applications, qualitative information about the hazard function helps in selecting a life distribution model. The shapes of hazard functions are qualitatively quite different and in medical applications, models with increasing hazard functions are used most. One reason being that interest often centers on a period of life of an individual over which some kind of gradual aging takes place, thus yielding an increasing hazard function (Lawless, 1982).

Lawless (1982), summarizing the usefulness of the hazard function said that, "the main point to be remembered is that the hazard function represents an aspect of a distribution that has a direct physical meaning and that information about the nature of the hazard function is helpful in selecting a model".

2.4 Survival Analysis With Covariates

In survival data analysis, many situations involve heterogeneous populations and it

is important to consider the relationship of time to death or an event to other factors which describe such heterogeneity. This can be achieved using regression models in which the dependence of time to death on concommitant variables is recognized.

Cox (1972), suggested an ingenious distribution-free approach to the analysis of data using a proportional hazards assumption. Proportional hazard models are models in which factors related to survival have a multiplicative effect on the hazard function and have the property that different individuals have hazard functions that are proportional to one another (i.e. the ratio of the hazard functions for two individuals with different sets of covariates does not depend on time).

The proportional hazards models is given by

$$\lambda(t;z) = \lambda_0(t) \, . \, \exp(\beta z)$$

where z is a known vector of covariates associated with an individual,

T is a continuous random variable representing time to death,

β is a vector of unknown regression coefficients, and

λ_0 is the baseline hazard function for the individual with covariate vector z = 0.

If $\lambda_0(t) = \lambda$, a constant, the model reduces to the exponential regression model and the Weibull is a special case when

$$\lambda_0(t) = \lambda p \, (\lambda t)^{p-1}$$ (Kalbfleisch and Prentice, 1980).

Cox's approach is distribution-free in that no specific form is assumed for $\lambda_0(t)$ and certain properties of the procedure do not depend on the underlying survival function or on $\lambda(t)$, thus making it flexible. However, if the data come from a specific hazards model such as the Weibull, there will be a loss in efficiency in using the distribution-free approach rather than the one based on the correct parametric

model.

To estimate the regression coefficients, β, let $t_1<t_2<...<t_k$ be k distinct times to death among n observed survival times. The conditional probability that an individual with covariate vector z_i dies at time t_i given that a single death occurs at t_i and given the set R_i - indices of the individuals at risk prior to t_i - is the ratio of the hazards

$$\exp(\beta z_i) / \sum_{j \epsilon R_i} \exp(\beta z_j)$$

Multiplying these probabilities together for all of the k death times gives the partial likelihood function (Cox, 1975)

$$L(\beta) = \prod_{i=1}^{k} \{\exp(\beta z_i) / \sum_{j \epsilon R_i} \exp(\beta z_j)\}$$

Cox(1975), suggests that for purposes of inferences about β, $L(\beta)$ can be treated as an ordinary likelihood function. In particular, maximization of $L(\beta)$ using numerical methods - Newton-Raphson - yields an estimate of β that is asymptotically normal with mean β and asymptotic covariance matrix estimated consistently by the inverse of the matrix of the second-order partial derivates of the log-likelihood function.

To test the hypothesis that all regression coefficients are identically zero, the Global Chi-square statistic which is used is defined as follows :

$$[U(0)]' I^{-1}(0) U(0)$$

where U(0) is the vector of first derivatives of the partial likelihood function evaluated at $\beta = 0$, and

I(0) is the negative of the matrix of second-order partial derivatives evaluated at $\beta = 0$.

The Global Chi-square statistic has asymptotic chi-square distribution with degrees of freedom equal to the number of covariates in the model (Dixon et al 1983). The proportionality assumption requires the ratio of the hazard functions of levels of an independent variable to be constant. To verify this assumption, one can plot a graph of log[-log[S(t;z)]] versus time where z is the mean of the covariates. Kalbfleisch and Prentice (1980), suggest stratifying the data based on an independent variable suspected of having a non-proportional effect on the hazard function. The plot should exhibit constant differences between strata if the proportionality assumption holds.

Finally it should be noted that sometimes a patient's prognosic status may change as a result of some event during the course of treatment. Such variables may be incorporated into the Cox model as time-dependent covariates. The idea of time-dependent covariates was introduced by Cox (1972) and work by Cox (1975), Kalbfleisch and MacKay (1978), and Kalbfleisch and Prentice (1980) have shown that the partial likelihood approach is still valid. Models with time-dependent covariates require specifying functions which define the values of each time-dependent covariate in terms of the survival time variable.

To carry out computations using the Cox proportional hazard model, a BMDP statistical package - BMDP2L - is available (Dixon et al, 1983), and a more detailed discussion of this model and other aspects of survival distributions can be found in Lawless (1982), Kalbfleisch and Prentice (1980), and Gross and Clarke (1975).

2.5 Incomplete Principal Components(IPC) Cox Regression

Like many biologically-oriented data sets, the Suloctidil trial offers a multitude of potential covariates. While stepwise application of the Cox model is possible, one must be concerned about over-fitting the model as a result of selecting variables which are spuriously related to survival. Possible solutions are the use of an independent validation data set obtained by random partitioning, the use of jack-knife procedures, or variable reduction techniques. The relative paucity of data precludes the first approach and computational complexity the second. We have thus opted to investigate the use of regression on principal to address this problem.

This is a method of placing restrictions on the parameters or covariates by reducing the effective number of parameters to estimate and can be applied to any regression model that is linear in covariates (Marquardt et al, 1975). It is performed as follows: for a set of p covariates X_1, X_2,..., X_p the principal components (pcs) are computed in order of amount of variation explained. A subset, say q, of the pcs explaining most of the variation in the covariates across patients are then taken and used as candidates in a variable selection program. The regression program - Cox's _ is forced to select the pcs in order of amount of variation explained until the pcs not in the model at a given step are jointly not significant at the 0.10 level of significance. The use of the residual chi-square as a stopping rule, allows one to select an adequate set of pcs for describing the response.

The imposition of the order of selection in accordance with the variability explained introduces stability and reduces noise in the model (Harrell, 1984). Also as noted by Marquardt et al (1975), "it is better to use a little bit of each variable than all of some variables and none of the remaining ones".

CHAPTER THREE

3. DATA SET

3.1 Design Of The Suloctidil Trial

As said earlier, the data was obtained from a multi-centre randomized controlled trial carried out to assess the potential of Suloctidil, a platelet reducing agent.

Patients of both sexes and all ages were considered eligible for the study if they had had a neurological deficit due to a well-documented thrombo-embolic stroke no less than 2 weeks or more than 4 months prior to entry into the trial. These included patients who had had atherothrombotic strokes, lacunar infarctions, and strokes that might have been due to emboli from the heart, provided the patient was not on an anti-coagulant therapy. Most important of all, the deficit had to be present at time of entry into the trial.

Eligible patients were allocated to one of two regimens - Suloctidil or placebo - according to randomization lists prepared separately for each of the four collaborating centres. Within each centre, patients were further stratified for sex and type of qualifying stroke.

To ensure eligibility criteria were met, initial assessment included a CAT scan and comprehensive review of all available documentation relating to the qualifying stroke, a neurological examination, a neurological history, an assessment of functional status, a general medical history, and potential cardiovascular risk factors. Also performed were haematologic and blood chemistry laboratory evaluations.

Patients were followed at three months and then three months thereafter for up to three years. At each follow-up visit, neurological and cardiovascular examinations were repeated; compliance and contamination were evaluated and a routine search made for possible toxicity and side-effects of treatment.

The set of outcomes on which the primary assessment of efficacy was based on was the recurrence of stroke or occurence of myocardial infarction or cardiovascular death. The outcome of stroke also included strokes due to haemorrhage, and the criteria for myocardial infarction included at least two of (i) typical chest pain, (ii) compatible ECG changes or myocardial scan and (iii) appropriate serum elevations. Death was categorized as stroke death and cardiac death - these were combined to form cardiovascular death - and death due to other causes. Death due pneumonia which had been clearly precipitated by a specific cardiac or cerebrovascular event were also classified as cardiovascular death.

3.2 Brief Review Of Relevant Literature

Although there is a good deal of information about factors related to the risk of developing complications of vascular disease, including stroke, information is meager regarding risk for subsequent stroke or death after recovery from a thrombo-embolic stroke. Several characteristics and attributes have been shown to indicate prognosis for survival and recovery after stroke. These include age (Goldner et al, 1967) , blood pressure (Eisenberg et al., 1964; Sacco et al., 1982), presence of cardiovascular diseases such as congestive heart failure (Robinson et al., 1959 ; Sacco et al., 1982) and coexisting diseases such as diabetes mellitus. Other important factors are the nature of stroke, ECG changes, and severity of stroke (Robinson et al., 1959).

It is also generally acknowledged that many of these charcteristics are related with one another. Because of this Goldner et al. (1967), conducted an epidemiological study to measure the relative effect of age, blood pressure and prior cerebrovascular accidents, as prognostic factors determining survival after stroke. They used 221 study patients - not confined to hospitalized cases - with an average follow-up of 15 months. Using the Chi-square test and case-fatality rates, they found no significant association between survival and sex, previous strokes nor hypertension (with age held constant). Age was the only factor found to be important.

Sacco et al. (1982), also assessed survival and recurrences after stroke in a general population of 5184 patients followed bienially for 26 years. They found that the risk of death or recurrence of stroke is profoundly and substantially influenced by sex, hypertension prior to the initial stroke and by cardiac co-morbidity - congestive heart failure and coronary heart disease.

Other studies have, however, found survival to be better for men as against what was found by Sacco et al.(eg Kabkin et al., 1978). This result was observed in a prospective study of a cohort in Manitoba. They also found high systolic blood pressure to be associated with decreased survival.

To establish the independent predictive effects of a range of personal and clinical characteristics, Shiekh et al. (1983), analysed data on 900 patients admitted to a hospital in the United Kingdom using multiple regression. 32 variables in all were analysed with three outcome variables - three-week survival,one-year survival and the level of disability on discharge. In the stepwise multiple linear regression, only 10 of the 32 variables were found to be related to prognosis for survival. These were increasing age, sensory deficit, abnormal pupils, speech defects, impaired conciousness, conjugate deviation of eyes, combined neurological deficits, extensive

motor deficits, history of unconciousness at onset of stroke, and severe disability. Sex made no independent contribution, nor did presence of visual field defects nor vascular disease presence.

Allen(1984), using a prognostic score derived from a prospective study of 148 consecutively admitted patients aged less than 76 years, who had survived the first 24 hours after an acute stroke, was used to compare the clinical features of the patients with their outcome after 2 months. Features found to predict functional dependence or death were older age, complete limb paralysis, depression of concious level, hemianopia and hemiplegia with higher cerebral dysfunction. The author notes that, there were imperfections in the scoring system and as a reminder, that, other factors not assessed in the study may affect patients' outcome after stroke.

Finally, Chen et al.(1985), studied the long-term prognosis of stroke by performing annual follow-up examinations on 306 patients who had survived cerebrovascular accidents. Patients were followed for one to four years and included 217 cases of cerebral thrombosis, 54 of cerebral haemorrhage, and 35 of TIA. The life-table method was used to determine cumulative survival rates and cumulative recurrence rates. They found that prognosis was not significantly influenced by, sex blood pressure on admission, or type of cerebrovascular accident. Age was an important prognostic factor, as usual.

This study makes use of the above-mentioned studies and others since continued attention to and modification of stroke and cardiovascular disease risk factors will help yield major dividends in the prevention of further strokes in patients who can be identified as being in the high risk group.

3.3 Available covariates

From the intial assessment, the available variables, after data reduction methods were used are presented in Table 1, and the four outcomes considered in this study are

a. stroke or myocardial infarction or cardivascular death;

b. stroke or myocardial infarction or death;

c. all deaths; and

d. cardiovascular deaths.

From Table 1, it should be noted that the names for the variables are only given in short, for convenience and that the full desription of each variable is given in Appendix 2. These shortened names will be used throughout the text. It should also be noted that these are not the original covariates, but that these are either transformed or combinations of some of the original covariates - which are more than 300 in number. The reduction of the original variables to the present number was one of the major problems encountered in this study. It is common knowledge that when there are many variables reflecting historical symptoms, physical signs and test results, relative to the number of cases, the identification of important prognostic factors is difficult. So, although the inclusion of a large number of factors in a model usually improves the prediction on which the model is developed,testing the fitted model on an independent sample often demonstrates a deterioration of its predictive ability (Harrel et al, 1984).

There was also the added problem of dealing with patients classified under the "not done" category for some of the test results, for example, the angiography

examination. Thus we were faced with the problem of deciding whether such patients were normal or did not have the exam done for other reasons. We have therefore assumed in general, that "no news is good news", so that a "no" response for the angiographic evidence of ulcer, for example combines subjects who had the test done and reported negative and those who did not have the exam done. This might underestimate the progostic significance of ulcer, but then it does seem preferable to omitting the "not done" cases entirely.

Next, although the factors used in this study were considered fixed, some of them,for example, right diastolic blood pressure, could change with time and could have been considered as time-dependent factors. In deciding to use only base-line characteristics we are thus avoiding the problem of time dependence and are thus looking at the prognosis in the eyes of a clinician assessing a patient shortly after a stroke. While subsequent changes in risk factors may occur and influence outcome, this seems to add to the complexity of the analysis unnecessarily. In addition, one is always concerned that subsequent changes are not independently causally important, but are related to earlier events. Their inclusion therefore, tends to "muddy" the causal interpretation, although other analysts might argue differently.

CHAPTER FOUR

4. ANALYSIS OF DATA

4.1 Data Reduction

As said earlier, with too many variables to contend with and problems of collinearity, it was decided to reduce the data set in terms of the number of variables of interest in order that the models used make any clinical or statistical sense. To aid us, two techniques were used

a. 2x2 tables - using as outcome, survival up to one year, and

b. principal component analysis (pca).

