C.1

STATISTICAL ANALYSIS OF SURVIVAL DATA:

AN APPLICATION TO PERHIPHERAL VASCULAR BYPASS SURGERY

BY

PREETHI NIRMALIE KOTTEGODA

B.Sc., The University of Sri-Jayawardenepura, Sri Lanka, 1981

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
(The Department of Statistics)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
November, 1985

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of STATISTICS

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date 06/01/86

ABSTRACT

A retrospective study was carried out on 535 patients who underwent bypass surgery for peripheral vascular disease. Survival data for 303 patients out of these 535 cases are subjected to quantitative analysis. The main interest is in survival of these patients in order to identify the risk factors. The importance of types of grafting technique in long-term survival is also considered.

Statistical methods used to ascertain the important prognostic variables include Cox's proportional hazards model, stepwise regression and all subsets regression in proportional hazards model discussed by Kuk (1984). In descending order of significance, the most important variables are myocardial infarction, presence or absence of hypertension, sex and whether or not a revision operation was done. The variable, history of a previous coronary bypass graft is highly correlated with survival but the comparison of its significance to the other significant variables is not possible with Cox's model. Age is also related to survival in this data set. However, since there is no control group, one cannot make a strong conclusion about the effect of age on survival of the patients who have had surgery for peripheral vascular disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE

ACKNOWLEDGEMENT

I would like to thank Dr. N. Reid for her guidance and assistance in producing this thesis. I am also indebted to Dr. M. Schulzer for his careful reading and helpful criticisms of this thesis. I would like to express my gratitude to Dr. M.T. Janusz, who kindly provided the data and offered advice on the medical aspects of this thesis.

The financial support of the University of British Columbia is gratefully acknowledged.

INTRODUCTION

Bypass surgery for peripheral vascular disease has been gaining wide acceptance as an effective alternative to amputations. Although there are controversies about the surgical techniques, less attention has been directed to the evaluation of risk factors. As a result of many people being interested in various survival studies, most surgical centres keep follow-up records of the survival experience of their patients. In this study, retrospectively obtained records of one such centre are subjected to quantitative analysis in order to identify factors affecting survival.

Clinical details of the bypass procedures are presented in section 1.1 while the background of data are given in section 1.2. In Chapters 2 and 3, answers to the following questions are sought:

1) What factors are the most important in predicting survival?

2) How does each bypass technique affect survival?

The statistical methods employed to answer these questions are Cox's proportional hazards regression, stepwise regression, all subsets regression in proportional hazards model and contingency table analysis. A method of detecting any influential observations is discussed in Chapter 4. Conclusions and suggestions are given in Chapter 5.

Chapter 1

DETAILS AND BACKGROUND OF DATA

Section 1.1 MEDICAL ASPECTS

Diseases involving peripheral blood vessels, that is blood vessels in the arms and legs,are known as peripheral vascular diseases. Bypass surgery for peripheral vascular disease is a highly accepted surgical treatment. This reduces the number of amputations, which had been the most common surgical procedure that was available. Different types of bypass procedures are used depending on the patient's condition. Each surgeon has somewhat different criteria in selecting patients. Another bias introduced is the surgeon's preference for one surgical technique over another.

While the results of these operative procedures have been studied extensively, less attention has been directed to the evaluation of risk factors. In this study we are interested in survival of patients undergoing surgery for peripheral vascular disease, in order to identify the risk factors. We are particularly interested in survival of the patients with deaths due to cardiac disease in order to identify high, medium, and/or low risk patient groups in the hope of identifying populations who are likely to benefit from aggressive investigation of their heart.

Aorotobifemoral bypass grafting has become the procedure of choice for most patients with occlusive disease of the aortic bifurcation, which is the junction where the abdominal aorta divides into the left and right branches. These two branches are the left and right common iliac arteries. In this type of technique, the graft is extended to the left femoral artery (in the left leg) and the right femoral artery (in the right leg) because aortic flow will be better when both sides are revascularized. By taking the graft to femoral arteries most of the disease is bypassed. The most popular and commonly used grafting material is Dacron, which can be woven or knitted. Usually, a Dacron tube with two limbs is used for Aorotobifemoral grafting. The proximal end of the graft is sutured to a small hole cut in the front of the aorta. This process is called a end-to-side anastomosis. Sometimes the aorta can be completely divided and the proximal end of the graft anastomosed end-to-end. Distally, one limb of the graft is sutured end-to-side to a hole cut in the right femoral artery and similarly, the other limb to the left femoral artery.

The Femoropopliteal bypass procedure is used to bypass occlusion of the superficial femoral artery, when there is an adequate flow in the popliteal artery in the leg. The most acceptable grafting material currently available, is the reversed saphenous vein. This vein possesses valves which only allow the flow of blood towards the heart. It is therefore necessary to remove an appropriate length of the vein and to reverse its direction, before grafting it to the artery. One end of the reversed vein is stitched to a small longitudinal incision

made in the popliteal artery and the other end to a similar cut made in the common femoral artery. Both anastomoses are performed end-to-side.

An aneurysm is an abnormal dilatation of a blood vessel, usually forming a pulsating tumour. Abdominal Aortic Aneurysm is the most commonly seen aneurysm. It consists of weakening of the arterial wall of the aorta so that it is likely to be stretched by the force of arterial blood pressure. When the wall is weakened, the whole vessel tends to dilate but if the vessel wall is weaker over one area, that part of the vessel is liable to blow out and form an aneurysm. Tube graft, end-to-end bifurcation graft from aorta to the right common iliac artery or end-to-side bifurcation graft to the left external iliac artery, are some of the possible types of resconstruction for this disease. Usually, a woven Dacron graft is preferred as the grafting material.

Other types of peripheral vascular operations include Axillofemoral bypass graft in which the axillary artery (in the arm) and the common femoral artery are involved. One end of the graft is stitched on to a small cut made in the axillary artery and the other end to a similar cut made in the common femoral artery. When one iliac artery in a leg is severely occluded and the other iliac artery in the other leg is a suitable donor-vessel, blood can be delivered to the ischemic end via a Femoral-Femoral bypass, Iliac-Iliac bypass or Iliac-Femoral bypass. There are several other operation techniques and bypass procedures for peripheral vascular disease, but the ones described above are the most common. In fact, Aorotobifemoral and

Femoropopliteal procedures account for the majority of bypasses in peripheral vascular disease.

Section 1.2 SOURCE OF DATA AND HOW IT WAS COLLECTED

The data analysed here is a collection of observations and measurements from reports on patients who had undergone peripheral vascular surgery at St. Paul's Hospital (Vancouver, B.C.) between 1975 and 1977. The data is recorded both on data sheets and on individual patient cards and the information contained on them is almost the same except the latter has only the summary.

A retrospective study on 535 patients was performed in October 1981 and information collected on each patient is name, age, sex, type of operation; whether it be Aorotobifemoral grafting (ABF), Femoropopliteal grafting (FP), Abdominal Aortic Aneurysm (AAA) or other peripheral vascular operations, the patient's preoperative symptoms; whether those be ischemia or claudication, whether the patient had a previous vascular operation and whether revisions of peripheral vascular operations were performed. Also recorded are the presence or absence of angina, history of a previous myocardial infarction or a previous coronary bypass graft and the presence or history of diabetes or hypertension. Patient deaths are recorded as being "early" which is within 30 days of surgery or "late" which is beyond 30 days. Cause of death is recorded on data sheets and noted on the cards as being cardiac or non-cardiac. The date of operation and date of death are

recorded by year and month (in 341 cases out of 535) although in some cases the day is also recorded.

The data was recorded manually on data sheets and then a summary of these details was noted on patient cards which are easy to read and handle. In one data sheet, there is information on more than one patient, whereas each patient has exactly one patient card. In February 1985, records of these 535 patients were converted to computer files. Reprints of the data sheet, patient card and the format used for converting to computer files are included in Appendix 1.

Section 1.3 CLEANING UP OF DATA

When the statistical analysis was carried out, all the 535 patients as well as all the variables were not used, for many reasons. There were 89 cases excluded from the study as their year of operation and/or death was unknown. Another 143 cases were deleted because some of their variables had missing observations.

It was noted that some patients had more than one operation type at the initial operation. Hence operation type was partitioned into 15 mutually exclusive subsets as shown in table III. The type OTHER includes peripheral vascular operations other than ABF, FP and AAA. The subsets 9, 10, 11 and 14 were automatically excluded because the patients belonging to those subsets were among the deleted 232 cases. Subsets 5,6,7 and 8 were pooled together and four indicator variables were defined to represent the differences in survival rates between the

five categories of operation type. Pooling was done to avoid having too many variables in the model.

Patients within a data sheet were ordered alphabetically and then their records entered into the computer file according to this order. In the analysis each patient was identified by two labels, namely, the sequence number and page number. The former in the order in which they were entered in to the computer file and the latter is the number corresponding to their data sheet.

Survival times were measured in months rather than in years because the former is more spread out. As noted in section 2.2, using the month or the year did not make any drastic changes in significance of variables nor in the estimated coefficients. There were 72 cases in which the month of operation and/or death was not recorded and in such situations it was assumed that month was June. This was done to avoid further deletion of cases which would have made the sample size small. As noted in section 2.2, assuming the unknown month to be January or December, did not make any drastic changes in significance of variables nor in estimating variable coefficients. Hence throughout the study the unknown month of operation and/or death was assumed to be June.

As we are particularly interested in survival with respect to cardiac disease, all non cardiac deaths and alive patients were treated as censored observations. There were 45 deaths, 255 censored and 3 losts to follow-up, out of 303 cases. From the 255 censored observations, there were 58 non cardiac deaths and 197 alive patients.

Even if the true survived patients were used, the results do not change drastically. This is noted in section 2.2.

The data file in its final form had 15 variables as well as follow-up information. Table I gives all the variable names and their description.