The former approach involved using 364 patients who had either survived a year or more or had an event before one year; all others who were followed for less than a year and did not have an event were considered censored.

Using the 2x2 tables, which are frequently employed in presenting statistical evidence, an attempt was made to see if survival varied across the levels or categories of the variables. Next similar or related variables were combined if survival patterns were similar. Otherwise, variables were treated as independent of other variables.

As an example, consider the variable, ARRYTHMA (the presence of arrythmia), which is a combination of two variables - presence of arrythmia (from cardiac exam) and presence of arrythmia (from ECG). The 2X2 tables corresponding to these variables are presented in Table 2 (a), (b), and (c) respectively. It can be seen from Table 2 that survival across the levels of the three variables are similar. Slight

differences, though exist in the proportions. This can be attributed to the fact that due to the two methods used to assess the presence of arrythmia, there would be slightly different numbers for each method. However, the differences are only slight and as such the proprtions across each of the three variables are similar.

Thus using this technique, the number of available variables was reduced from over 300 to 52, after which no further reduction could be achieved. The 52 variables (or covariates) have already been presented in Table 1. It should be noted that some of the variables which had too many patients in the "not done" category were eliminated. Others were combined because the type of neurological deficit did not really matter. For example, MUSCLE, which was used to represent muscle weakness in any part of the patient's body and SENSORY which represented sensory weakness in any part of the patient's body. Thus some of the variables represent general neurological deficits.

The main advantage, therefore, in the use of the 2X2 tables, was in helping decide if the variation in the survival across each variable made any clinical sense. It also helped identify the most influential factors of survival after a stroke.

4.2 Identification of Important Prognostic Variables

4.2.1 Testing the assumptions of the Cox proportional hazard model

The first goal in using Cox's proportional hazards model is to check the model assumptions, which is only possible if the factors are dichotomous in nature, or have been defined as such. From Chapter Two, it can be recalled that the proportionality assumption requires the hazard functions for two individuals with different sets of

covariates or factors does not depend on time. This assumption can thus be checked by plotting $\ln[-\ln\hat{S}(t;z)]$ against time, and if the proportionality assumption holds, the plot should exhibit constant differences between the levels of the factor (Kalbfleisch and Prentice, 1980). This was done univariately.

Since it will not be possible to present the plots for all the 50 or so factors, four are presented in Figures 1(a), (b), (c), and (d) for the variables CHF, PTNAGE, STROKES and PREVTIAS respectively using as outcome Stroke, MI or Death as examples. As will be seen from these four plots, across each factor or variable we can see the differences are approximately constant.

The main advantage in going through the check is that if any particular variable does not exhibit constant differences across its levels, one can then stratify the data using that factor so that cases within each strata then conform to the proportional hazards model. The model is then defined as:

$$\lambda_j(t:z) = \lambda_{0j}(t)\exp(\beta z)$$

where j represents one of the s different strata corresponding to the factor. Under this model, the arbitrary hazard function, $\lambda_{0j}(t)$ is allowed to be different for each stratum while the regression coefficients remain the same across strata.

4.2.2 Important Prognostic Factors

After ascertaining that the factors do indeed satistfy the proportionality assumption of the Cox model, the next goal was then to quantify the relationship between the variables and survival for all four outcomes:

a. Stroke or MI or CV death

b. Stroke or MI or Death

c. Death

d. CV Death

A set of regression coefficients which relates the effect of each variable to survival was computed for each outcome. These coefficients indicate the relationship between the variables and survival, in particular, a positive coefficient indicates a negative relationship with survival and as such increases the value of the hazard function. Tables 4(a), (b), (c), and (d) show the relationship between each factor and survival for each of the four outcomes respectively.

From the p-values of the tests of the hypotheses that the regression coefficients are identically equal to zero, it is obvious that the values of 0.0139, 0.021, 0.000, 0.004 indicate that all the regression coefficients are not jointly identical to zero. Thus it is necessary to do a stepwise regression to remove the variables which are least significantly related to the outcomes.

It is also evident from Table 4, on the whole, that the factors PTNAGE, MI, PREVTIAS, SMOKEST, ANGINA, ARRYTHMA, CHF, STRKTYP2, ONSET, INFTSIZE, ULCER, STENOSIS, DYSPHAT, MUSCLE, SENSORY, MENTATN, VISION, and DPREFX are consistently negatively related to survival and as such increase the hazard for all four outcomes. On the other hand, the following factors were also found to be consistently positively related to survival after a stroke : SEX, RCAROTID, LCAROTID, OCCLSION, LIMB, STRROKES, CARDSURG, VETEBR, EYES, VERTIGO, SENSYS, TYPESMKR, and DYSPHST.

Consistent with earlier findings, the absence of any previous strokes, STROKES, is positively related to survival and thus causes a reduction in the hazard function, which is definitely a good sign. Also consistent with earlier studies by Wolf (1985), Sacco et al (1982), and others, PTNAGE, ANGINA, MI, and CHF are negatively related with survival and thus lead to an increase in the hazard function. Since SEX and STROKE are both positively related with survival, this means one should expect that females with no previous strokes, except the qualifying one, stand a better chance of surviving a stroke. Finally similar to findings by Sheikh et al (1983), SENSORY and VISION are also negatively related to survival.

One inconsistency with earlier findings, however, is that although HYPERT was negatively related to survival for three outcomes - Stroke or MI or CV death, CV Death, and Death - it was positively related to survival for the outcome Stroke or MI or Death. No explanation could be found for this inconsistency, although it is known that hypertension is the major risk factor for the occurrence of the first stroke (Sacco, 1982). Thus apart from this, the data set and results conform somewhat to what earlier studies have found on stroke survival.

As said earlier, since the initial investigation showed that we could not assume all the coefficients were jointly equivalent to zero, the stepwise selection procedure using Cox's model was used to select 'significant' factors. The restriction placed on the selection procedure was that a variable was selected if it had a p-value of 0.10 or less and was jointly significant with the variables already selected. The variables so selected are presentd in Table 5 (a), (b), (c), and (d) for all four outcomes.

From Table 5, the most consistent significant predictor of survival is ONSET - for all outcomes. Apart from that, the relative risk of an event for a subject whose stroke onset was fluctuating or abrupt is at least twice that for a subject whose stroke

onset was gradual or stepwise. Evident also from Figure 2 is the difference in survival for patients in the two groups (p=0.0274). It should be noted that the outcome used for the survival curves was Stroke or MI or Death and that the curves for other outcomes were similar.

CHF was also found to be a consistent significant predictor of survival for all the outcomes. This is in agreement with what other epidemiological studies have identified. For example, Wolf (1985) identified people with cardiac disease impairments including coronary heart disease (CHD), congestive heart failure (CHF), left ventricular hypertrophy (LVH), and fibrillation at an increased risk of strokes and transient ischemic attacks, although this is second to hypertension. In the Framingham study, the leading cause of death was found to be cardiovascular diseases and survival was found to be worse in patients with CHD or CHF (Sacco et al., 1962).

It is also evident from Table 5 that CHF - history of congestive heart failure, or left ventricular hypertrophy or cardiomegaly - also increases one's risk of having an event. The increase in the relative risk is by at least 60%and as can be seen in Figure 3, patients without CHF definitely do survive better than those with CHF (p=0.0004)

Although the absence of any previous strokes prior to the qualifying stroke, STROKES, improves one's chances of surviving after a stroke, it is not however, a significant predictor of for the outcomes CV Death nor Death. One might say then that it is a major risk factor for the recurrence of stroke and myocardial infarction. From Table 5(a) and (b) we can see that the absence of any previous strokes does indeed reduce one's risk of having an event (see Figure 4). The relative risk reduction of having an event is about 45% for patients with no previous strokes as against patients who have had one or more stroke prior to the qualifying stroke.

According to Sacco et al (1982), recurrence of stroke or myocardial infarction is common following an atherothrombolic brain infarction and is usually the same type as the initial stroke. They also found that recurrence was strongly influenced by cardiac comorbidity (CHD, CHF, LVH) prior to the stroke and by the presence of hypertension prior to the stroke.

However, in this study, HYPERT, that is, the history of hypertension or abnormal right diastolic blood pressure, was a significant risk factor for only CV Deaths. This contradicts what earlier studies found somehow, and can be attributed to differences in the definition of hypertension and also differences in the time frames used since some authors used abnormal blood pressure or hypertension prior to the stroke and not at time of entry into the study. Thus although hypertension prior to the stroke is a major risk factor for stroke, it is however only significant as a risk factor for cardiovascular death. From Table 5(c), we also realise it is indeed one of the major significant risk factors for cardiovascular deaths after a stroke with a relative risk of having an event being about three times that if one was not hypertensive.

Another significant risk factor was ULCER, the presence of ulcer(from the angiography exam), for the outcomes Stroke or MI or CV Death and Stroke or MI or Death. It must be noted, however, that those who did not have an angiography done were assumed to be normal and as such were considered not to have an ulcer. It should also be noted that out of the 438 patients entered in the study, only 7 had an ulcer present, and 4 out of the 7 had en event, as is portrayed in Figure 7. Thus the effect of ULCER as a risk factor might not be as precise as one would have liked and it is obvious from the estimates of relative risk in Table 5(a) and (b) that the risk of a patient with ulcer having an event is at least four times that for a "normal" patient.

Next, PTNAGE, which is generally considered an important predictor of one's survival was however not a significant risk factor for the outcome Stroke or MI or CV Death, although it was a significant risk factor for the three other outcomes. No reason could be found for this apparent discrepancy. However, since it is known that a patient's survival in general depends on his age, we can at least be sure that it is an important risk factor for the stroke-related death outcomes. Chen and Ling (1985), found a similar result using as outcome deaths from cerebrovascular accidents or immediate complications, that is, stroke-related deaths. Khaw et al (1984), also found similar results using as outcome , stroke-related deaths.

Similar to findings from other studies, cigarette smoking - SMOKEST and TYPESMKR - was not found to be a significant risk factor for any outcome, although one study in Finland did find that the number of cigarettes smoked was significantly associated with stroke risk using multiple logistic analysis (Johnson et al., 1967; Ostfield, 1980). In young men in the Framingham study, univariate analysis showed a positive association between stroke and cigarette smoking, but a subsequent 24-year follow-up found no significant association with stroke (Wolf et al., 1978). Thus although cigarette smoking is a risk factor of stroke, it does not look like it is a risk factor for the recurrence of stroke nor myocardial infarction nor death.

Thus the major risk factors for the recurrence of stroke or myocardial infarction or death from Table 5 are attributes and factors associated with patient age (PTNAGE), onset of stroke (ONSET), cardiac comorbidity (CHF, PREVTIAS). Other important risk factors are the absence of any previous strokes (STROKES), history of diabetes (DIABETES), SEX, and the presence of ulcer (ULCER).

It should be noted that whereas STROKES bring about a reduction in the relative risk of having an event, the other factors pose major risk problems . These factors

thus provide the potential for intervention and prevention of the recurrence of stroke or myocardial infarction or death after a stroke. SEX, also brings about a change in the relative risk of having an event.

4.3 Incomplete Principal Components (IPC) Cox Regression

As said earlier, the principal components (pcs) were computed for the 52 covariates and 20 were selected in order of amount of variation explained. These were used in a variable selection program using Cox's model and the regression program was forced to select the pcs in order until the pcs not in the model were jointly not significant at the 0.10 level of significance using the residual Chi-square statistic.

The imposition of the selection of the pcs in order of variation explained was to ensure stability and reduce noise in the model. The use of the residual Chi-square statistic as a stopping rule, allows one to select an adequate set of pcs for describing the response while avoiding the problem that if the second pc is not significant while the third is, one would probably select the first three pcs.

Table 6 presents the the results from fitting all 20 pcs into the Cox model for all four outcomes. It is thus obvious that the pcs can exlplain a significant amount of variation. By comparing the Global Chi-square in Table 6 to those in Table 4, we see that 20 pcs do a reasonable job in explaining survival sompared to the 50 odd original variables. Thus the 20 pcs are enough to adequately describe the outcomes and with this in mind the stepwise selection procedure was performed.

The use of the pcs are supposed to help build better predictive models. The results of the stepwise procedure are presented in Table 7 along with residual Chi-

square at every step. From this we realize that more pcs than covariates were selected by the model for each of the four outcomes. This can be attributed to the fact that the individual variables were largely uncorrelated with one another and as such the proportions of the variation explained by each pc is nearly constant and so the Principal Component Analysis merely found components which are close to the original variables, but arranged in order of decreasing variance. Some analysts argue that when this is the case, it is better not to use or perform Principal Component Analysis.

Apart from this, there are also certain drawbacks to the use of the pcs in the variable selection program. First, it is difficult to interpret the models involving the pcs beyond the first pc and secondly, it is not possible to delineate the factors from the pcs with sufficient clarity for a thorough understanding and clinical insight.

Some solutions to these drawbacks which have been suggested is to use ridge regression or variable clustering techniques. Ridge regression suffers from being arbitrary and difficut to apply outside the field of ordinary linear regression (Harrell et al., 1984). Both were not attempted due to constraints.

In Table 8, and from Table 7, it thus becomes evident that the models with the individual variables are as good as those with the pcs although we have a better understanding of the model and also clearer clinical insight when the individual variables are in the model. It might thus be suggested that pcs should only be used when there is strong evidence of correlation between some of the variables, or when one intends on building a better predictive model. Alternatively, one can involve clinicians in the study to help derive indices for the clusters of variables and then proceed with the modelling.

CHAPTER FIVE

5. SUMMARY AND CONCLUSIONS

Stroke is the third major cause of death in North America, however, over the last decade, there has been a reported decline of approximately five per-cent per year in stroke mortality. This has been attributed to the identification of hypertension as the major risk factor for stroke - whether the stroke mechanism is a haemorrhage or infarction - and the demonstration that treatment will reduce stroke recurrence and stroke death (Kagan et al., 1980; Kannel et al., 1970).