TABLE I. Variables Associated with the Study

VARIABLE NAME	VARIABLE DESCRIPTION
AGE	Age; range is 30 to 97 years.
SEX	Sex; 0 = males 1 = females
ISCH	Symptoms of ischemia; 1 = yes 0 = no
CLAUD	Symptoms of claudication; 1 = yes 0 = no
PVOP	A previous vascular operation done; 1 = yes 0 = no
ANGINA	Presence or absence of angina; 1 = present 0 = absent
MI	History of myocardial infarction; 1 = yes 0 = no
DIAB	History of diabetes; 1 = yes 0 = no
HYPT	History of hypertension; 1 = yes 0 = no
ADDOP	Revisions of peripheral vascular operations; 1 = yes 0 = no
PCBG	Previous coronary bypass graft done; 1 = yes 0 = no
(D1, D2, D3, D4)	(0,0,0,0) if FP only (1,0,0,0) if ABF only (0,1,0,0) if OTHER only (0,0,1,0) if AAA only (0,0,0,1) if ANY TWO
D1	Indicator variable representing the difference between operation type FP and ABF

TABLE I. Variables Associated with the Study (cont'd.)

VARIABLE NAME	VARIABLE DESCRIPTION
D2	Indicator variable representing the difference between operation type FP and OTHER
D3	Indicator variable representing the difference between operation type FP and AAA
D4	Indicator variable representing the difference between operation type FP and ANY TWO

Section 1.4 SUMMARY STATISTICS

Statistics given in the following tables are based on the 15 variables and 535 cases. The values in parentheses correspond to the used 303 cases.

TABLE II. Frequency Distribution of Variables

VARIABLE NAME		Present	Absent	Missing
ISCH		101 (80)	434 (223)	0 (0)
CLAUD		96 (55)	439 (248)	0 (0)
PVOP		106 (72)	429 (231)	0 (0)
ANGINA		79 (45)	456 (258)	0 (0)
MI		100 (65)	435 (238)	0 (0)
DIAB		46 (30)	489 (273)	0 (0)
HYPT		150 (87)	382 (216)	3 (0)
ADDOP		120 (49)	415 (254)	0 (0)
PCBG		15 (6)	520 (297)	0 (0)
D1,D2,D3,D4	= 0,0,0,0	128 (80)	386 (223)	21 (0)
	= 1,0,0,0	97 (60)	417 (243)	21 (0)
	= 0,1,0,0,	53 (34)	461 (269)	21 (0)
	= 0,0,1,0	73 (40)	441 (263)	21 (0)
	= 0,0,0,1	163 (89)	351 (214)	21 (0)
SEX	males = 388 (217),		females = 147 (86)	

Table III. Frequency Distribution of Operation Type

Subset	Subset Name		Frequency
1	FP only		128 (80)
2	ABF only		97 (60)
3	OTHER only		53 (34)
4	AAA only		73 (40)
5	ABF + FP	ANYTWO	15 (8)
6	ABF + OTHER		52 (30)
7	ABF + AAA		30 (10)
8	FP + OTHER		66 (41)
9	FP + AAA		2
10	OTHER + AAA		10
11	ABF + FP + OTHER		3
12	ABF + FP + AAA		0
13	FP + OTHER + AAA		0
14	ABF + OTHER + AAA		6
15	ALL FOUR		0

TABLE IV. Frequency Distribution of Follow-up Information

		Early Death	Late Death	Alive	Unknown	
CAUSE OF DEATH	Non Cardiac	25 (23)	38 (35)	0 (0)	0 (0)	63 (58)
	Cardiac	7 (4)	54 (41)	0 (0)	0 (0)	61 (45)
	Alive	0 (0)	0 (0)	387 (197)	0 (0)	387 (197)
	Unknown	0 (0)	8 (0)	0 (0)	16 (3)	24 (3)
		32(27)	100 (76)	387 (197)	16 (3)	535 (303)

Chapter 2

COX'S REGRESSION MODEL

Section 2.1 GENERAL THEORY FOR COX'S MODEL

There have been many articles in the recent literature on the application of regression analysis to data with censored observations (e.g. Cox, 1972; Miller, 1981; Kalbfleish and Prentice, 1980).

Let T denote the random failure time with a density function f(t) and distribution function F(t). The survival function S(t) is defined to be the cummulative probability of survival past time t and given by

$$S(t) = Pr\ \{T > t\} = 1 - F(t)$$

The hazard function $\lambda(t)$ has the interpretation

$$\lambda(t)\ dt = Pr\{t \leq T \leq t + dt \mid t \leq T\}$$

Then,

$$\lambda(t) = \frac{f(t)}{[1 - F(t)]}$$

Hence we have

$$S(t) = \exp\left\{ -\int_0^t \lambda(x)\, dx \right\}$$

One of the important goals is to estimate the survival function. If the parametric form of f(t) is known and once we have the maximum likelihood estimates for the parameters, s(t) can be estimated. For example, if

$$f(t) = \mu\, e^{-\mu t}, \text{ then } \lambda(t) = \mu \text{ and } S(t) = e^{-\mu t}.$$

If the maximum likelihood estimate of μ is $\hat{\mu}$, then, the maximum likelihood estimate of S(t) is $e^{-\hat{\mu}t}$. However, if the parametric form of f(t) is unknown, a non parametric estimate for S(t) can be obtained using the empirical survival function. If there is no censoring, the empirical survival function based on a sample of size n is given by

$$\hat{S}(t) = \frac{1}{n}\left\{ \text{Number of observations} \geq t \right\} \quad ; \; t \geq 0$$

When dealing with censored data, this equation has to be modified.

Consider n individuals and assume that $t_1 < t_2 < \cdots < t_K$ are K($\leq$n) distinct times at which deaths occur. Let

d_i = number of deaths at time t_i

n_i = number of individuals "at risk" at time t_i^-;

that is the number of individuals alive just prior to time t_i. In addition to life times $t_1, t_2, \cdots, t_k$, there are also censoring times c_j's for individuals whose life times are not observed. Then an estimate of S(t) is defined as

$$S(t) = \prod_{i:\, t_i < t} \left[1 - \frac{d_i}{n_i}\right]$$

This is called the Kaplan-Meier estimate of the survival function and is a kind of a non parametric maximum likelihood estimate. (Kaplan and Meier, 1958). This estimate is a step function with a unit value at t = 0 and drops by a factor $(1 - \frac{d_i}{n_i})$ after $t = t_i$. It does not change at C_j's. However, the effect of censoring times is incorporated into the n_i's and hence, into the sizes of the jumps in $\hat{S}(t)$.

Typically, the failure time depends upon quantitative or qualitative explanatory variables known as covariates, such as age, sex, type of medical treatment. Effects of these covariates on the life times can be studied using a kind of regression model called Cox's model.

Let $\underline{Z}$ be the vector of covariates and $\underline{\beta}$ be a vector of unknown coefficients. Then, Cox's model specifies

$$\lambda\,(t;\, \underline{Z}) = \lambda_o(t)\, \exp\,\{\underline{Z}^T\beta\},$$

where $\lambda(t ; \underline{Z})$ is the hazard rate with covariate vector $\underline{Z}$ and $\lambda_o(t)$ is the hazard rate with $\underline{Z} = \underline{0}$ (Cox, 1972). The regressor variables here are the covariates and changes in these, change the hazard function in a multiplicative way. Such a model is called a proportional hazards model. When $\underline{\beta}$ is estimated and tested for significance, one can finally select a set of significant covariates that would predict the hazard rate.

Estimates of the regression parameters are obtained by maximizing the partial likelihood function given by (Cox, 1975)

$$L(\underline{\beta}) = \prod_{i=1}^{K} \left[\exp\ (\underline{Z}_i^T\ \underline{\beta}) \ / \sum_{j \in R_i} \exp\ (\underline{Z}_j^T\ \underline{\beta}) \right]$$

where

$\underline{Z}_i$ = covariate vector of the i^{th} individual

R_i = set of individuals at risk just prior to t_i

when there are ties among the death times, the partial likelihood function proposed by Breslow (1974):

$$L(\underline{\beta}) = \prod_{i=1}^{K} \left[\exp\ (\underline{S}_i^T\ \underline{\beta}) \ / \left\{ \sum_{j \in R_i} \exp\ (\underline{z}_j^T\ \underline{\beta}) \right\}^{d_i} \right]$$

is maximized. Here, d_i in the number of deaths at time t_i and $\underline{S}_i$ is the vector sum of the covariates of d_i individuals.

Section 2.2 APPLICATIONS AND RESULTS FROM COX'S MODEL

The sample size for this analysis was 303. The variable, HISTORY OF PREVIOUS CORONARY BYPASS GRAFT, was perfectly ordered with time; that is all people who have had a previous coronary bypass had survival times less than 28 months and the patients who had not undergone a coronary bypass graft had survival times greater than 28 months. Hence it is clear that HISTORY OF PREVIOUS CORONARY BYPASS GRAFT is a variable which is highly correlated to survival. Due to the fact that the variable was ordered with time, the partial likelihood is maximized at infinity. Therefore the coefficient cannot be estimated. Since Cox's model cannot be used with such a variable in the model, it was excluded from the computer analyses. The other 14 variables used for this analysis were AGE, SEX, ISCH, CLAUD, PVOP, ANGINA, MI, DIAB, HYPT, ADDOPT, D1, D2, D3 and D4 where D1, D2, D3 and D4 are dummy variables defined in Table I.

The regression analysis was carried out using the computer package BMDP program 2L. The logarithm of the maximized partial likelihood function, the global chi-square and its p-value as well as the estimated coefficients, their asymptotic standard errors and the standardized coefficients for each covariate are presented in Table V. Here, the unknown month of operation and/or death was assumed to be June.

The global chi-square statistic tests the hypothesis that all

coefficients are identically zero. This statistic is defined as

$$U^T(\underline{0})\ I^{-1}\ U\ (\underline{0})$$

where $U(\underline{0})$ represent the vector of first derivative of the partial likelihood function evaluated at $\underline{\beta} = \underline{0}$ and $I(\underline{0})$ denotes the observed information matrix evaluated at $\underline{\beta} = \underline{0}$. The global chi-square has an asymptotic chi-square distribution with degrees of freedom equal to the number of covariates in the model.