However, this study found that hypertension in the sample used was a major significant risk factor for only cardiovascular deaths The most consistent significant risk factors identified for the recurrence of stroke or myocardial infarction or death were congestive heart failure (i.e. history of congestive heart failure or left ventricular hypertrophy or cardimegaly), and the onset of the qualifying stroke (i.e whether abrupt or fluctuating). This means that patients with cardiac diseases whose qualifying stroke onset was abrupt or fluctuating stand a higher risk of having an event - stroke or myocardial infarction or cardiovascular death or even death.

A third and very important factor found to be a significant risk factor was the absence of any previous strokes prior to the qualifying stroke since it reduces one's risk of any further events by about a half. This makes clinical sense and from the point of view of a clinical trial can be considered the most important.

Other factors found to be significant predictors of stroke survival were patient age, previous TIAs, presence of ulcers, diabetes, sex. Significant risk factors for

stroke-related deaths were, including the factors already mentioned, presence of angina, throat defects, valvular diseases, functional status of the patient and carotid endarectomy.

The use of the Incomplete Principal Components Cox regression did not help much due to the fact that the individual variables were highly uncorrelated among themselves. Its major drawback is the difficulty in interpreting the principal components beyond the first pc. An alternative to this procedure is the variable clustering techniques which would involve the use of cardiologists and other clinicians to derive indices for each cluster as suggested by Harrell et al., (1984).

Concluding then, one can say that among stroke survivors, congestive heart failure along with other cardiac impairments pose the major risks. Therefore early diagnosis and treatment of these impairments for stroke survivors may offer the opportunity to prevent any recurrence of stroke or myocardial infarction or death.

TABLES AND APPENDICES

TABLE 1. VARIABLES USED IN THE STUDY

VARIABLE	CODE	LEVEL
SEX	1	male
	2	female
PTNAGE		continuous
STROKES	0	one/more
	1	none
CERESURG	0	no
	1	yes
PREVTIAS	0	no
	1	yes
SMOKEST	0	no
	1	yes
TYPESMKR	0	none/light
	1	heavy
HYPERT	0	no
	1	yes
ANGINA	0	no
	1	yes
MI	0	no
	1	yes
ARRYTHMA	0	no
	1	yes

TABLE 1 (cont.).

VARIABLE	CODE	LEVEL
CHF	0	no
	1	yes
VALVDIS	0	no
	1	yes
CARDSURG	0	no
	1	yes
DIABETES	0	no
	1	yes
STRKTYP1	0	no
	1	yes
STRKTYP2	0	no
	1	yes
STRKTYP3	0	no
	1	yes
LCAROTID	0	no
	1	yes
RCAROTID	0	no
	1	yes
VETEBR	0	no
	1	yes

TABLE 1 (cont.).

VARIABLE	CODE	LEVEL
CINFRCT	0	no
	1	yes
INFTSIZE		continuous
INFCTAPP	0	no
	1	yes
ULCER	0	no
	1	yes
STENOSIS	0	no
	1	yes
OCCLSION	0	no
	1	yes
MENTCH	0	no
	1	yes
DYSPHAT	0	no
	1	yes
ATAXIAT	0	no
	1	yes
EYES	0	no
	1	yes
VERTIGO	0	no
	1	yes

TABLE 1 (cont.).

VARIABLE	CODE	LEVEL
THROAT	0	no
	1	yes
MUSCLE	0	no
	1	yes
SENSORY	0	no
	1	yes
FNSTAT		continuous
FNSTAT1		continuous
MENTATN	0	no
	1	yes
DYSPHST	0	no
	1	yes
VISION	0	no
	1	yes
FACE	0	no
	1	yes
LIMB	0	no
	1	yes
MUSCTONE	0	no
	1	yes

TABLE 1 (cont.).

VARIABLE	CODE	LEVEL
CEREBR	0	no
	1	yes
DPREFX	0	no
	1	yes
PLREFX	0	no
	1	yes
SENSYS	0	no
	1	yes
GAIT	0	no
	1	yes
CDAPEX		continuous
DYSART	0	no
	1	yes
ONSET	0	no
	1	yes
MUSCPOWR	0	no
	1	yes

TABLE 2. One-year Survival Status By Arrythmia

(a) Arrythmia (from cardiac exam)

STATUS	ABSENT	PRESENT
Yes	82.5	77.1
No	17.5	22.9
TOTAL	100.0	100.0

(b) Arrythmia (from ECG)

STATUS	ABSENT	PRESENT
Yes	82.4	78.6
No	17.6	21.4
TOTAL	100.0	100.0

(c) Arrythmia (from cardiac exam or ECG)

STATUS	ABSENT	PRESENT
Yes	82.1	80.4
No	17.9	19.6
TOTAL	100.0	100.0

TABLE 3. Principal Components With Characteristic Roots And Cumulative Amount Of Variation Explained

PC	CHARACTERISTIC ROOT OF PC	CUM. AMOUNT OF VARIATION EXPLAINED
1	4.38	0.09
2	3.14	0.15
3	2.62	0.20
4	2.38	0.25
5	2.23	0.29
6	1.96	0.33
7	1.68	0.37
8	1.60	0.40
9	1.54	0.43
10	1.47	0.46
11	1.38	0.49
12	1.31	0.51
13	1.27	0.54
14	1.25	0.56
15	1.15	0.59
16	1.09	0.61
17	1.06	0.63
18	1.03	0.65
19	1.02	0.67
20	0.97	0.69

TABLE 4. Relationships between variables and survival for all outcomes

(a) STROKE or MI or CV DEATH

(i) negative relationship with survival

PTNAGE	CDAPEX	DYSPHAT
PREVTIAS	STRKTYP2	MUSCLE
SMOKEST	STRKTYP3	SENSORY
HYPERT	ONSET	FNSTAT
ANGINA	INFCTAPP	MENTATN
MI	ULCER	DYSART
ARRYTHMA	STTENOSIS	VISION
CHF	MENTCH	FACE
DIABETES	ATAXIAT	MUSCTONE
PLREFX	CEREBR	DPREFX
MUSCPOWR	INFTSIZE	

(ii) positive relationship with survival

SEX	STROKES	TYPESMKR
VALVDIS	CARDSURG	RCAROTID
LCAROTID	VETEBR	CINFRCT
OCCLSION	EYES	VERTIGO
THROAT	FNSTAT1	DYSPHST
LIMB	SENSYS	GAIT

GLOBAL CHI-SQUARE = 74.52 D.F.= 50 P-VALUE=0.0139

TABLE 4 (cont.)

(b) STROKE or MI or DEATH

(i) negative relationship with survival

PTNAGE	STRKTYP2	SENSORY
PREVTIAS	STRKTYP3	FNSTAT1
SMOKEST	ONSET	MENTATN
ANGINA	MI	ARRYTHMA
CHF	VALVDIS	DIABETES
CDAPEX	CINFRCT	ULCER
STENOSIS	MENTCH	ATAXIAT
DYSPHAT	MUSCLE	DYSART
VISION	MUSCTONE	CEREBR
DPREFX	INFTSIZE	

(ii) positive relationship with survival

SEX	HYPERT	LCAROTID
OCCLSION	EYES	THROAT
FACE	PLREFX	STROKES
CARDSURG	VETEBR	FNSTAT
LIMB	SENSYS	TYPESMKR
RCAROTID	INFCTAPP	VERTIGO
DYSPHST	MUSCPOWR	GAIT

GLOBAL CHI-SQUARE = 83.53 D.F.= 50 P-VALUE=0.0021

TABLE 4. (cont.)

(c) DEATH

(i) negative relationship with survival

PTNAGE	SMOKEST	MI
PREVTIAS	ANGINA	VALVDIS
ARRYTHMA	CHF	ONSET
DIABETES	STRKTYP2	STENOSIS
CINFRCT	ULCER	MUSCLE
DYSPHAT	THROAT	MENTATN
SENSORY	FNSTAT1	GAIT
VISION	DPREFX	INFTSIZE
HYPERT		

(ii) positive relationship with survival

SEX	STROKES	STRKTYP3
CARDSURG	CDAPEX	VETEBR
RCAROTID	LCAROTID	MENTCH
INFCTAPP	OCCLSION	VERTIGO
ATAXIAT	EYES	DYSART
FNSTAT	DYSPHST	MUSCTONE
FACE	LIMB	PLREFX
MUSCPOWR	CEREBR	SENSYS
TYPESMKR		

GLOBAL CHI-SQUARE = 119.08 D.F.= 50 P-VALUE=0.0000

TABLE 4. (cont.)

(d) CV DEATH

(i) negative relationship with survival

PTNAGE	ONSET	VISION
PREVTIAS	CINFRCT	CEREBR
SMOKEST	STENOSIS	DPREFX
HYPERT	DYSPHAT	PLREFX
ANGINA	MUSCLE	INFTSIZE
MI	SENSORY	ARRYTHMA
FNSTAT1	CHF	MENTATN
STRKTYP2	DYSART	

(ii) positive relationship with survival

SEX	STROKES	TYPESMKR
VALVDIS	DIABETES	CARDSURG
CDAPEX	STRKTYP3	RCAROTID
LCAROTID	VETEBR	INFCTAPP
OCCLSION	MENTCH	ATAXIAT
EYES	VERTIGO	THROAT
FNSTAT	DYSPHST	FACE
LIMB	MUSCTONE	MUSCPOWR
SENSYS	GAIT	

GLOBAL CHI-SQUARE = 88.98 D.F.= 49 P-VALUE=0.0004

--

TABLE 5. Significant Factors and Relative Risks of Having An Outcome

(a) STROKE or MI or CV DEATH

STEP	VARIABLE ADDED	COEFFICIENT	STANDARD ERROR	RELATIVE RISK	P-VALUE
1	CHF	.6588	.2163	1.9325	.001
2	ONSET	.8669	.3388	2.3796	.010
3	PREVTIAS	.5470	.2275	1.7281	.025
4	MI	.4969	.2152	1.6436	.029
5	STROKES	-.5432	.2351	.5814	.042
6	ULCER	1.3659	.6026	3.9191	.046
7	DIABETES	.4059	.2331	1.5061	.089

(b) STROKE or MI or DEATH

STEP	VARIABLE ADDED	COEFFICIENT	STANDARD ERROR	RELATIVE RISK	P-VALUE
1	PTNAGE	.0364	.0097	1.0371	.001
2	CHF	.5229	.1875	1.6870	.006
3	ONSET	.6793	.2703	1.0724	.018
4	STROKES	-.5153	.2031	.5973	.022
5	DIABETES	.4892	.1991	1.6310	.020
6	PREVTIAS	.5444	.1992	1.7236	.017
7	ULCER	1.5011	.5291	4.4868	.020
8	SEX	-.4300	.1923	.6505	.023

TABLE 5. (cont.)

(c) DEATH

STEP	VARIABLE ADDED	COEFFICIENT	STANDARD ERROR	RELATIVE RISK	P-VALUE
1	PTNAGE	.0629	.0139	1.6049	.000
2	FNSTAT1	.1608	.0428	1.1744	.004
3	MENTATN	.5601	.2855	1.7509	.020
4	ANGINA	.9545	.2917	2.5974	.040
5	ONSET	1.2219	.4039	3.3937	.025
6	THROAT	.6143	.2438	1.8484	.040
7	VALVDIS	.8117	.4215	2.2517	.064
8	EYES	-.7964	.3343	.4510	.051
9	SEX	-.5369	.2625	.5845	.072
10	VISION	.6431	.2990	1.9023	.070
11	RCAROTID	.4810	.2368	1.6177	.071
12	FACE	-.5652	.2572	.5682	.077
13	FNSTAT	-.1421	.0687	.8676	.060
14	CHF	.5078	.2502	1.6616	.046

TABLE 5. (cont.)

(d) CV DEATH

STEP	VARIABLE ADDED	COEFFICIENT	STANDARD ERROR	RELATIVE RISK	P-VALUE
1	MI	1.0793	.3496	2.9427	.000
2	CHF	1.0775	.3331	2.9374	.002
3	ONSET	1.7688	.7469	5.8636	.012
4	HYPERRT	1.0986	.3904	2.9999	.010
5	LCAROTID	-.8863	.3901	.4122	.017
6	ANGINA	.8725	.3784	2.3928	.035
7	PTNAGE	.0310	.0174	1.0315	.035
8	INFTSIZE	.1913	.0852	1.2108	.064
9	VISION	1.1548	.3800	3.1734	.048
10	MUSCLE	1.2491	.5369	3.4873	.025
11	FNSTAT	-.1944	.1096	.8233	.058

TABLE 6. Results of fitting all 20 pcs into the Cox model

OUTCOME	LOG LIKELIHOOD	GLOBAL CHISQUARE	P-VALUE
Stroke/MI/CVDeath	-542.2453	36.34	.0140
Stroke/MI/Death	-728.8029	48.24	.0000
Death	-411.9829	75.03	.0000
CV Death	-208.4972	54.46	.0000

TABLE 7. Pcs Selected in the IPC procedure

(a) STROKE or MI or CV DEATH

STEP	PC ADDED	RESIDUAL CHI-SQUARE	DF	P-VALUE
1	pc1	74.7762	49	.0103
2	pc2	74.7270	48	.0081
3	pc3	73.3258	47	.0083
4	pc4	72.6164	46	.0074
5	pc5	72.3832	45	.0059
6	pc6	67.9938	44	.0116
7	pc7	66.7026	43	.0117
8	pc8	61.2682	42	.0276
9	pc9	51.5810	41	.1244

(b) STROKE or MI or DEATH

STEP	PC ADDED	RESIDUAL CHI-SQUARE	DF	P-VALUE
1	pc1	76.1046	49	.0078
2	pc2	75.2540	48	.0077
3	pc3	73.5024	47	.0080
4	pc4	72.2162	46	.0081
5	pc5	71.1670	45	.0077
6	pc6	64.3140	44	.0245
7	pc7	62.5358	43	.0273
8	pc8	57.8640	42	.0524
9	pc9	52.4490	41	.1084

TABLE 7. (cont.)