The regression coefficient indicates the relationship between the covariate and the hazard function. The effect of a unit change in variable X_i on the hazard function is estimated by $e^{\hat{\beta}i}$; all other X's held fixed. A positive coefficient increases the value of the hazard function and therefore survival deteriorates with increasing values of the variable provided that the covariates are reasonably independent of one another. (A negative coefficient has the reverse interpretation).

TABLE V. Regression Coefficents for Cox's Model

Log likelihood = -218.2350

Global Chi-square = 53.2400, D.F = 14, p-value = 0.0000

VARIABLE NAME	COEFFICIENT	STANDARD ERROR	STANDARDIZED COEFFICIENT	P-VALUE
AGE	0.0368	0.0163	2.27	0.005
SEX	-0.8954	0.4302	-2.08	0.009
ISCH	0.2854	0.3689	0.77	0.180
CLAUD	-0.5094	0.4932	-1.03	0.200
PVOP	-0.2491	0.3819	-0.65	0.780
ANGINA	0.2599	0.3865	0.67	0.340
MI	1.1720	0.3346	3.50	0.000
DIAB	0.4045	0.4135	0.99	0.110
HYPT	0.9028	0.3354	2.69	0.004
ADDOP	-0.7097	0.5018	-1.41	0.09
D1	-0.4222	0.5568	-0.76	0.24
D2	0.7896	0.4536	1.74	0.04
D3	-0.8628	0.5624	-1.53	0.07
D4	-0.2016	0.4242	-0.48	0.34

A p-value of zero for the global chi-square statistic indicates that not all the coefficients are zero. AGE, SEX, MI, HYPT and D2 are highly significant whereas D3 and ADDOP appear to have a fairly significant effect on the hazard function.

The above regression analysis was carried out similarly with the unknown month of death and/or operation assumed to be January and December and the results are shown in Table VI. (The values in parentheses correspond to December). It is clear that the values of the coefficients do not change very much when compared to the values given in Table V.

The significant variables turn out to be the same. Hence all further analyses are done with the assumption of unknown month of death and/or operation to be June.

The same regression analysis was repeated, once with survival times measured in years and true survived patients and again with survival times measured in months and true survived patients. The corresponding results are presented in Table VII and Table VIII respectively. It is clear that in both these tables, the values of the coefficients do not change very much when compared to the values given in Table V. The significant variables turn out to be the same. Hence, we use the survival times in months and consider non cardiac deaths and alive patients as censored observations.

TABLE VI. Regression Coefficients for Cox's Model; Varying Month of Death and/or Operation

Log Likelihood = -222.3807 (-210.2653)

Global chi-square = 52.32 (53.05), D.F = 14, p-value = 0.00 (0.00)

VARIABLE NAME	COEFFICIENT	STANDARD ERROR	STANDARDIZED COEFFICIENT
AGE	0.0355 (0.0358)	0.0162 (0.0160)	2.19 (2.24)
SEX	-0.8929 (-0.8820)	0.4281 (0.4365)	-2.09 (-2.02)
ISCH	0.2523 (0.3553)	0.3677 (0.3705)	0.67 (0.96)
CLAUD	-0.5596 (-0.4384)	0.4921 (0.4886)	-1.14 (-0.90)
PVOP	-0.2202 (-0.2479)	0.3783 (0.3884)	-0.58 (-0.64)
ANGINA	0.2458 (0.2414)	0.3850 (0.3893)	0.64 (0.62)
MI	1.1895 (1.1427)	0.3332 (0.3394)	3.58 (3.37)
DIAB	0.4273 (0.3208)	0.4139 (0.4187)	1.03 (0.77)
HYPT	0.8415 (0.9040)	0.3312 (0.3399)	2.54 (2.66)
ADDOP	-0.6064 (-0.9413)	0.4909 (0.5243)	-1.24 (-1.80)
D1	-0.4991 (-0.3987)	0.5532 (0.5559)	-0.90 (-0.72)
D2	0.7520 (0.7632)	0.4483 (0.4521)	1.68 (1.69)
D3	-0.8791 (-0.8325)	0.5609 (0.5609)	-1.57 (-1.48)
D4	-0.2490 (-0.1465)	0.4183 (0.4290)	-0.60 (-0.34)

TABLE VII. Regression Coefficients for Cox's Model; Survival Time in Years and True Survived Patients.

Log Likelihood = -243.8209

Global chi-square = 59.94, D.F. = 14, p-value = 0.0000

VARIABLE NAME	COEFFICIENT	STANDARD ERROR	STANDARDIZED COEFFICIENT
AGE	0.0362	0.0156	2.32
SEX	-0.8399	0.4123	-2.04
ISCH	0.2206	0.3562	0.62
CLAUD	-0.5209	0.4812	-1.08
PVOP	-0.2106	0.3568	-0.59
ANGINA	0.2303	0.3684	0.63
MI	1.1806	0.3244	3.64
DIAB	0.3916	0.4084	0.96
HYPT	0.9097	0.3107	2.93
ADDOP	-0.6639	0.4912	-1.35
D1	-0.4417	0.5477	-0.81
D2	0.7892	0.4442	1.78
D3	-0.8270	0.5224	-1.58
D4	-0.2354	0.4109	-0.57

TABLE VIII. Regression Coefficients for Cox's Model; Survival Time in Months and True Survived Patients

Log Likelihood = -233.5218

Global chi-square = 54.82, D.F. = 14, p-value = 0.0000

VARIABLE NAME	COEFFICIENT	STANDARD ERROR	STANDARDIZED COEFFICIENT
AGE	0.0364	0.0156	2.33
SEX	-0.8876	0.4325	-2.05
ISCH	0.2630	0.3618	0.72
CLAUD	-0.5318	0.4812	-1.11
PVOP	-0.2311	0.3880	-0.65
ANGINA	0.2581	0.3922	0.66
MI	1.1650	0.3210	3.63
DIAB	0.3921	0.4025	0.97
HYPT	0.9130	0.3218	2.84
ADDOP	-0.6528	0.4931	-1.32
D1	-0.4350	0.5529	-0.78
D2	0.7725	0.4512	1.71
D3	-0.8510	0.5583	-1.52
D4	-0.2182	0.4217	-0.52

TABLE IX. Estimated Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
AGE (1)	1.0													
SEX (2)	-.11	1.0												
ISCH (3)	-.05	-.10	1.0											
CLAUD (4)	-.09	.02	.10	1.0										
PVOP (5)	.01	-.05	.08	.16	1.0									
ANGINA (6)	-.09	.01	.05	.07	-.01	1.0								
MI (7)	-.05	.09	-.02	.04	-.01	-.09	1.0							
DIAB (8)	-.11	.07	-.10	-.09	-.13	.05	-.10	1.0						
HYPT (9)	.03	-.02	.08	-.15	-.18	.07	.11	-.05	1.0					
D1 (10)	.04	-.04	-.07	-.02	-.01	-.08	-.07	.16	.05	1.0				
D2 (11)	.10	.05	-.22	-.07	-.20	-.11	.02	.16	.07	.03	1.0			
D3 (12)	-.11	.06	.14	.13	.16	.11	-.11	.09	-.12	.17	.19	1.0		
D4 (13)	-.09	.02	-.10	.13	-.10	-.15	-.14	.12	-.13	.09	.41	0.02	1.0	
ADDOP (14)	.05	.06	.10	-.04	-.16	.07	-.09	.05	-.05	.02	.07	.09	.06	1.0

According to medical reports it was suspected that ISCH and AGE were correlated. To examine this first order association, a simple contingency table was constructed and the null hypothesis of independence of ISCH and AGE tested by Pearson's chi-square goodness of fit test.

For the purpose of this analysis AGE was categorized into 4 groups. The 2x4 contingency table for ISCH vs. AGE is

AGE

	< 40 yr	41-60 yr	61-80 yr	⩾ 80 yr	
ISCH present	6	139	275	14	434
ISCH absent	0	26	70	5	101
	6	165	345	19	535

with a Pearson χ_3^2 of 3.58 with a significance level of 0.31. From Table IX we have the correlation coefficient between AGE and ISCH as -0.05. Significance of this sample correlation coefficient can be tested using the following test (Anderson, 1984, p.109). If γ is the sample correlation coefficient between two variables, then the null hypothesis of the population correlation between the two variables being equal to zero, is rejected if

$$\frac{(N-2)^{\frac{1}{2}}}{(1-\gamma^2)^{\frac{1}{2}}}|\gamma| > t_{N-2}(\alpha)$$

where N is the sample size and $t_{N-2}(\alpha)$ is the two-tailed

significance point of the t-distribution with (N-2) degrees of freedom for significance level α. Using this test for sample correlation coefficient between AGE and ISCH, the significance level turned out to be 0.2. Similarly the other coefficients between each pair of variables were tested and the significance levels appeared to be in the range of 0.3 - 0.1.

A stepwise logistic regression was also carried out using BMDP program LR, with ISCH as the binary response variable and AGE as the independent variable. This stepwise procedure did not select AGE as a significant variable since the p-value was 0.62. Hence there is no evidence for any association between AGE and ISCH for this data set.

Section 2.3 THEORY FOR STEPWISE REGRESSION IN COX'S MODEL

As a more efficient way of identifying the independent variables which are significantly related to the hazard function, stepwise regression procedure was used.