(c) DEATH

STEP	PC ADDED	RESIDUAL CHI-SQUARE	DF	P-VALUE
1	pc1	95.9802	49	.0001
2	pc2	91.1142	48	.0002
3	pc3	87.5722	47	.0003
4	pc4	85.1240	46	.0004
5	pc5	82.4044	45	.0006
6	pc6	76.8848	44	.0016
7	pc7	76.4612	43	.0013
8	pc8	67.5406	42	.0075
9	pc9	66.1228	41	.0077
10	pc10	65.5672	40	.0066
11	pc11	61.6496	39	.0119
12	pc12	60.6122	38	.0113
13	pc13	53.9520	37	.0355
14	pc14	51.6048	36	.0444
15	pc15	47.6810	35	.0748
16	pc16	47.3082	34	.0643
17	pc17	46.8824	33	.0554
18	pc18	42.4298	32	.1028

TABLE 7. (cont.)

(d) CV DEATH

STEP	PC ADDED	RESIDUAL CHI-SQUARE	DF	P-VALUE
1	pc1	88.6964	48	.0003
2	pc2	87.0176	47	.0004
3	pc3	84.4992	46	.0005
4	pc4	81.9144	45	.0006
5	pc5	79.9252	44	.0008
6	pc6	76.2138	43	.0013
7	pc7	76.0048	42	.0010
8	pc8	65.9096	41	.0081
9	pc9	61.3988	40	.0164
10	pc10	60.1670	39	.0163
11	pc11	53.9780	38	.0447
12	pc12	53.9778	37	.0353
13	pc13	48.8774	36	.0744
14	pc14	46.7930	35	.0878
15	pc15	41.9356	34	.1646

TABLE 8. Comparison of the number of variables chosen

OUTCOME	No OF VARIABLES SELECTED	GLOBAL CHI-SQUARE	P-VALUE
Stroke/MI/CVDeath	(a) 7 factors	40.16	.0000
	(b) 9 pcs.	23.75	.0047
Stroke/MI/Death	(a) 8 factors	52.01	.0000
	(b) 9 pcs.	30.77	.0003
Death	(a) 14 factors	91.42	.0000
	(b) 18 pcs.	74.03	.0000
CV Death	(a) 11 factors	65.84	.0000
	(b) 15 pcs.	49.35	.0000

Figure 1(a). Plot of log(-log(cumulative survival)) against survival time for the variable PTNAGE.

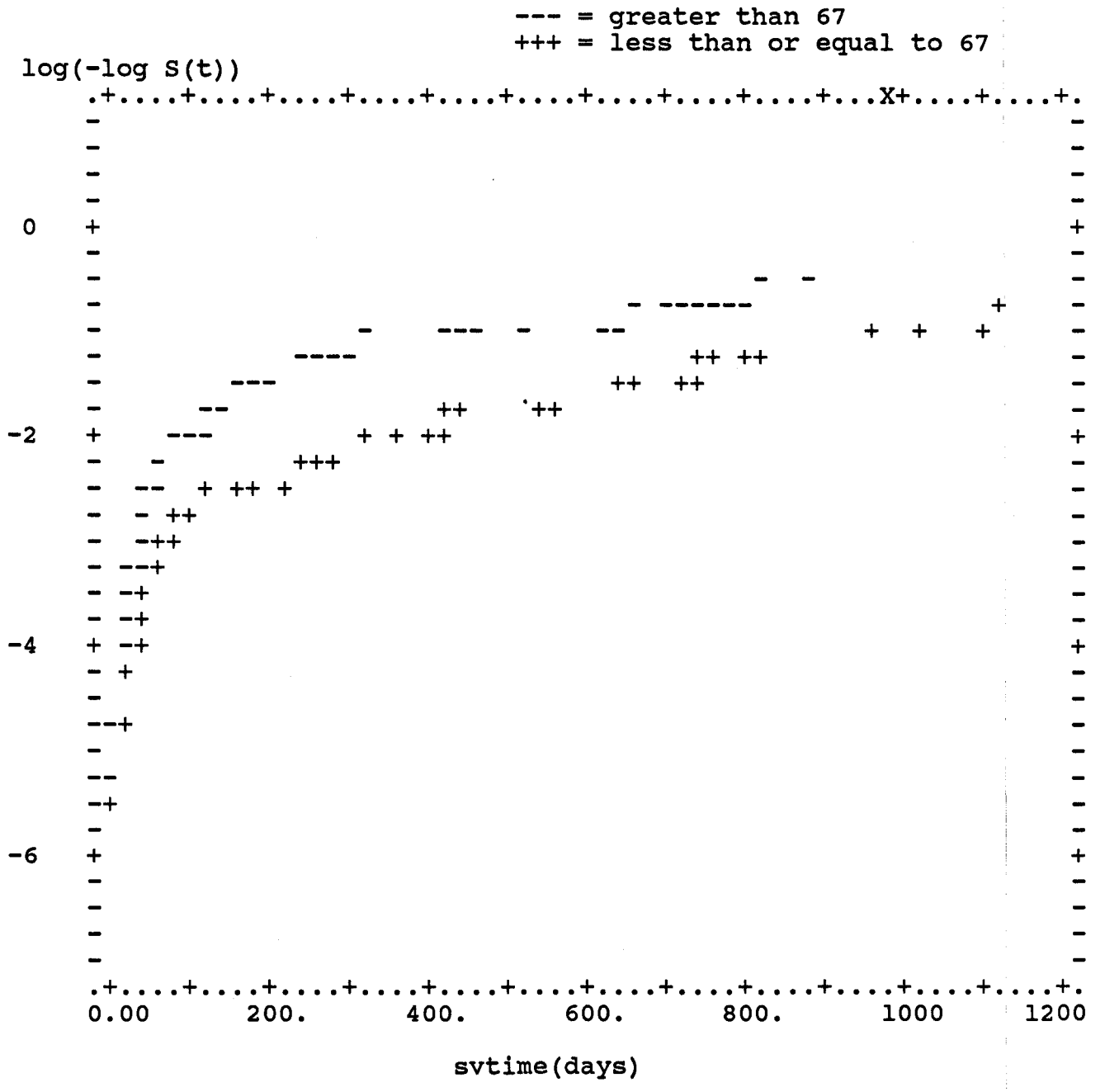


Figure 1(b). Plot of log(-log(cumulative survival) against survival time for the variable STROKES.

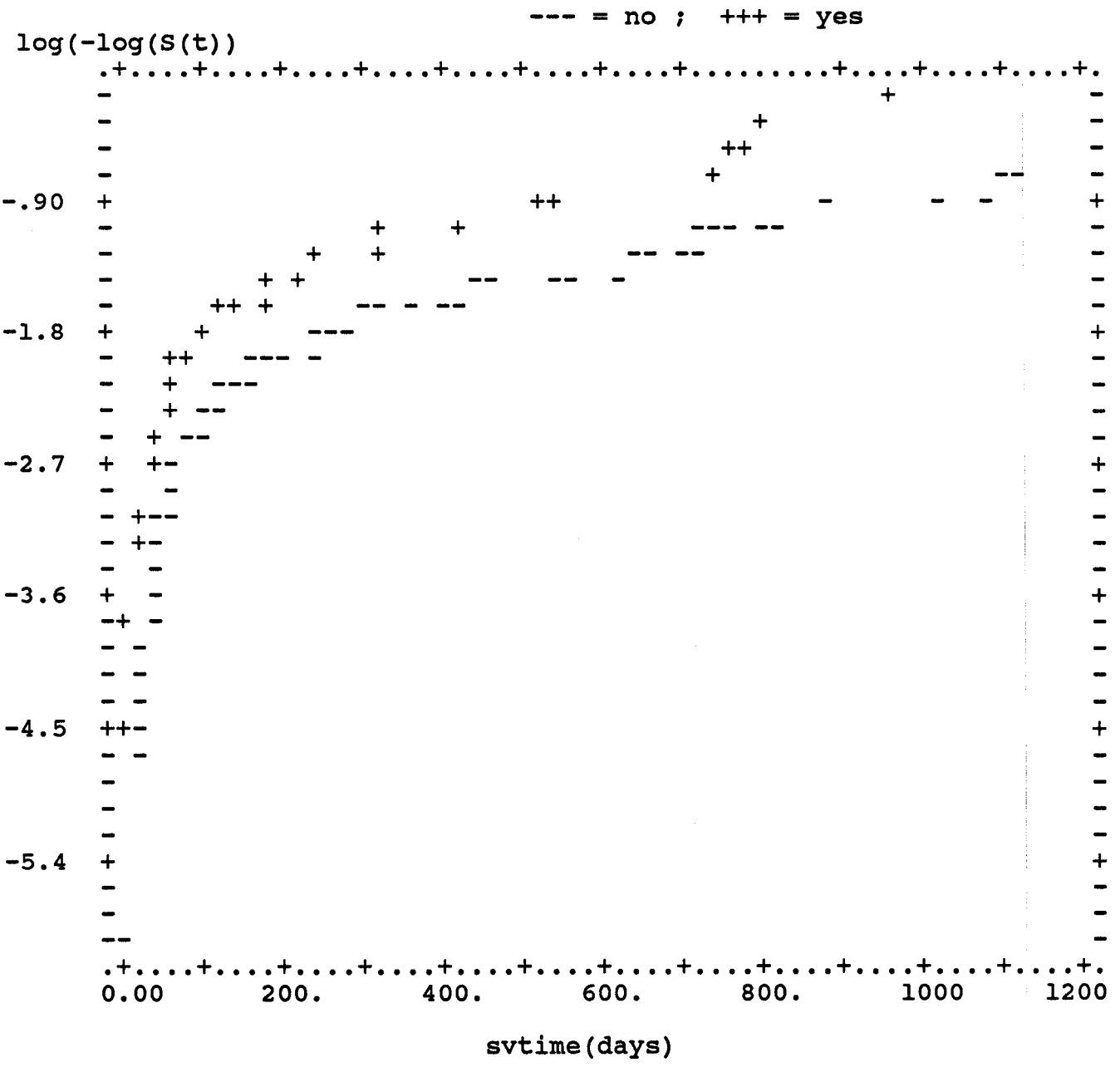


Figure 1(c). Plot of log(-log(cumulative survival) against survival for the variable PREVTIAS.

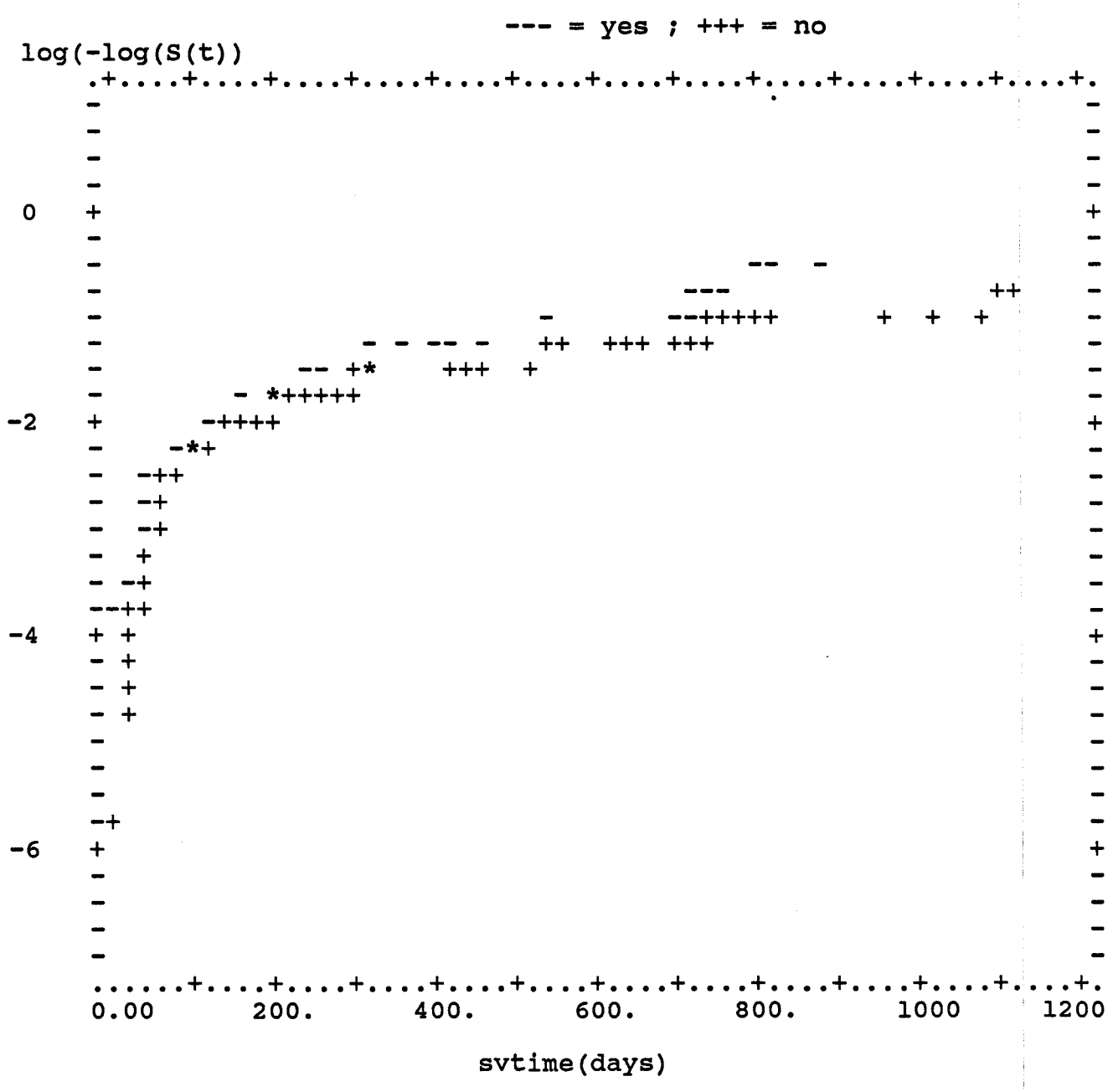


Figure 1(d). Plot of log(-log(cumulative survival) against survival time for the variable CHF.

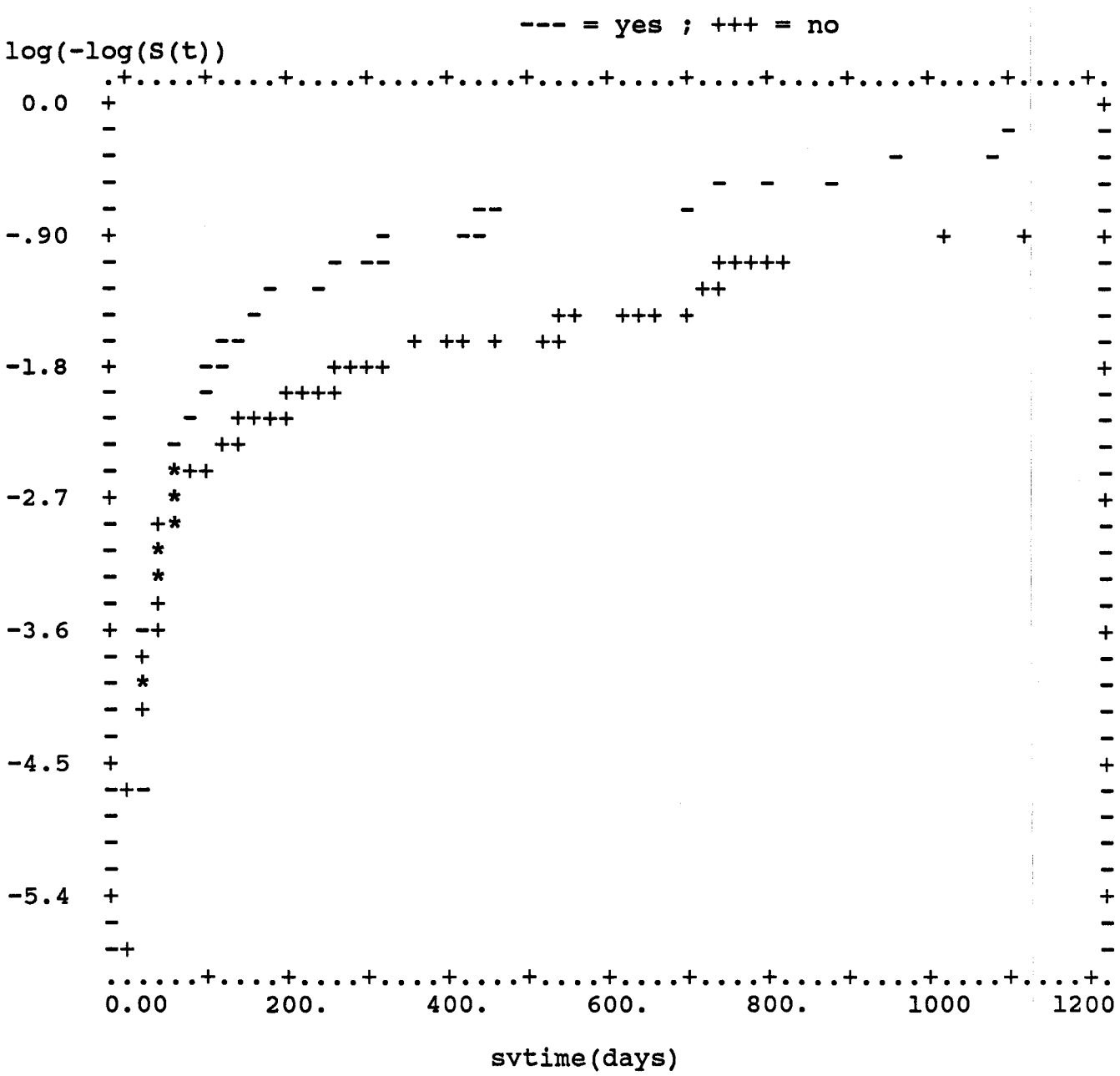
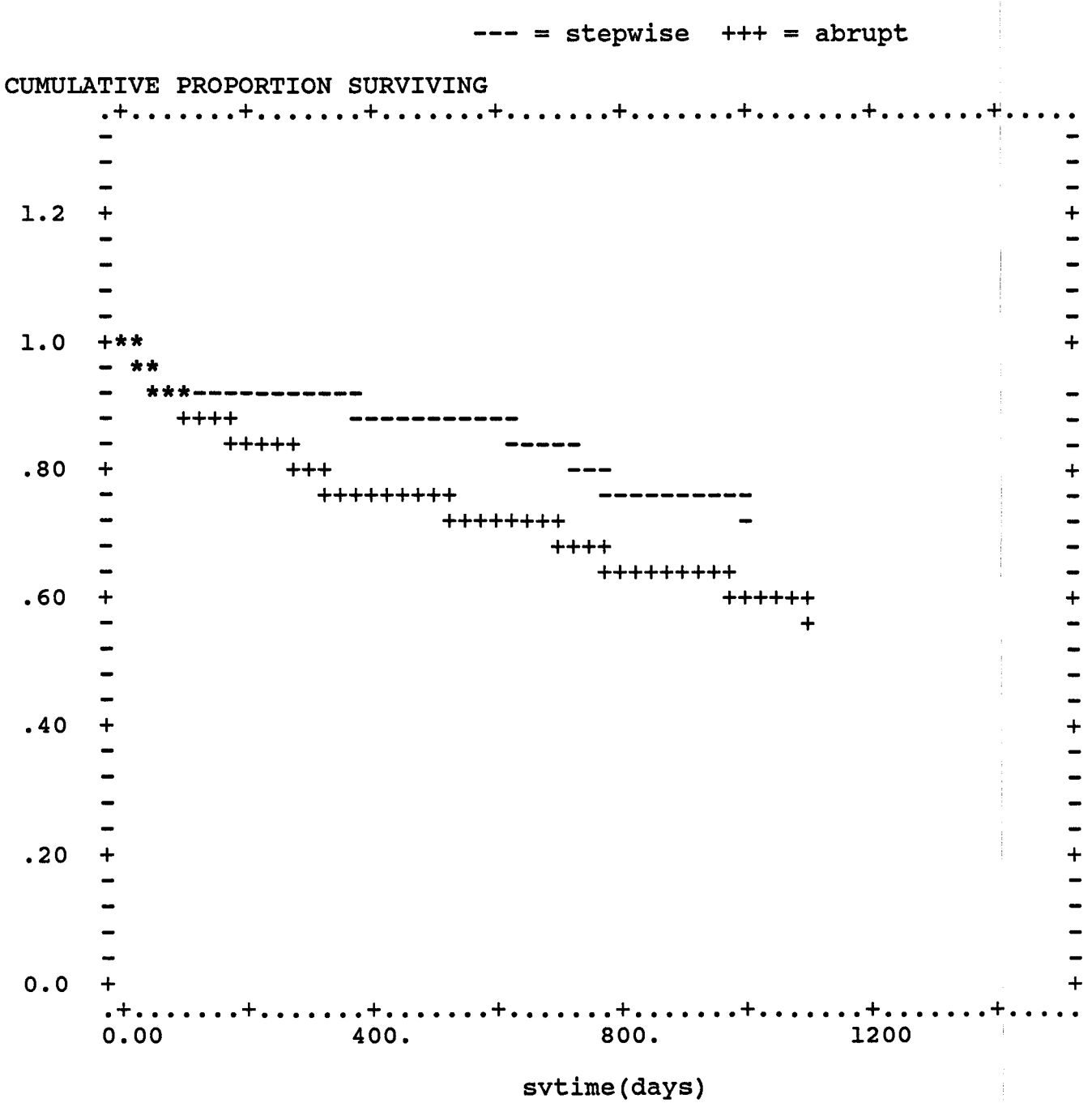
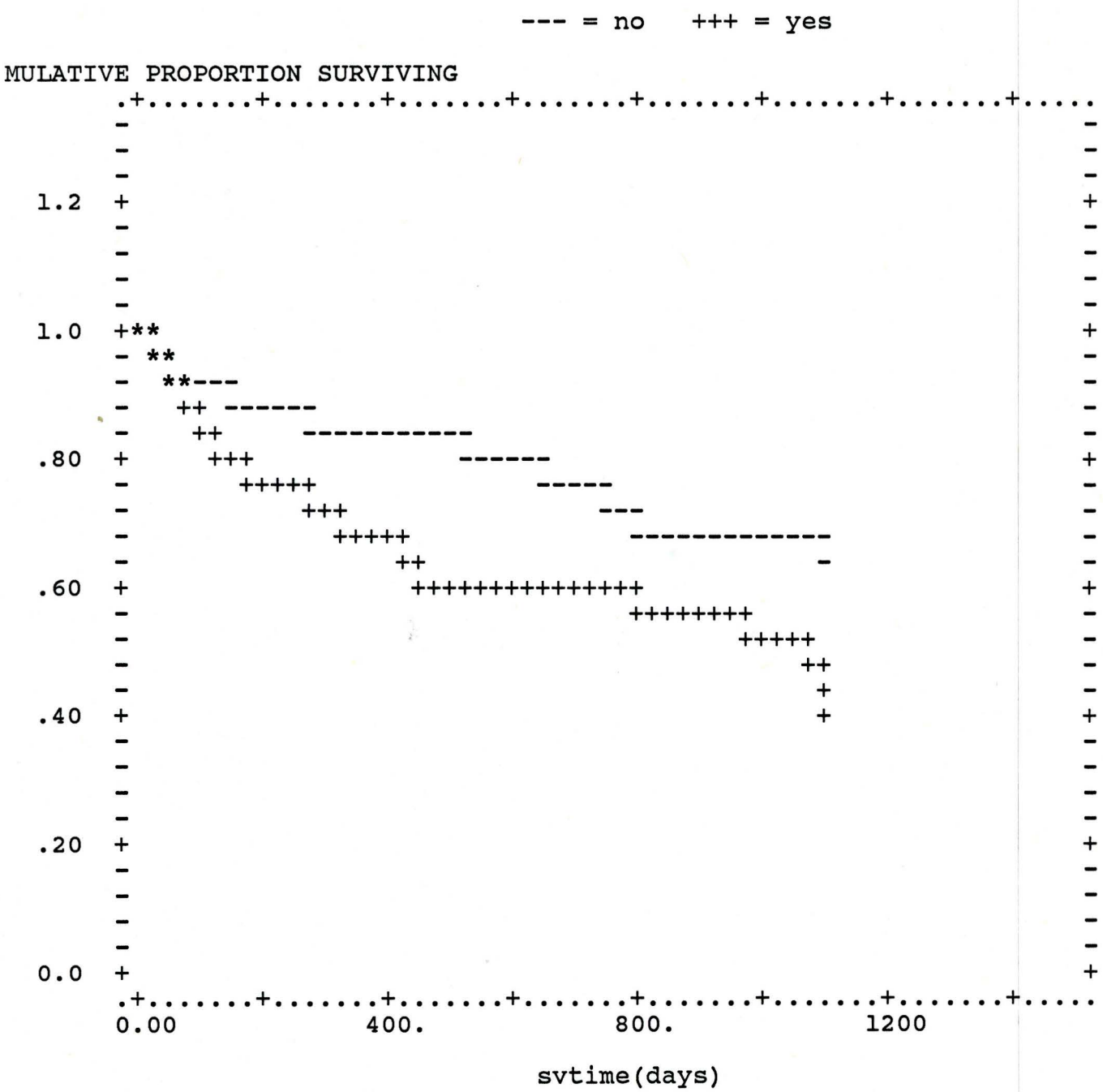


Figure 2. Survival curve for ONSET, for outcome -stroke or mi or death.