In the stepwise process significant probabilities are computed on the basis of a large sample partial likelihood ratio test using the chi-square value calculated from the log of the ratio of two maximized partial likelihood functions. This is known as the MPLR method. Let M represent the set of indices of the covariates in the regression model at any given step and L_M denote the maxmized partial likelihood function based on the covariates belonging to set M. The MPLR method

removes the variables corresponding to the index $K \varepsilon M$ for which

$$\chi_1^2 = -2 \ln\left\{\frac{L_M(\hat{\beta})}{L_{M^-}(\hat{\beta})}\right\}; \quad M^- = M - \{K\}$$

is smallest if $Pr(\chi_1^2) >$ limit to remove or enter the variable corresponding to index $K \notin M$ for which

$$\chi_1^2 = -2 \ln\left\{\frac{L_M(\hat{\beta})}{L_{M^+}(\hat{\beta})}\right\}; \quad M^+ = M \cup \{K\}$$

is largest if $Pr(\chi_1^2) <$ limit to enter. The remove and enter limits used for this analysis are 0.15 and 0.10 respectively.

Section 2.4 RESULTS FROM STEPWISE REGRESSION

Computer package BMDP program 2L was used to carry out the analysis. Following this procedure MI was the first variable to enter the model with a χ_1^2 of 19.24. With MI in the model, the variable that was added next is AGE. The χ_1^2 for this stage was 6.72 with a significance level of about 0.009. The next variable to enter was D2 and the χ_1^2 was 3.58 with a significance level of 0.058. HYPT was entered at the fourth step with a χ_1^2 of 3.67, significance level 0.055. The variable SEX which had a χ_1^2 of 4.18 and significance level 0.041 was entered at the fifth stage and the stepwise process terminated after the sixth step in which ADDOP was entered with χ_1^2 of 2.85, significance level 0.092. The coefficient values, their asymptotic standard errors and the

standardized coefficients are given in Table X. These values do not change drastically, when compared to the values given in Table V. Thus at this stage we choose as the model

$$\hat{\lambda}(t ; \underline{Z}) = \lambda_o(t) \exp\{1.33 \times MI + 0.04 \times AGE + 0.77 \times HYPT + 0.89 \times D2 - 0.82 \times SEX - 0.76 \times ADDOP\} \qquad (**)$$

Recall that

SEX = 0 ; males
1 ; females

ADDOP = 0 ; revision operation not done
= 1 ; revision operation done

Hence it is clear that the hazard rate for males is almost twice that for females ($e^{-0.82} = 0.44$) and performing a revision operation tends to halve the hazard rate ($e^{-0.76} = 0.47$). Patients who have had femoropopliteal grafting technique (FP) have a better survival than the patients who had undergone any peripheral vascular surgery belonging to the category "OTHER". This is indicated by the estimated coefficient of D2 ($e^{-0.89} = 2.44$) which is a measurement of the difference in hazard rates between operation type FP and OTHER. The estimated coefficients for MI and HYPT are positive, as expected, since presence

of these is related to poorer patient functioning. A positive coefficient for age implies that the older people tend to die earlier than young ones but since this study does not have a control group, that is there are no age matched patients who have not been operated for the disease, one cannot make a strong conclusion about the effect of age on the survival of the patients who had undergone surgery for peripheral vascular disease.

TABLE X. Parameter Estimates from Stepwise Regression

VARIABLE NAME	COEFFICIENT	STANDARD ERROR	STANDARDIZED COEFFICIENT
MI	1.3261	0.3031	4.38
AGE	0.0400	0.0149	2.68
HYPT	0.7661	0.3168	2.42
D2	0.8893	0.3995	2.23
SEX	-0.8171	0.4217	-1.94
ADDOP	-0.7619	0.4938	-1.54

Section 2.5 CHECKING FOR PROPORTIONALITY ASSUMPTION AND ADEQUACY OF THE FIT IN COX'S MODEL

Recall that the survival function $S(t; \underline{Z})$ is given by

$$S(t ; \underline{z}) = \exp \{ - \int_0^t \lambda(x ; \underline{Z}) \, dx\}$$

Hence with Cox's proportional hazards model we get

$$-\ell n \, S(t ; \underline{Z}) = -\ell n \, S_o(t) \cdot \exp \{\underline{Z}^T \beta\}$$

$$\ell n \, [-\ell n \, S(t ; \underline{Z})] = \ell n [-\ell n \, S_o(t)] + \underline{Z}^T \beta$$

Thus, the logarithm of the minus logarithm of survival function for a particular covariate pattern, when plotted against time is a straight line, if the proportionality assumption is true. When we plot this on the same scale for the categories of a particular variable, such as males and females of variable SEX, then the two lines should be parallel, if the proportionality assumption holds for that variable.

Figures 1 through 6 show the plots of logarithm of the minus logarithm of estimated survival function for the six significant variables, evaluated with the mean covariate vector (Kalbfleish and Prentice, 1980, p. 92). The mean covariate vector has elements which are equal to the mean of each covariate and it was used to avoid having too many plots which would correspond to each possible value of the six

variables. The proportional hazards assumption is met by the variables AGE, SEX, MI and HYPT as the corresponding curves are parallel. The proportionality assumption does not seem to hold for D2 and ADDOP since those curves have slight departures from parallelism.

Once the model is fitted, the overall adequacy of the model can be checked by plotting the survival curve estimates computed from the residuals. The estimated residuals for the i^{th} individual is given by

$$\hat{e}_i = -\ell n\, \hat{S}(t_i\ ;\ \underline{Z}_i)\ ;\ i = 1,2,\cdots,\ n$$

where $\hat{S}(t_i\ ;\ \underline{Z}_i)$ is the estimated survival function for the i^{th}individual (Kalbfleish and Prentice, 1980, p.96). If the model fits the data, the $\hat{e}_i$'s should behave as a random sample of censored unit exponential variates. Thus when the survival curve estimates based on these residuals are plotted on a log scale, the resulting plot should yield approximately a straight line with slope -1. In this analysis, the estimated residuals were obtained from the output of BMDP program IL which computed the Kaplan-Meier survival curve estimates based on the residuals. The corresponding plot which is illustrated in Figure 7, supports the adequacy of the fit.