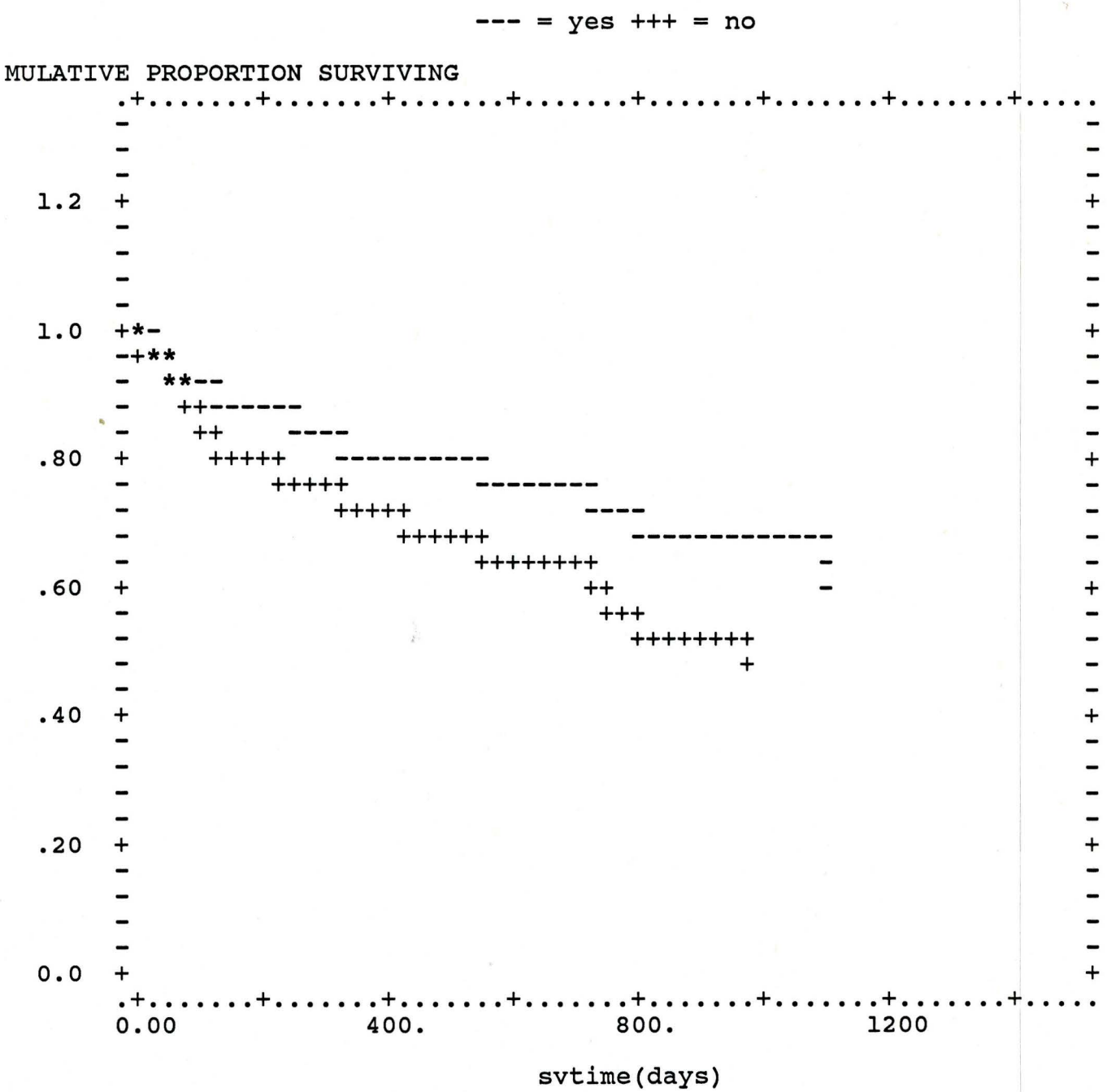
Mantel-Haenszel test p = .0274

Figure 3. Survval curves for CHF for outcome stroke or mi or death.



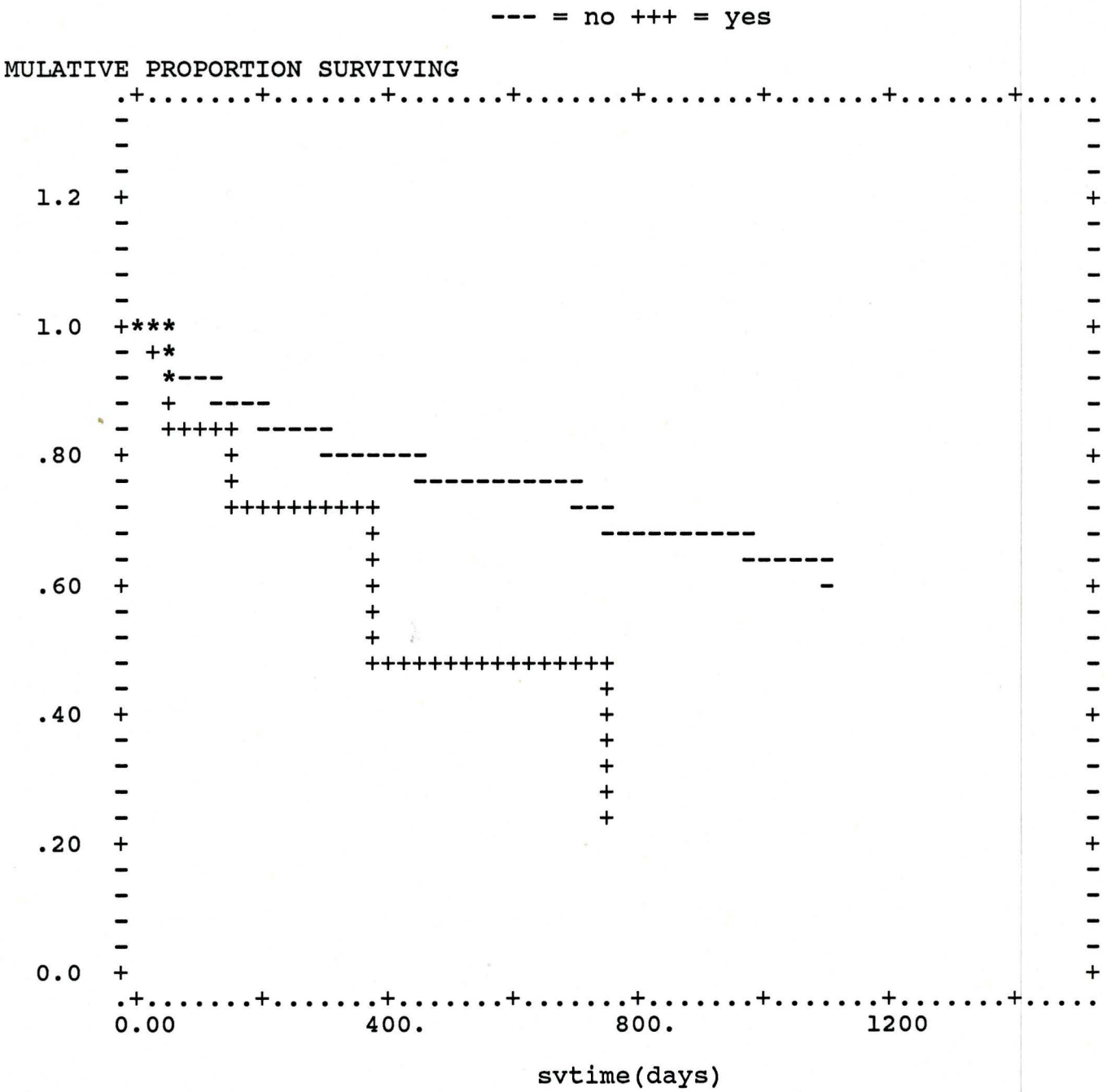
Mantel-Haenszel test p = .0004

Figure 4. Survival curve for STROKES for outcome stroke or mi or death.



Mantel-Haenszel test p= .0110

Figure 5. Survival curve for ULCER for outcome stroke or mi or death.



Mantel-Haenszel test p = .0788

Appendix 1.

(a) Distribution of patients by type of stroke and centre

	Centre				
Type*	Hamilton	London	Toronto	Montreal	TOTAL
I	110	74	127	57	368
II	1	5	26	2	34
III	6	3	15	21	45
TOTAL	117	82	168	80	447

* (I = atheroembolic stroke; II = cardiac embolism; III = lacunr infarction)

(b) Efficacy analysis - excluding events that occurred more than 28 days after complete withdrawal from study

Outcome	Suloctidil	Placebo	p-value*
Stroke/MI/CVdeath	38	47	0.17
Stroke/MI/deaths	47	58	0.08
CVdeaths	10	18	0.06
Deaths	21	29	0.04

* based on the Mantel-Haenszel test.

Appendix 1 (cont.).

(c) Percentage distribution of patients by sex and type of stroke

	Sex		
Type*	male	female	TOTAL
I	81.7	83.2	82.3
II	6.3	9.5	7.6
III	11.9	7.3	10.1
TOTAL	100.0	100.0	100.0

* (I = atheroembolic stroke; II = cardiac embolism; III = lacunar infarction)

Appendix 2. Decription of Variables.

SEX : patient's sex.

PTNAGE :patient's age

STROKES : previous number of strokes(excluding qualifying stroke)

CERESURG : any cerebrovascular surgery(previous carotid endarterectomy or previous cerebrovascular surgery)

PREVTIAS : previous transient ischemic attacks

SMOKEST : smoking status

TYPESMKR : type of smoker

HYPERT : hypertensive(right diastolic blood pressure greater than 95mmHg or history of hypertension)

ANGINA : history of angina

MI : myocardial infarction(history of myocardial infarction or ECG evidence of myocardial infarction)

ARRYTHMA : arrythmia (from cardiac exam or ECG)

CHF : congestive heart failure(history of congestive heart failure or left ventricular hypertrophy or cardiomegaly)

VALVDIS : evidence of valvular disease

CARDSURG : history of by-pass surgery

DIABETES : history of diabetes mellitus

Appendix 2 (cont.)

STRKTYP2 : stroke type 2(cardiac embolic stroke)

STRKTYP3 : stroke type 3(lacunar infarction)

LCAROTID : vascular origin of stroke - left carotid

RCAROTID : vascular origin of stroke - right carotid

VETEBR : vascular origin of stroke - vetebrobasilar

ONSET : onset of stroke - abrupt or fluctuating

CINFRCT : CAT infarct present

INFTSIZE : infarct size

INFCTAPP : angiography done and infarct appropriate to qualifying stroke

ULCER : angiography done and ulcer present

STENOSIS : angiography done and stenosis present

OCCLSION : angiography done and occlsion present

MENTCH : mental changes

DYSPHAT : dysphasia present

ATAXIAT : ataxia present

EYES : eye defects present - presence of diplopia or hemianopia or monocular vision

VERTIGO : presence of vertigo

THROAT : throat ailments present - dysarthria or dysphagia

Appendix 2 (cont.)

MUSCLE : muscle weakness - facial, upper or lower limb on right or left side

SENSORY : sensory weakness - facial, upper or lower limb on right or left side

FNSTAT : functional status - performing each of the following items with difficulty - sitting, eating, getting out of bed, shopping, dress, climbing, employment, and toileting

FNSTAT1 : performing each of above items with difficulty or requiring assistance

MENTATN : intellectual function impaired and related to stroke

DYSPHST : dysphasia related to stroke

VISION : vision defects present - retinal arteriolar occlusion or visual abnormality or extraocular movement on either side - related to stroke

FACE : facial power abnormality on either side - related to stroke

LIMB : any limb abnormality - related to stroke

MUSCTONE : muscle tone abnormality related to stroke

MUSCPOWR : muscle power abnormality related to stroke

CEREBR : cerebrellar function abnormality related to stroke

DPREFX : deep reflex abnormality due to stroke

PLREFX : plantar reflex abnormality due to stroke

Appendix 2 (cont.)

SENSYS : any abnormal symptoms in the sensory system due to stroke

GAIT : abnormal gait due to stroke

CDAPEX : apex rate

DYSART : dysarthria related to stroke

STRKTYP1 : stroke type 1 (atherothrombolic stroke)

Appendix 3. Principal Components Used in the IPC Cox Procedure

```
pc1=((.0886*sex)+(.0786*ptnage)+(-.0093*strokes)+
        (.0525*prevtias)+(-.0963*smokest)+(-.0743*typesmkr)+
        (.0616*hypert)+(-.0246*angina)+(.0080*mi)+
        (.0178*arrythma)+(.0315*chf)+(.0108*valvdis)+
        (.0667*diabetes)+(-.0596*cardsurg)+(.0388*cdapex)+
        (.0391*strktyp2)+(-.0452*strktyp3)+(.1408*rcarotid)+
        (.1267*lcarotid)+(-.2857*vetebr)+(-.0856*onset)+
        (.1353*cinfrct)+(.1471*infctapp)+(-.0767*inftsize)+
        (.0445*ulcer)+(.0674*stenosis)+(.0272*occlsion)+
        (.0098*mentch)+(-.2162*ataxiat)+(.1576*dysphat)+
        (-.1361*eyes)+(-.1567*vertigo)+(.0347*throat)+
        (.2886*muscle)+(.0363*sensory)+(.0587*fnstat)+
        (.2469*fnstat1)+(.0294*mentatn)+(.1760*dysphst)+
        (.0529*dysart)+(-.0912*vision)+(.2349*face)+
        (.3093*limb)+(.2823*musctone)+(.3465*muscpowr)+
        (-.0928*cerebr)+(.0044*dprefx)+(.2678*plrefx)+
        (.1205*sensys)+(.1341*gait)).

pc2=((-.1271*sex)+(-.1110*ptnage)+(.0012*strokes)+
        (-.0557*prevtias)+(.1716*smokest)+(.1048*typesmkr)+
        (-.0397*hypert)+(.0378*angina)+(.0929*mi)+
        (.0095*arrythma)+(-.0167*chf)+(-.0580*valvdis)+
        (-.0093*diabetes)+(.0794*cardsurg)+(-.0014*cdapex)+
        (.0501*strktyp2)+(.0164*strktyp3)+(-.1263*rcarotid)+
        (.3289*lcarotid)+(-.1790*vetebr)+(.0122*onset)+
        (.1895*cinfrct)+(.1974*infctapp)+(-.0709*inftsize)+
        (.0540*ulcer)+(.1385*stenosis)+(.0406*occlsion)+
        (.0123*mentch)+(-.1612*ataxiat)+(.3113*dysphat)+
        (-.0602*eyes)+(-.1724*vertigo)+(-.1872*throat)+
        (-.0883*muscle)+(-.1319*sensory)+(-.2727*fnstat)+
        (-.2681*fnstat1)+(.0268*mentatn)+(.2794*dysphst)+
        (-.1134*dysart)+(-.0646*vision)+(-.0635*face)+
        (-.1302*limb)+(-.0560*musctone)+(-.0367*muscpowr)+
        (-.1925*cerebr)+(-.0505*dprefx)+(-.0531*plrefx)+
        (-.1688*sensys)+(-.2941*gait)).

pc3=((-.1420*sex)+(-.1744*ptnage)+(-.0273*strokes)+
        (.1028*prevtias)+(.2116*smokest)+(.1906*typesmkr)+
        (-.0560*hypert)+(.0411*angina)+(.0238*mi)+
        (.1085*arrythma)+(-.0359*chf)+(.0437*valvdis)+
        (-.0722*diabetes)+(-.0081*cardsurg)+(-.0414*cdapex)+
        (.0024*strktyp2)+(-.1923*strktyp3)+(-.0624*rcarotid)+
        (-.0562*lcarotid)+(.1704*vetebr)+(-.0269*onset)+
        (.2675*cinfrct)+(.2762*infctapp)+(-.1920*inftsize)+
        (.0503*ulcer)+(.0663*stenosis)+(.1713*occlsion)+
```

(.1413*mentch)+(.1206*ataxiat)+(.1126*dysphat)+
(.3473*eyes)+(.0756*vertigo)+(-.0188*throat)+
(-.1005*muscle)+(.1146*sensory)+(.0925*fnstat)+
(.1729*fnstat1)+(.1975*mentatn)+(.1559*dysphst)+
(-.0097*dysart)+(.2842*vision)+(.0804*face)+
(-.0590*limb)+(.1812*musctone)+(-.0932*muscpowr)+
(.1161*cerebr)+(.0701*dprefx)+(-.0342*plrefx)+
(.1670*sensys)+(.2371*gait)).

pc4=((.0780*sex)+(.0314*ptnage)+(-.0203*strokes)+
(-.0307*prevtias)+(-.0713*smokest)+(-.0709*typesmkr)+
(.0234*hypert)+(-.0995*angina)+(-.1609*mi)+
(.0362*arrythma)+(-.0570*chf)+(-.0759*valvdis)+
(.0287*diabetes)+(-.1608*cardsurg)+(.0569*cdapex)+
(.0507*strktyp2)+(-.0954*strktyp3)+(-.3212*rcarotid)+
(.1890*lcarotid)+(.1121*vetebr)+(.0245*onset)+
(-.2908*cinfrct)+(-.2947*infctapp)+(.2391*inftsize)+
(.0846*ulcer)+(.0700*stenosis)+(.0665*occlsion)+
(.0387*mentch)+(.1671*ataxiat)+(.3013*dysphat)+
(-.0107*eyes)+(.1857*vertigo)+(-.2623*throat)+
(-.0002*muscle)+(-.0696*sensory)+(.1024*fnstat)+
(.1656*fnstat1)+(.0391*mentatn)+(.3175*dysphst)+
(-.2647*dysart)+(-.0434*vision)+(-.1173*face)+
(.0256*limb)+(.0941*musctone)+(-.0123*muscpowr)+
(.0679*cerebr)+(.0534*dprefx)+(.0917*plrefx)+
(-.0133*sensys)+(.1266*gait)).

pc5=((.2458*sex)+(.2807*ptnage)+(-.0368*strokes)+
(-.1533*prevtias)+(-.3846*smokest)+(-.3952*typesmkr)+
(.0635*hypert)+(.1039*angina)+(.0832*mi)+
(.0940*arrythma)+(.1384*chf)+(.1357*valvdis)+
(.1178*diabetes)+(.0307*cardsurg)+(.0794*cdapex)+
(.0777*strktyp2)+(-.1631*strktyp3)+(.0872*rcarotid)+
(-.0979*lcarotid)+(.0098*vetebr)+(.0803*onset)+
(.1670*cinfrct)+(.1414*infctapp)+(-.0143*inftsize)+
(-.0422*ulcer)+(-.0379*stenosis)+(.0114*occlsion)+
(.1484*mentch)+(-.0995*ataxiat)+(-.0149*dysphat)+
(.1955*eyes)+(-.0671*vertigo)+(-.2001*throat)+
(-.1373*muscle)+(-.1226*sensory)+(-.0093*fnstat)+
(.0313*fnstat1)+(.0948*mentatn)+(-.0179*dysphst)+
(-.1749*dysart)+(.2427*vision)+(-.1326*face)+
(-.1658*limb)+(-.0062*musctone)+(-.0778*muscpowr)+
(-.1753*cerebr)+(-.0491*dprefx)+(.0349*plrefx)+
(-.0215*sensys)+(-.0123*gait)).

pc6=((.0195*sex)+(.1335*ptnage)+(-.2244*strokes)+
(-.0239*prevtias)+(.0157*typesmkr)+(.3461*hypert)+
(.0852*angina)+(.0740*mi)+(.0211*arrythma)+
(.2313*chf)+(.0712*valvdis)+
(.1296*diabetes)+(.0667*cardsurg)+(-.0171*cdapex)+
(.0658*strktyp2)+(.1534*strktyp3)+(-.2938*rcarotid)+

(.1893*lcarotid)+(.1184*vetebr)+(-.0169*onset)+
(.0269*cinfrct)+(-.0491*infctapp)+(-.0318*inftsize)+
(.0820*ulcer)+(-.0462*stenosis)+(-.1292*occlsion)+
(.2329*mentch)+(.0887*ataxiat)+(-.0002*dysphat)+
(-.0758*eyes)+(.0077*vertigo)+(.3199*throat)+
(-.0370*muscle)+(-.2311*sensory)+(-.0394*fnstat)+
(.0191*fnstat1)+(.2695*mentatn)+(.0535*dysphst)+
(.3410*dysart)+(.0342*vision)+(.1444*face)+
(-.0061*limb)+(.0159*musctone)+(-.0758*muscpowr)+
(.0935*cerebr)+(.1479*dprefx)+(.0617*plrefx)+
(-.2409*sensys)+(.0158*gait)).

pc7=((.1073*sex)+(-.0819*ptnage)+(.0950*strokes)+
(.1220*prevtias)+(-.2287*smokest)+(-.2507*typesmkr)+
(-.0279*hypert)+(.2119*angina)+(-.0186*mi)+
(-.3538*arrythma)+(-.0441*chf)+(-.1828*valvdis)+
(.1279*diabetes)+(.1179*cardsurg)+(-.1474*cdapex)+
(-.2388*strktyp2)+(-.0094*strktyp3)+(-.1818*rcarotid)+
(.1190*lcarotid)+(.0560*vetebr)+(-.1691*onset)+
(.0059*cinfrct)+(.0047*infctapp)+(-.0116*inftsize)+
(.1427*ulcer)+(.3096*stenosis)+(.2613*occlsion)+
(.0787*mentch)+(.0559*ataxiat)+(-.0315*dysphat)+
(.1255*eyes)+(.1065*vertigo)+(.1001*throat)+
(.0611*muscle)+(.2306*sensory)+(-.1670*fnstat)+
(-.1778*fnstat1)+(.0527*mentatn)+(-.0440*dysphst)+
(.0977*dysart)+(.0901*vision)+(-.0666*face)+
(.1127*limb)+(-.0487*musctone)+(.1250*muscpowr)+
(.1189*cerebr)+(-.1348*dprefx)+(.0206*plrefx)+
(.0573*sensys)+(-.0778*gait)).

pc8=((.1101*sex)+(-.0813*ptnage)+(.2301*strokes)+
(-.1892*prevtias)+(-.1078*smokest)+(-.1346*typesmkr)+
(-.0092*hypert)+(-.3091*angina)+(-.3073*mi)+
(.1072*arrythma)+(-.3049*chf)+(-.1662*valvdis)+
(.0375*diabetes)+(-.2589*cardsurg)+(.1139*cdapex)+
(.1405*strktyp2)+(.0742*strktyp3)+(-.0998*rcarotid)+
(.0113*lcarotid)+(.0937*vetebr)+(-.1023*onset)+
(.1705*cinfrct)+(.2383*infctapp)+(-.1772*inftsize)+
(.1067*ulcer)+(-.1279*stenosis)+(-.0850*occlsion)+
(.1108*mentch)+(.1461*ataxiat)+(-.0274*dysphat)+
(-.0399*eyes)+(.1069*vertigo)+(.1269*throat)+
(.0220*muscle)+(-.0963*sensory)+(-.0754*fnstat)+
(-.0472*fnstat1)+(-.0577*mentatn)+(-.0324*dysphst)+
(.1063*dysart)+(-.0303*vision)+(.1003*face)+
(-.0450*limb)+(.0584*musctone)+(.0001*muscpowr)+
(-.0128*cerebr)+(-.2640*dprefx)+(-.1010*plrefx)+
(-.1274*sensys)+(-.0104*gait)).

pc9=((.1157*sex)+(.0711*ptnage)+(-.1179*strokes)+
(.2188*prevtias)+(-.1268*smokest)+(-.1296*typesmkr)+
(-.2317*hypert)+(.1617*angina)+(.0886*mi)+

(.0373*arrythma)+(-.1291*chf)+(-.1009*valvdis)+
(-.1781*diabetes)+(.0556*cardsurg)+(-.2971*cdapex)+
(.1310*strktyp2)+(-.2023*strktyp3)+(-.1294*rcarotid)+
(.0913*lcarotid)+(.0680*vetebr)+(.0649*onset)+
(.1205*cinfrct)+(.0919*infctapp)+(.1239*inftsize)+
(-.1637*ulcer)+(-.1057*stenosis)+(-.0792*occlsion)+
(-.3773*mentch)+(.0810*ataxiat)+(.1339*dysphat)+
(-.0080*eyes)+(.0375*vertigo)+(.1790*throat)+
(-.0220*muscle)+(-.1927*sensory)+(-.0541*fnstat)+
(.1167*fnstat1)+(-.3514*mentatn)+(.1337*dysphst)+
(.1993*dysart)+(.0858*vision)+(.1133*face)+
(-.0368*limb)+(-.0572*musctone)+(-.0648*muscpowr)+
(.0139*cerebr)+(-.0289*dprefx)+(.0297*plrefx)+
(-.0586*sensys)+(.0580*gait)).

pc10=((.0445*sex)+(.0006*ptnage)+(-.1493*strokes)+
(.1154*prevtias)+(.0662*smokest)+(.0374*typesmkr)+
(.0012*hypert)+(.1727*angina)+(.0650*mi)+
(.2678*arrythma)+(.0118*chf)+(.1426*valvdis)+
(-.0083*diabetes)+(-.0649*cardsurg)+(.1448*cdapex)+
(.1817*strktyp2)+(.0316*strktyp3)+(.0208*rcarotid)+
(-.1717*lcarotid)+(.1737*vetebr)+(-.4035*onset)+
(.0522*cinfrct)+(.0020*infctapp)+(-.0349*inftsize)+
(.1905*ulcer)+(.1054*stenosis)+(.3011*occlsion)+
(-.0692*mentch)+(-.0043*ataxiat)+(-.0745*dysphat)+
(-.1041*eyes)+(-.1258*vertigo)+(-.0634*throat)+
(.0399*muscle)+(-.1920*sensory)+(-.0565*fnstat)+
(-.0969*fnstat1)+(-.2215*mentatn)+(-.0892*dysphst)+
(-.1173*dysart)+(-.0094*vision)+(-.1938*face)+
(.1588*limb)+(.0742*musctone)+(.0536*muscpowr)+
(.2736*cerebr)+(-.0702*dprefx)+(.0873*plrefx)+
(-.2884*sensys)+(.0396*gait)).

pc11=((-.2198*sex)+(.0120*ptnage)+(.1856*strokes)+
(-.1940*prevtias)+(.0366*smokest)+(.0081*typesmkr)+
(-.1217*hypert)+(.2898*angina)+(.2705*mi)+
(.1154*arrythma)+(.1244*chf)+(-.2813*valvdis)+
(.1341*diabetes)+(-.0564*cardsurg)+(-.1273*cdapex)+
(.2765*strktyp2)+(-.1591*strktyp3)+(.1528*rcarotid)+
(-.1818*lcarotid)+(.0400*vetebr)+(.0738*onset)+
(-.0694*cinfrct)+(-.0439*infctapp)+(.0794*inftsize)+
(.0489*ulcer)+(-.0738*stenosis)+(-.0166*occlsion)+
(.1768*mentch)+(.0677*ataxiat)+(.0680*dysphat)+
(-.0564*eyes)+(.1462*vertigo)+(.0421*throat)+
(-.0297*muscle)+(-.1304*sensory)+(-.1118*fnstat)+
(-.0130*fnstat1)+(.1835*mentatn)+(.0667*dysphst)+
(.0471*dysart)+(-.2366*vision)+(.0124*face)+
(.0791*limb)+(-.0807*musctone)+(.0779*muscpowr)+
(.0879*cerebr)+(-.3479*dprefx)+(.0618*plrefx)+
(.1620*sensys)+(.0548*gait)).

```
pc12=(-.0655*sex)+(.0178*ptnage)+(-.2270*strokes)+
        (-.2324*prevtias)+(.0718*smokest)+(.0687*typesmkr)+
        (.1663*hypert)+(.1002*angina)+(.0185*mi)+
        (-.3398*arrythma)+(-.0664*chf)+(-.2031*valvdis)+
        (.0510*diabetes)+(-.0182*cardsurg)+(-.0441*cdapex)+
        (-.3852*strktyp2)+(.0963*strktyp3)+(.0623*rcarotid)+
        (-.1513*lcarotid)+(.1163*vetebr)+(-.1191*onset)+
        (.1170*cinfrct)+(.1584*infctapp)+(.0243*inftsize)+
        (.0275*ulcer)+(-.1414*stenosis)+(-.1833*occlsion)+
        (-.1520*mentch)+(.1632*ataxiat)+(-.0303*dysphat)+
        (-.0323*eyes)+(.0649*vertigo)+(-.1649*throat)+
        (-.1360*muscle)+(-.1512*sensory)+(-.1809*fnstat)+
        (.0751*fnstat1)+(-.0643*mentatn)+(-.0010*dysphst)+
        (-.1801*dysart)+(-.1088*vision)+(-.0283*face)+
        (.0010*limb)+(.1161*musctone)+(.0387*muscpowr)+
        (-.0835*cerebr)+(.0126*dprefx)+(.2655*plrefx)+
        (-.0657*sensys)+(.2084*gait)).

pc13=((-.0403*sex)+(.0062*ptnage)+(.0616*strokes)+
        (-.0127*prevtias)+(.0361*smokest)+(.1062*typesmkr)+
        (.0159*hypert)+(-.2352*angina)+(.0373*mi)+
        (-.1193*arrythma)+(-.0573*chf)+(.1052*valvdis)+
        (.0012*diabetes)+(-.1055*cardsurg)+(.1489*cdapex)+
        (.0041*strktyp2)+(-.0090*strktyp3)+(.0877*rcarotid)+
        (-.1632*lcarotid)+(.0645*vetebr)+(-.1741*onset)+
        (-.1972*cinfrct)+(-.1374*infctapp)+(.2997*inftsize)+
        (-.0953*ulcer)+(.1896*stenosis)+(.1615*occlsion)+
        (-.1027*mentch)+(-.1128*ataxiat)+(-.0142*dysphat)+
        (.0983*eyes)+(-.1428*vertigo)+(.1492*throat)+
        (-.1644*muscle)+(-.1145*sensory)+(-.2711*fnstat)+
        (.0603*fnstat1)+(.0644*mentatn)+(.1220*dysphst)+
        (.2498*dysart)+(.2252*vision)+(.1055*face)+
        (-.1819*limb)+(.1718*musctone)+(-.0914*muscpowr)+
        (-.1938*cerebr)+(-.2459*dprefx)+(.2430*plrefx)+
        (.0850*sensys)+(-.0607*gait)).

pc14=((.0473*sex)+(-.1120*ptnage)+(.3603*strokes)+
        (.0652*prevtias)+(.0659*smokest)+(.0214*typesmkr)+
        (.0424*hypert)+(.0867*angina)+(.2152*mi)+
        (-.1020*arrythma)+(.1602*chf)+(.0680*valvdis)+
        (.4104*diabetes)+(-.0397*cardsurg)+(.2314*cdapex)+
        (-.1875*strktyp2)+(-.0948*strktyp3)+(-.0280*rcarotid)+
        (.0599*lcarotid)+(-.0382*vetebr)+(.0978*onset)+
        (-.0536*cinfrct)+(-.0184*infctapp)+(-.1720*inftsize)+
        (.2342*ulcer)+(.0756*stenosis)+(.0138*occlsion)+
        (-.2956*mentch)+(.0224*ataxiat)+(-.0137*dysphat)+
        (.0603*eyes)+(.0333*vertigo)+(.0118*throat)+
        (.0027*muscle)+(-.1987*sensory)+(.2292*fnstat)+
        (.1800*fnstat1)+(-.1682*mentatn)+(.0133*dysphst)+
        (-.0063*dysart)+(-.0212*vision)+(.1847*face)+
        (-.1076*limb)+(-.1095*musctone)+(-.1036*muscpowr)+
```