FIGURE 1. Log minus log survival function for MI

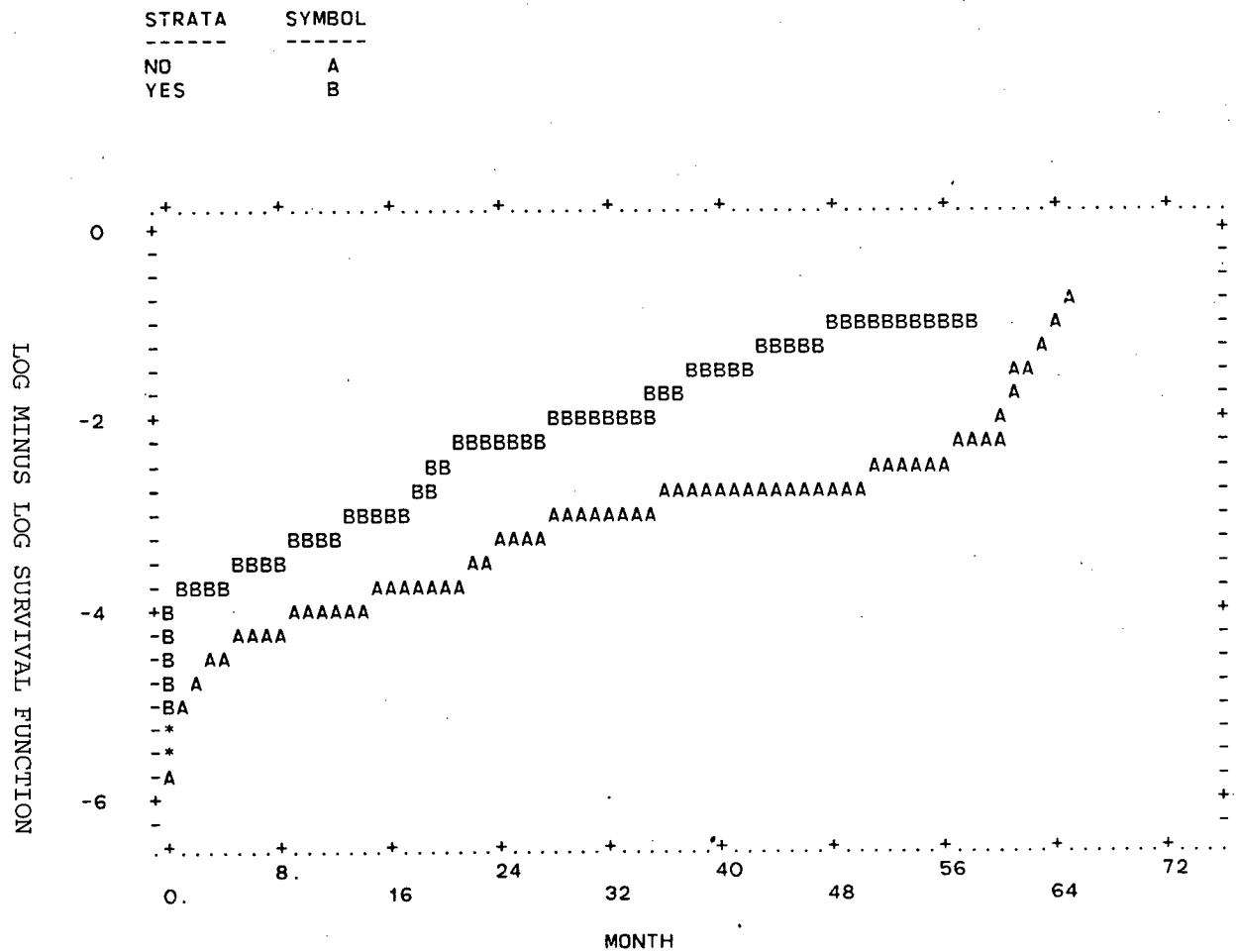


FIGURE 2. Log minus log survival function for AGE

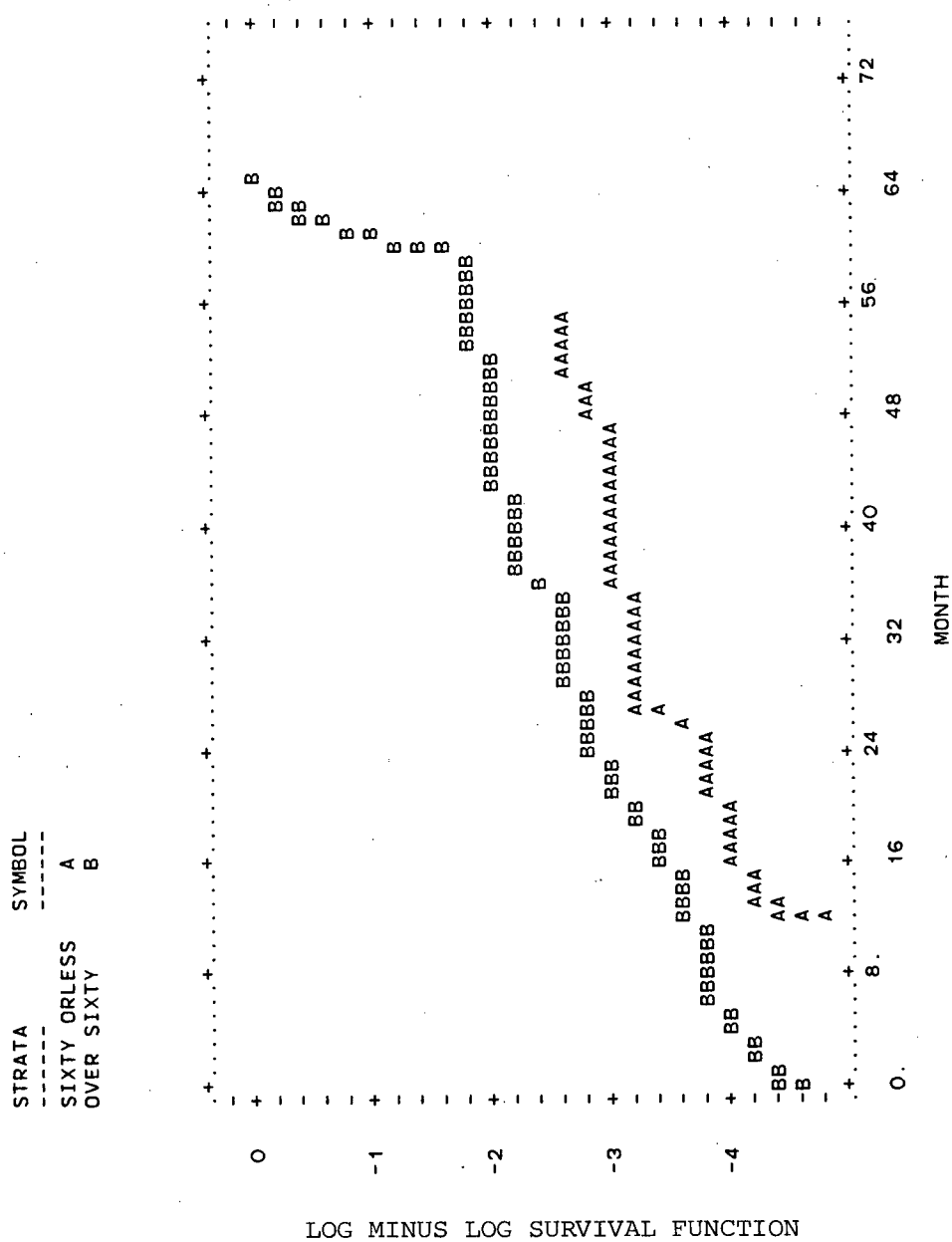


FIGURE 3. Log minus log survival function for HYPT

STRATA	SYMBOL
NO	A
YES	B

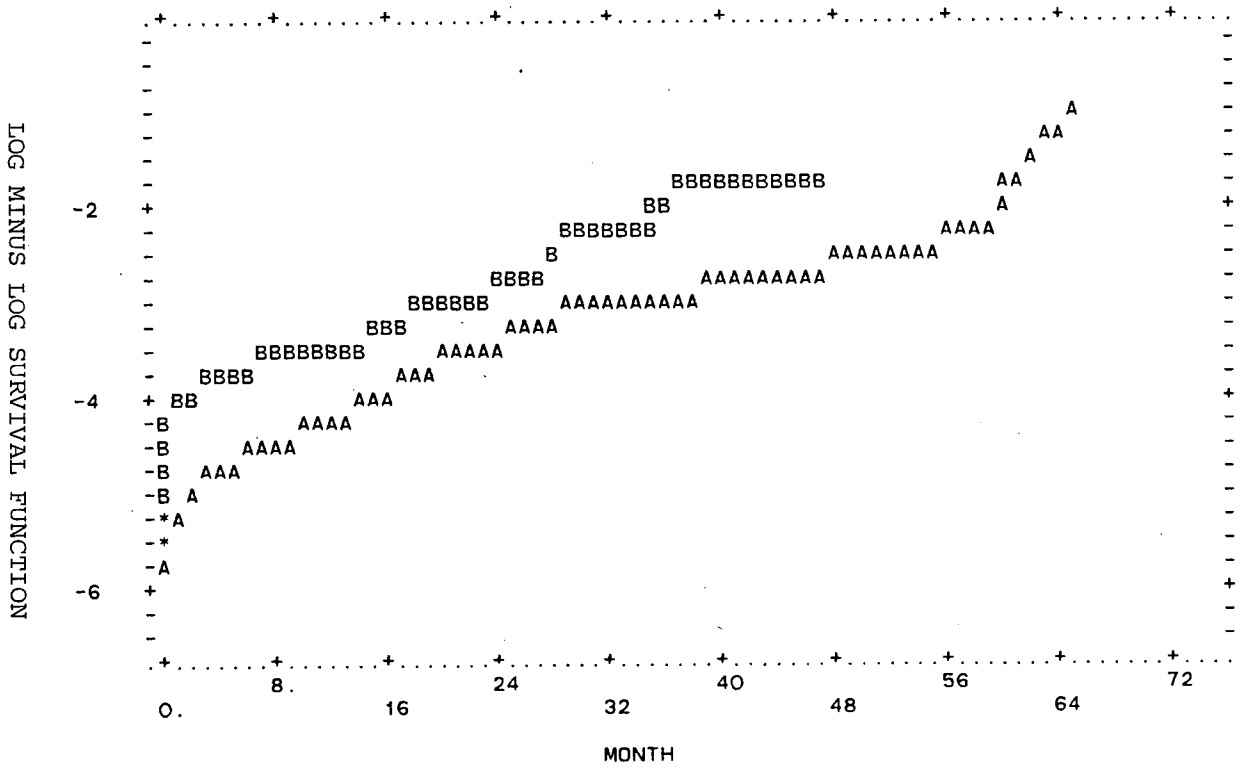


FIGURE 4. Log minus log survival function for SEX

STRATA	SYMBOL
MALE	A
FEMALE	B

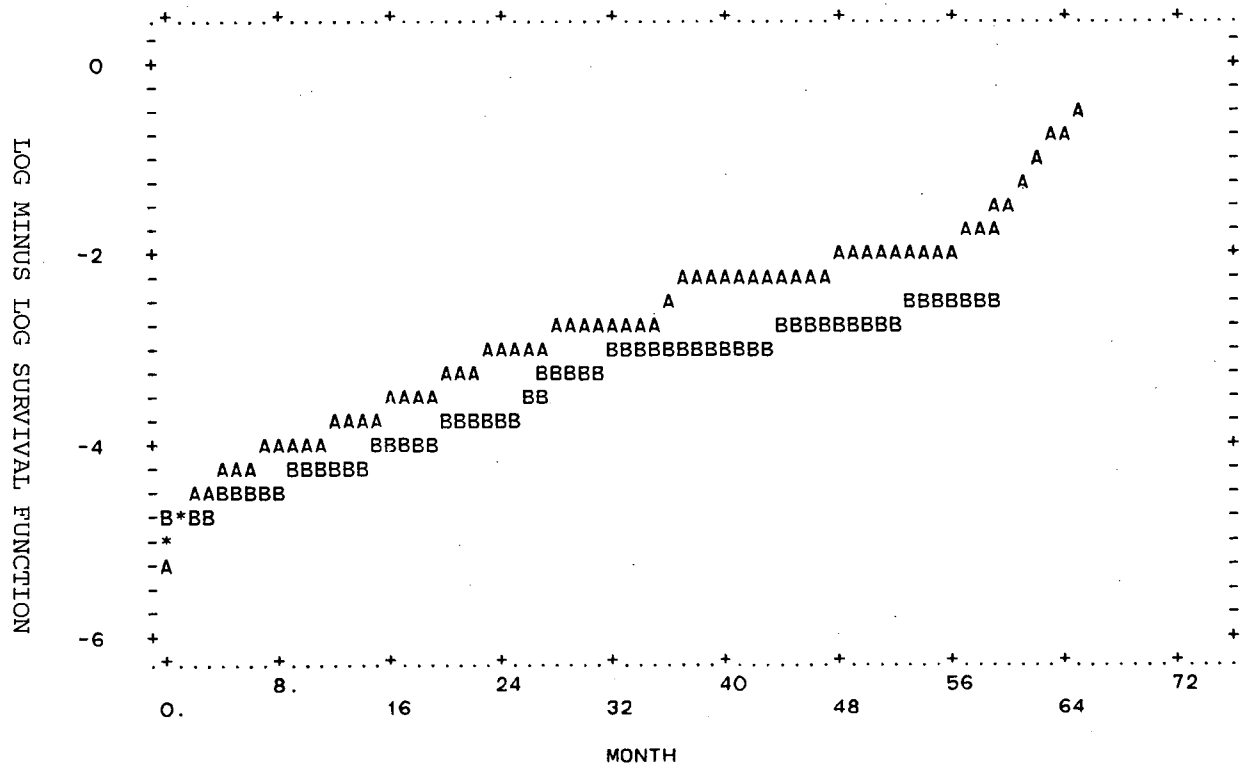


FIGURE 5. Log minus log survival function for ADDOP

STRATA	SYMBOL
NO	A
YES	B

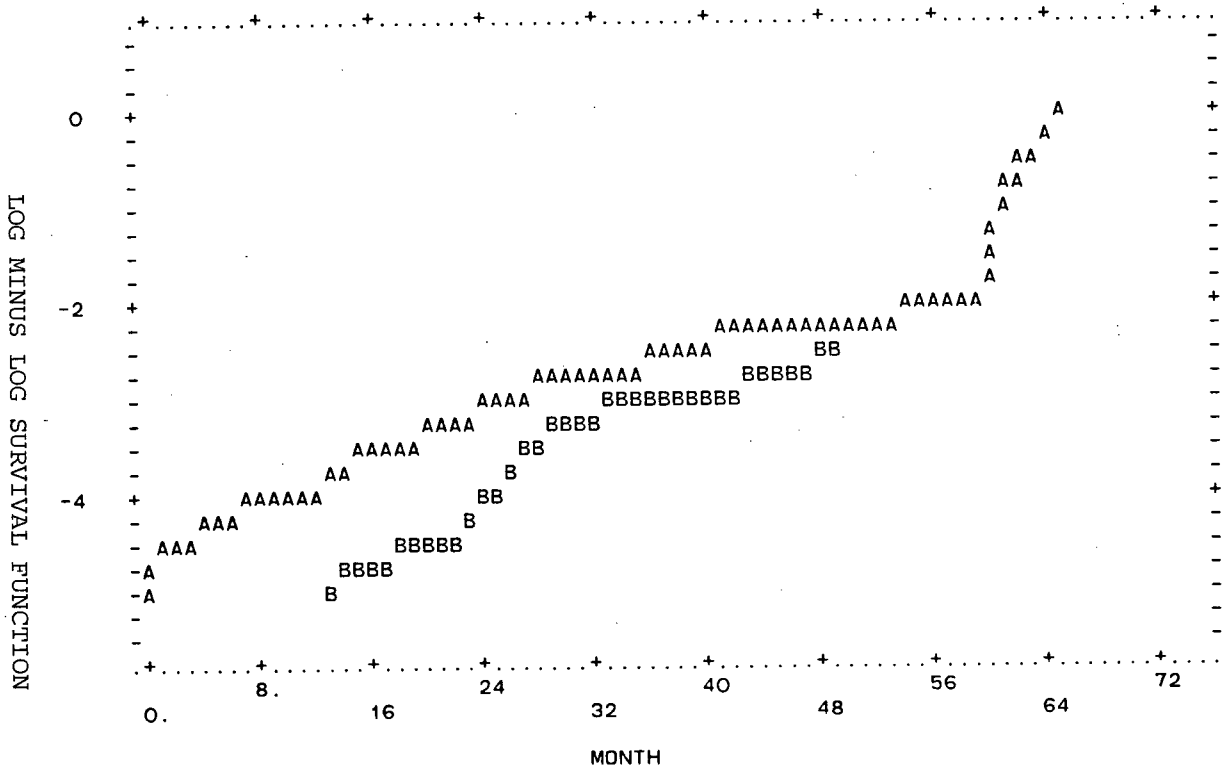


FIGURE 6. Log minus log survival function for D2

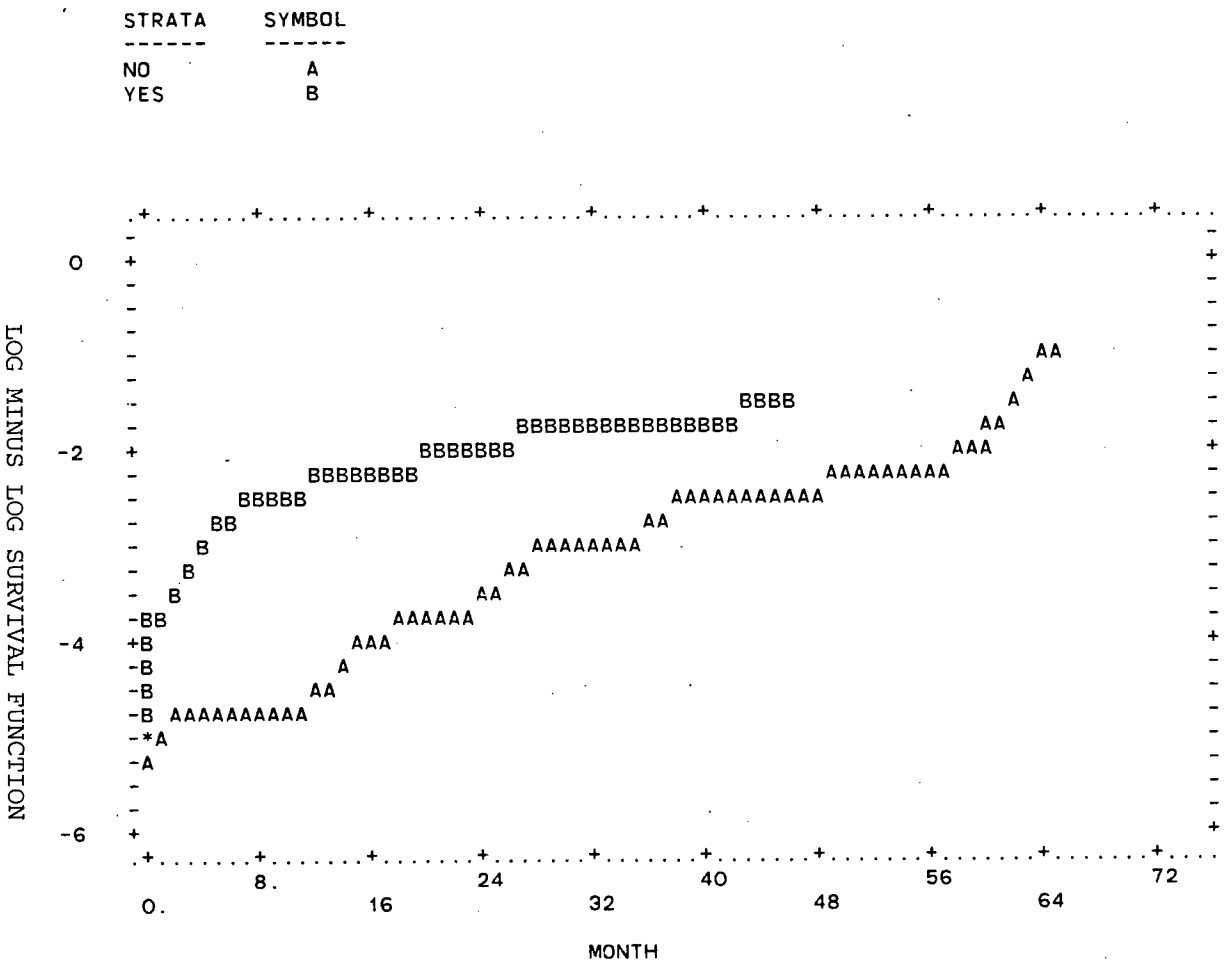
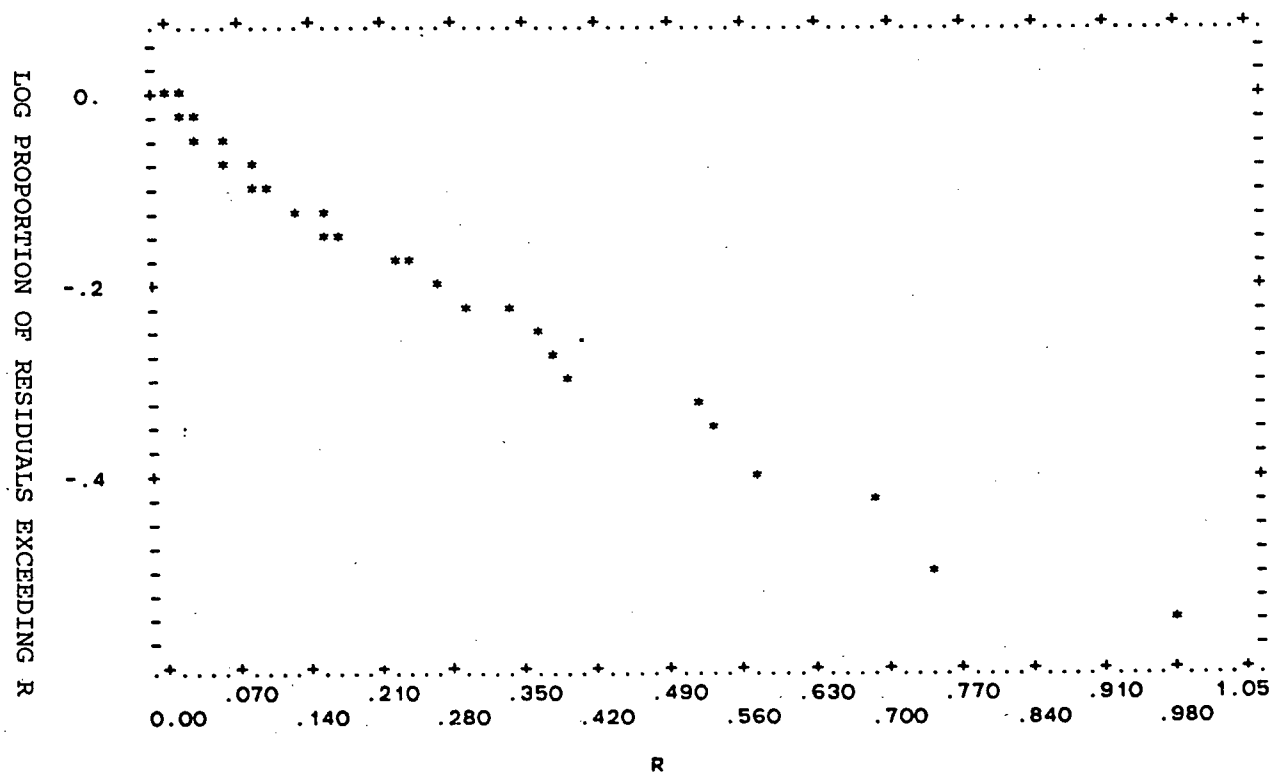


FIGURE 7. Residual plot for checking proportional hazards model



Chapter 3

ALL POSSIBLE SUBSETS REGRESSION IN COX'S MODEL

Section 3.1 THEORY

Although stepwise procedures are often used to select significant variables in regression with censored data, all possible subsets regression is preferred as a more reliable and informative method, provided that it is computationally feasible (Kuk, 1984; Draper and Smith, 1981). This is because stepwise procedures lead to a single subset of variables and do not suggest alternative good subsets. A criterion that is based on the Wald statistic and which is equivalent to Mallow's Cp statistic is used for selecting the best subset.