(-.0499*cerebr)+(-.1093*dprefx)+(-.1460*plrefx)+
(-.1724*sensys)+(.0620*gait)).

pc15=(.0015*sex+.1189*ptnage+.1608*strokes-.0228*prevtias-
.0565*smokest-.0287*typesmkr+.0647*hypert+.0284*angina-
.1020*mi+.0775*arrythma+.3715*chf+.2398*valvdis-
.1667*diabetes+.0447*cardsurg-.1997*cdapex-
.0596*strktyp2+.2322*strktyp3-.0754*rcarotid+
.0456*lcarotid+.0202*vetebr-.2866*onset+.0285*cinfrct+
.0053*infctapp-.0501*inftsize+.1691*ulcer-
.0804*stenosis-.1327*occlsion-.0391*mentch-
.0360*ataxiat+.1515*dysphat-.0892*eyes-.0139*vertigo-
.0443*throat-.0391*muscle+.2321*sensory-.0380*fnstat+
.027*fnstat1-.1945*mentatn+.0772*dysphst-
.0877*dysart-.1054*vision+.1482*face-.2342*limb+
.0020*musctone-.2286*muscpowr+.1296*cerebr-
.2536*dprefx+.0593*plrefx+.2942*sensys+
.0933*gait).

pc16=(.1482*sex+.2219*ptnage-.0374*strokes+.2011*prevtias+
.0982*smokest+.1557*typesmkr+.2283*hypert-
.0700*angina-.3035*mi-.0175*arrythma+.3258*chf-
.0516*valvdis+.1020*diabetes-.3989*cardsurg-
.1498*cdapex-.0193*strktyp2-.1783*strktyp3+
.2136*rcarotid-.1065*lcarotid-.1237*vetebr+
.1407*onset+.1148*cinfrct+.1197*infctapp+
.0796*inftsize-.0694*ulcer+.3044*stenosis-
.0082*occlsion-.0819*mentch+.1461*ataxiat+
.0304*dysphat+.0364*eyes+.1122*vertigo+.0195*throat-
.0704*muscle-.0630*sensory-.1017*fnstat-
.1386*fnstat1-.0452*mentatn+.0560*dysphst+
.0096*dysart-.0580*vision-.0465*face+.0852*limb-
.0897*musctone-.0222*muscpowr+.1109*cerebr+
.0459*dprefx-.1093*plrefx+.0391*sensys-
.1050*gait).

pc17=(-.2455*sex-.1197*ptnage-.2264*strokes-.3620*prevtias-
.1448*smokest-.2100*typesmkr-.1461*hypert-.0826*angina+
.0614*mi-.1719*arrythma+.0755*chf+.0791*valvdis-
.0268*diabetes-.2808*cardsurg-.0547*cdapex-
.1249*strktyp2+.0390*strktyp3+.0171*rcarotid+
.0200*lcarotid+.0066*vetebr+.1437*onset+.0853*cinfrct+
.0476*infctapp+.0553*inftsize-.1893*ulcer+
.1445*stenosis+.2830*occlsion+.0184*mentch+
.0015*ataxiat+.0109*dysphat-.1700*eyes-.2724*vertigo+
.0698*throat+.0457*muscle-.0481*sensory+.1928*fnstat+
.0210*fnstat1-.0220*mentatn+.0597*dysphst-.0051*dysart-
.0837*vision+.1318*face-.0559*limb+.0085*musctone-
.1576*muscpowr+.3129*cerebr-.0277*dprefx-.1628*plrefx-
.0261*sensys+.0365*gait).

pc18=(-.1985*sex-.1210*ptnage-.1616*strokes+.1091*prevtias-
.1529*smokest-.1406*typesmkr-.0081*hypert-.1225*angina+
.1871*mi-.0663*arrythma+.0012*chf+.3511*valvdis+
.1476*diabetes-.0660*cardsurg+.1298*cdapex+
.2240*strktyp2-.1416*strktyp3-.0883*rcarotid-
.0614*lcarotid+.1534*vetebr+.1458*onset+.0076*cinfrct-
.0313*infctapp-.2355*inftsize-.0116*ulcer+
.1026*stenosis-.1064*occlsion-.0681*mentch+
.2792*ataxiat-.0361*dysphat-.0717*eyes+.1469*vertigo-
.0347*throat+.0415*muscle+.1835*sensory-.3173*fnstat-
.1335*fnstat1-.1224*mentatn-.0182*dysphst-
.0293*dysart-.2205*vision+.0761*face+.0419*limb+
.1803*musctone-.0185*muscpowr-.1583*cerebr+
.1197*dprefx+.0701*plrefx+.1410*sensys+
.0629*gait).

pc19=(-.0360*sex-.0353*ptnage+.1445*strokes+.0556*prevtias+
.0405*smokest+.0294*typesmkr+.1798*hypert+.0551*angina-
.0627*mi+.0672*arrythma+.1366*chf-.0061*valvdis+
.1601*diabetes+.0615*cardsurg+.0386*cdapex+
.0916*strktyp2+.3567*strktyp3-.1747*rcarotid+
.0867*lcarotid+.0744*vetebr+.1262*onset+
.0748*cinfrct+.0500*infctapp-.0722*inftsize-
.6338*ulcer-.0996*stenosis+.3269*occlsion-.1473*mentch-
.0279*ataxiat-.0030*dysphat-.0034*eyes+.1716*vertigo-
.1073*throat+.0304*muscle-.0245*sensory-.0180*fnstat-
.0392*fnstat1-.0181*mentatn-.0549*dysphst-.0226*dysart+
.0344*vision-.0661*face+.0839*limb+.0317*musctone+
.0704*muscpowr-.0689*cerebr-.1707*dprefx+.0630*plrefx+
.0401*sensys+.1482*gait).

pc20=(-.3149*sex-.1520*ptnage-.0124*strokes+.1338*prevtias-
.1374*smokest-.1743*typesmkr+.2327*hypert+.1007*angina-
.1337*mi+.0768*arrythma-.0907*chf-.1145*valvdis+
.2842*diabetes-.1232*cardsurg-.1436*cdapex-
.0544*strktyp2-.0314*strktyp3+.1822*rcarotid-
.1168*lcarotid-.0607*vetebr-.2975*onset-.1509*cinfrct-
.1160*infctapp-.1131*inftsize-.0686*ulcer-
.3493*stenosis+.0591*occlsion-.0431*mentch-
.1419*ataxiat+.1767*dysphat+.0831*eyes-.0072*vertigo+
.0437*throat+.0806*muscle-.0753*sensory-.0773*fnstat-
.0493*fnstat1-.0412*mentatn+.1436*dysphst-.0135*dysart+
.1177*vision+.0283*face-.0623*limb+.0249*musctone-
.0953*muscpowr+.0197*cerebr+.2803*dprefx-.1499*plrefx+
.0781*sensys-.1460*gait).

--

REFERENCES

Allen C. M. C.(1984) Predicting the outcome of acute stroke : a prognostic score. **Journal of Neurology, Neurosurgery and Psychiatry,** 47:475-480.

Chen Q., and R. Ling.(1985) A 1-4 year followup study of 306 cases of stroke. **Stroke,** 16:323-327.

Cochran, W.(1954) Some methods for strengthening the common Chi-square tests. **Biometrics** 10:417-451.

Cox, D. R.(1972) Regression models and lifetables. **J. Roy. Stat Soc.,** Series B 34:187-202.

Cox, D. R.(1975) Partial likelihood. **Biometrka** 62:269-276.

Dixon, W. J., M. B. Brown, L. Engelman, J. W. Frane, M. A. Hill, R. I. Jenrich, and J. D. Toporek (eds) (1983) **BMDP Statistical Software,** University of California Press, Berkeley.

Dyken M. L., and P. A. Wolf.(1984) Risk factors in stroke. **Stroke,** 15:1105-1111.

Eisenberg H., J. T. Morrison, P. Sullivan, and F. M. Foote.(1964) Cerebrovascular accidents : incidence and survival rates in a defined population, Middlesex,Connecticut. **J. Am. Med. Assoc.,** 189:883-888.

Friedman, L. M., C. D. Furberg, and D. L. DeMets.(1981) **Fundamentals of Clinical Trials,** John Wright P.S.G.

Gehan, E. A.(1965) A generalized Wilcoxon test for comparing arbitrarily single censored samples. **Biometrika,** 52:203-223.

Greenwood, M.(1926) The natural duration of cancer. **Reports on Public Health and Medical Subjects, 33,** H. M. Stationery Office, London.

Gent M., J. A. Blakely, V. Hachinski, R. S. Roberts, H. J. M. Barnett, N. H. Bayer, S. G. Carruthers, S. M. Collins, M. G. Gawel, M. Giroux-Klimek, M. B. Hopkins, P. Jain, M. Lamy, J. P. Meloche, E. Saerens, J. Sicurella, and A. G. G. Turpie.(1985) A secondary prevention, randomised trial of Suloctidil in patients with a recent history of thromboembolic stroke. **Stroke** 16:416-424.

Goldner, J. C., G. H. Payne, F. R. Watson, and H. M. Parrish.(1967) Prognosis for survival after stroke. **American Journal of Medical Sciences,** 129-133.

Gross, A., and V. Clark.(1975) **Survival Distributions: Reliability Applications in the Biomedical Sciences,** Wiley & Sons, New York.

Harrel, F. E., and K. L. Lee(1984) Regression modelling strategies for improved prognostic prediction. **Statistics in Medicine,**3:143-152.

Kablan S. W., F. A. L. Mathewson, and R. B. Tate.(1978) The relation of blood pressure to stroke prognosis. **Annals of Internal Medicine,** 89:15-20.

Kalbfleisch,J. D., and R. J. MacKay.(1978) Censoring and the immutable likelihood. **University of Waterloo - Department of Statistics, Technical Report 78-09.**

Kalbfleisch, J. D., and R. L. Prentice.(1980) **The Statistical Analysis of Failure Time Data,** Wiley & Sons, New York.

Kaplan, E. L., and P. Meier.(1958) Non-parametric estimation from

incomplete observations. **J. Amer. Stat. Assoc.**, 53:457-481.

Khaw, K., E. Barrett-Connor, L. Suarez, and M. H. Criqui.(1984) Predictors of stroke-associated mortality in the elderly. **Stroke**, 15:244-248.

Lawless, J. F.(1982) **Statistical Models and Methods for Lifetime Data**, Wiley & Sons, New York.

Mantel, N.(1966) Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**, 50:153-170.

Mantel, N., and W. Haenszel.(1959) Statistical aspects of the analysis of data from retrospective studies of disease. **J. Natl. Cancer Inst.**, 22:719-748.

Marquardt, D. W., and R. D. Snee.(1975) Ridge regression in practice. **American Statistician**, 29:3.

Robinson R. W., W. D. Cohen, N. Higano, R. Meyer, G. H. Lukowsky, R. B. McLaughlin, and H. R. MacGilpin.(1959) Lifetable analysis of survival after cerebral thrombosis : 10-year experience. **J. Am. Med Assoc.**, 169:1149-1152.

Sacco R. L., P. A. Wolf, W. B. Kannel, and P. M. McNamara.(1982) Survival and recurrence following stroke : The Framingham Study. **Stroke**, 13:290-295.

Sheikh K., P. J. Brennan, T. W. Meade, D. S. Smith, and E. Goldenberg.(1983) Predictors of mortality and disability in stroke. **Journal of Epidemiology and Community Health**, 37:70-74.

Wolf P. A.(1985) Risk factors for stroke. **Stroke**, 16:359-360.