Consider Cox's proportional hazard model discussed in section 2.1. Let $\beta^T = (\beta_1^T, \beta_2^T)$ and let model α correspond to $\beta_2 = 0$. Then W_α, the wald statistic of the full model against model α is defined as

$$W_\alpha = \hat{\beta}_2^T C_{22}^{-1} \hat{\beta}_2$$

where $\hat{\beta}^T = (\hat{\beta}_1^T, \hat{\beta}_2^T)$ is obtained from the full model and

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} = A^{-1}$$ is also obtained

from the full model as the estimated covariance matrix of $\hat{\beta}$. So, to get W_α,
from the full fit, extract the second component of $\hat{\beta}$ and bottom corner of C; this last needs to be inverted. Then a selection criterion V_α, suggested by Kuk is given as

$$V_\alpha = W_\alpha + 2p_\alpha$$

where P_α is the number of covariates in the model α.

To begin with, the following matrix

$$\begin{pmatrix} A & A^T\hat{\beta} \\ \hat{\beta}^T A & (N-p-1)+\hat{\beta}^T A\hat{\beta} \end{pmatrix} \qquad (3.1.1)$$

where again, $\hat{\beta}$ is obtained from full fit

A^{-1} = C = estimated covariance matrix of $\hat{\beta}$

N is an arbitrary integer > P

was constructed by Kuk in order to show the equivalence of V_α and C_p statistic.

If x,y are the independent and dependent variables from an ordinary multiple regression and M is the matrix of corrected sums of squares and crossproducts defined as

$$M = \begin{pmatrix} x^T x & x^T y \\ y^T x & y^T y \end{pmatrix},$$

then the residual sums of squares, RSS is

$$RSS = y^T y - y^T x \, (x^T x)^{-1} x^T y.$$

By treating (3.1.1) as if it were a matrix of corrected sums of squares and crossproducts of independent and dependent variables computed from a sample size N, the residual sums of squares obtained by this matrix is

$$\begin{aligned} RSS(full) &= (N - p - 1) + \hat{\beta}^T A \, \hat{\beta} - \hat{\beta}^T \, A \, A^{-1} A^T \hat{\beta} \\ &= (N - p - 1) + \hat{\beta}^T \, A \, \hat{\beta} - \hat{\beta}^T A \, \hat{\beta} \\ &= (N - p - 1) \end{aligned}$$

The residual sums of squares for the model α is

$$RSS(\alpha) = RSS(full) + \hat{\beta}_2^T \, C_{22}^{-1} \hat{\beta}_2 \qquad (3 \cdot 1 \cdot 2)$$

and the Mallows' Cp statistic for the model α is

$$Cp(\alpha) = \frac{RSS(\alpha)}{S^2} + 2 \, (P\alpha + 1) - N$$

where

$$S^2 = \frac{RSS(full)}{(N - p - 1)} = 1 \text{ by the choice of } (3\cdot1\cdot1).$$

Substituting (3•1•2) for RSS(α) the above equation for $C_p(\alpha)$ can be simplified as

$$\begin{aligned} C_p(\alpha) &= RSS(full) + \hat{\beta}_2^T C_{22}^{-1} \hat{\beta}_2 + 2(P_\alpha + 1) - N \\ &= (N - p - 1) + (2 - N) + \hat{\beta}_2^T C_{22}^{-1} \hat{\beta}_2 + 2p_\alpha \\ &= -p + 1 + W_\alpha + 2p\alpha \\ &= V_\alpha - p + 1 \end{aligned}$$

Hence it is clear that the criterion V_α is formally equivalent to Mallows' C_p. The problem can now be handled by the standard statistical package BMDP program 9R, which does all possible subsets linear regression. The subset that minimizes C_p is chosen to be the best subset.

Section 3.2 APPLICATIONS AND RESULTS

For this analysis all 14 variables were used. The estimated coefficients and the estimated covariance matrix of $\hat{\beta}$ were obtained

from the output of BMDP program 2L. The estimated covariance matrix was then inverted with the help of a Fortran subroutine. (See Appendix 2) Using this inverted matrix and $\hat{\beta}$, the matrix (3.1.1) was constructed. In this study, (3.1.1) was a 15 X 15 symmetric matrix. The matrix is used as a covariance matrix for input to the BMDP program P9R. In the control language for this program, the value of the sample size N, should be specified in the INPUT paragraph. (See Appendix 2)

The best subset selected by this method was SEX, MI, HYPT, D2 and ADDOP which had a Cp value of 5.18. The second best was the model AGE, SEX,MI,HYPT,D2 and ADDOP with a Cp value of 5.58. The second best was the subset selected by the stepwise procedure. The difference between the Cp values for the best subset and the second best subset is very small. The coefficient for age in the second best subset was 0.0396 and the corresponding standardized coefficient was 2.61. When these values are compared to the corresponding values obtained from stepwise regression, it is clear that AGE is a significant variable.

From the results (discussed in section 2.2) on the significance of correlation coefficients, it appears that there is no evidence for any association between AGE and the other variables. Separate stepwise logistic regressions were carried out for each variable; taken as a binary response and the independent variable as AGE. All these regressions indicated a p-value greater than 0.6 for AGE. Several contingency tables were constructed for AGE vs the other five significant variables and a Pearson's chi-square goodness of fit test was carried out. According to the results presented in Table XI, there

is no evidence for any association between AGE and other variables in this data set. MI, AGE, HYPT D2, SEX and ADDOP were selected as the significant variables for the final model.

TABLE XI. Two-way Contingency Tables

AGE

		< 40 yr	41-60 yr	61-80 yr	>80 yr	
SEX	Male	4	68	139	6	217
	Female	1	23	57	5	86
		5	91	196	11	303

Pearsons's $\chi^2_3 = 2.23$, significance level = 0.53

AGE

		< 40 yr	41-60 yr	61-80 yr	>80 yr	
HYPT	No	4	72	132	8	216
	Yes	1	19	64	3	87
		5	91	196	11	303

Pearsons's $\chi^2_3 = 4.41$, significance level = 0.22

TABLE XI. (continued)

AGE

		< 40 yr	41-60 yr	61-80 yr	>80 yr	
D2	No	4	79	176	10	269
	Yes	1	12	20	1	34
		5	91	196	11	303

Pearsons's χ^2_3 = 0.99, significance level = 0.80

AGE

		< 40 yr	41-60 yr	61-80 yr	>80 yr	
ADDOP	No	5	76	162	11	254
	Yes	0	15	34	0	49
		5	91	196	11	303

Pearsons's χ^2_3 = 3.29, significance level = 0.35

AGE

		< 40 yr	41-60 yr	61-80 yr	>80 yr	
MI	No	4	73	154	7	238
	Yes	1	18	42	4	65
		5	91	196	11	303

Pearsons's χ^2_3 = 1.61, significance level = 0.66

Chapter 4

CHECKING FOR INFLUENTIAL OBSERVATIONS

In some data sets, one of the cases may have sufficient impact upon the regression such that, if that case were deleted, different results would have been obtained. Such cases are known as influential observations. It is suggested that empirical influence functions computed for each covariate and each observation in the proportional hazards regression model, can be useful to identify these influential observations. (Reid and Crépeau, 1985). The theory and method discussed in the above reference was applied to this study. Influence function values are computed for each case (patient) and each covariate. Since it is difficult to consider influence function values for all the 14 variables and 303 observations, attention was restricted only to the six significant variables. The estimated coefficients were obtained from BMDP program 2L and a Fortran program was used to calculate the influence function.

From the summary in Table XII it is seen that case 160 (the case numbers are with respect to all 535 cases) has the largest value of the influence function for covariate HYPT and D2. Observation 1 had the smallest value for covariates MI and D2. Table XIII summarizes the proportional hazards regression models; the first using all observations and the others excluding different cases.

The magnitude of the influence function for each case is roughly consistent with the magnitude of $(\hat{\beta} - \hat{\beta}_{-i})$ where $\hat{\beta}_{-i}$ is the

estimated coefficient when the i^{th} case is deleted. In this study, one unit on the influence function scale correspond to $|\hat{\beta} - \hat{\beta}_{-i}|$ approximately equal to 0.003. From the values given in Table XIII, it is clear that the estimated coefficients and their standard errors do not change very much and this indicates that none of these specified cases seem to have very strong influence on the estimated parameters. This also agrees with the proportional hazards plots and the residual plot of section 2.5 because none of these cases show up on either of these plots.

TABLE XII. Summary of Influence Function Values

Covariate	Maximum of influence function (case no:)	Minimum of influence function (case no:)
AGE	0.5069 (431)	-1.2242 (324)
SEX	44.9926 (225)	-24.0415 (257)
MI	17.2703 (106)	-23.3969 (1)
HYPT	20.1358 (160)	-19.6145 (30)
D2	33.6886 (160)	-45.0780 (1)
ADDOP	64.2133 (46)	-40.4665 (162)

TABLE XIII. Proportional Hazards Regression Model

Estimated coefficient

(standard error)

	AGE	SEX	MI	HYPT	D2	ADDOP
All data with model(**)	0.0400 (0.015)	-0.8171 (0.422)	1.3261 (0.303)	0.7661 (0.317)	0.8893 (0.400)	-0.7619 (0.494)
Case 160 deleted	0.0394 (0.015)	-0.7770 (0.423)	1.3636 (0.306)	0.6943 (0.322)	0.7575 (0.421)	-0.7376 (0.495)
Case 1 deleted	0.0393 (0.015)	-0.8290 (0.421)	1.3894 (0.306)	0.7474 (0.315)	1.0316 (0.401)	-0.7916 (0.494)
Case 46 deleted	0.0398 (0.015)	-0.7916 (0.423)	1.3817 (0.307)	0.8113 (0.320)	0.9340 (0.402)	-0.7502 (0.495)
Case 162 deleted	0.0404 (0.015)	-0.8355 (0.422)	1.3525 (0.303)	0.7962 (0.316)	0.8827 (0.399)	-0.6446 (0.494)
All cases specified in Table XII deleted	0.0495 (0.016)	-0.9765 (0.487)	1.4662 (0.332)	0.6816 (0.344)	1.1120 (0.430)	-0.7984 (0.555)

Chapter 5

CONCLUSIONS

An outline of the surgical techniques of bypass surgery for peripheral vascular disease is presented. The data analysed in this study is based on 303 patients surgically treated for peripheral vascular disease at St. Paul's Hospital, Vancouver, B.C., between 1975-1977. A subset of the recorded variables was used for the analysis due to problems with incomplete records. When the month of death and/or operation was unknown, it was assumed to be June.

Statistical procedures such as Cox's regression, stepwise regression, all subsets regression for the proportional hazards model as well as contingency tables are used to isolate important variables in predicting survival and to discover associations among variables. The conclusions of these analyses are:

1) the most important variables in descending order of their significance are myocardial infarction, presence or absence of hypertension, sex and whether or not a revision operation was done. History of a previous coronary bypass graft is highly correlated with survival but the comparison of its significance to the other significant variables is not possible since the coefficient corresponding to history of a previous coronary bypass graft could not be estimated.

2) age is also related to survival in this data set. However, since there is no control group; that is we do not have a group of age matched patients who have not undergone surgery for peripheral vascular disease, one cannot make a strong conclusion about the effect of age on survival of the patients who have had surgery for peripheral vascular disease.

3) patients who have had Femoropopliteal grafting technique have a better survival than the patients who had undergone any peripheral vascular surgery belonging to the category "OTHER".

4) in this data set hazard rate for males is almost twice that for females.

5) performing a revision operation tends to halve the hazard rate.

6) presence of myocardial infarction or hypertension is related to poorer patient functioning.

7) although pairwise correlation between some of the variables (example; age and ischemia, ischemia and claudication) is suspected, tests used in this study did not indicate it.

One of the difficulties in this study was that there was no control group available. Hence strong conclusions could not be made in

certain instances. The other problem was that the data was not completely recorded, especially the date of death and/or operation. Although there was data for 535 patients, 89 cases had to be deleted because their year of death and/or operation was not known. Then, another 143 cases were excluded from the study since their variables had missing values. Hence, if we had more complete and accurate data, the results could have been more accurate. Also, the type of operation should be clearly specified. In this data set, only the types ABF, FP and AAA were clearly noted. The operation types belonging to category "OTHER" were noted very poorly; especially if the specific operation type in this category was more accurately recorded, one could have seen if there was a difference in survival rates between those types. One could have also checked for interactions between operation type and other variables. The other problem was that in certain cases the hand writing in the data sheets and patient cards was illegible. It would have been much better if the people who were engaged in the survey or the medical staff could have entered the records into computer files and then given them to the statistician for statistical analyses.

BIBLIOGRAPHY

Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis. 2nd Ed. New York: Wiley.

Cooperman, M. Pflung, B., Martin, E.W. and Evans, W.E. (1978). Cardiovascular risk factors in patients with peripheral vascular disease. Surgery 84, 505-509.

Cox. D.R. (1972). Regression models and life tables (with discussion). J.R. Statist. Soc. B 34, 187-202.

Draper, N.R. and Smith, H. (1981). Applied Regression Analysis.

Gaspar, M.R. and Barker, W.F. (1981). Peripheral Vascular Disease. 3rd Ed.

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

Kuk, A.Y.C. (1984). All subsets regression in proportional hazards model. Biometrika, 71, 587-92.

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

Miller R.G. (1982). Survival Analysis. New York: Wiley.

Reid, N. and Crépeau, H. (1985). Influence functions for proportional hazards regression. Biometrika 72, 001-009.

Weisberg, S. (1980). Applied Linear Regression.

Wright, C.B. (1983). Vascular Grafting: Clinical Applications and Techniques.

APPENDIX 1 DATA SHEET

10

CORONARY ARTERY DISEASE — Peripheral vascular study

NAME (+ Next of Kin)	ADDRESS	Unit No. / Tel. No	Age	Sex	Year	Surgeon	Other Dis. (incl. CAD)	SURGERY	Complications	Comments	Add. Surg	Illnesses / Hosp.	Follow-up Cardiac: AMI	Ang/Class	CAB	Meds.	Mortality: Yr.	Etiology
	3716 W 19th Ave Vanc	195615 224-7739	68	M	77	G T	MVA '58 Thromb.	F-PT (G) Ⓡ	Ⓡ B-K Amp.	'79 Ⓛ Profund. Exp Ⓛ tibial	'80 Ⓛ B-K Amp.		UBC [illegible]	[illegible]	[illegible] -OK		—	
	6891 Rupert St. Van	200137 438-7819	56	M	77	G	Claud.	Ⓛ F-P (v)	Seroma				Circ -	100% I	Nov. '80 PA Class I		—	
	Box 47 Falkland	195903 379-2222	72	F	77	M	Claud.	ABF (s)	—	Incisional Hernia '78							LTF	
	426 5th St N West	196088 521-0218	59	M	77	G	Sympt.	AAA	—			none	Circ -OK Heart -OK				—	
	3340 Garden City Way Richmond	198023 278-2951	76	M	77	G	Claud.	Ⓛ F-P (v)	—			[illegible] '79 ?	Circ - [illegible] no SOB or angina				—	
	RR3 Stave Lake Rd Mission	198112 826-7913	55	M	77	M	Insuff Ⓡ Leg Diab.	Ⓡ F-P	Acute GB (D) GB - Etol.	—	'79 Ⓛ F-P		Circ - Cold - Walk	good normal slowly	[illegible] 3-4 [illegible]		—	
	15631 Moffat Lane White Rock	198654 531-0882	64	F	77	M	PVD —	ABF	Hypoxia BP↓ Neurol. ↓								0'	CVA C-R Failure
	5142 Patrick St Burnaby	199325 437-5052	61	M	77	M	Claud. Carotid bruits.	ABF B:1 Sym.	—								LTF	
	31250 King Rd Abtsfd.	199821 853-8191	50	M	Nov 77	G	Claud	Ⓡ F-P T (v)	—				●				'79 6/1/79	[illegible] MSA [illegible] abbotsford

APPENDIX 1 (continued)

PATIENT CARD

CAD — PVD Study

Page No. 45

Name **Age** 66 **Sex** F

- ☐ ABF
- ☐ FP
- ☐ Other

} ASOD Operation

☒ AAA '98

- ☐ Isch.
- ☐ CLaudication
- ☐ Other

} Symptoms

☐ Prev. Vasc. Op.

- ☒ Angina
- ☒ MI '16
- ☐ ACBG

} CAD

coronary angio

☐ Diabetes

☐ Hypertension

5 Duration of follow-up

- ☐ Cardiac
- ☐ Non-Cardiac

} Early Death

- ☐ Cardiac
- ☐ Non-Cardiac

} Late Death

MI - 1 week p.o.

APPENDIX 1 (continued)

FORMAT FOR COMPUTER FILES

(1) SEQUENCE NUMBER:

(2) PAGE NUMBER:

(3) AGE: 99 = Missing

(4) SEX: 0 = Male , 1 = Female

(5) OPERATION TYPE:

ABF	0 = NO , 1 = YES
FP	0 = NO , 1 = YES
AAA	0 = NO , 1 = YES
OTHER	0 = NO , 1 = YES

(6) ADDITIONAL SURGERY: 0 = NO , 1 = YES

(7) SYMPTOMS:

Ischemia	0 = NO , 1 = YES
Claudication	0 = NO , 1 = YES

(8) HISTORY:

Previous vascular operation	0 = NO , 1 = YES
Angina	0 = NO , 1 = YES
Myocardial infarction	0 = NO , 1 = YES
Previous coronary bypass	0 = NO , 1 = YES
Diabetes	0 = NO , 1 = YES
Hypertension	0 = NO , 1 = YES

(9) STATUS OCTOBER '81:

Early death	= 0
Late death	= 1
Still alive	= 2
Unknown	= 9

(10) CAUSE OF DEATH:

Non cardiac	= 0
Cardiac	= 1
Still alive	= 2
Unknown	= 9

(11) DATE OF DEATH: DD/MM/YR

if DD unknown leave blank
if MM unknown leave blank
if YR unknown type 99
if patient is still alive leave all columns blank.

(12) DATE OF OPERATION: DD/MM/YR

if DD unknown leave blank
if MM unknown leave blank
if YR unknown type 99

If information on (3) to (8) is known to be missing, then this was indicated with a code of 9

APPENDIX 2

FORTRAN SUBROUTINE FOR MATRIX INVERSION

```
      REAL*8 DA, DT, DDET, DCOND
      DIMENSION DA (8,8), DT (10,10), IPERM (16)
C **READ IN MATRIX DATA**
      READ (5,10) N
10    FORMAT (I2)
      READ (5,20) ((DA (I,J), I=1, N), J=1, N)
20    FORMAT (F5.0)
C **FIND THE INVERSE**
      CALL INV (N, NDIMA, DA, IPERM, NDIMT, DT, DDET, JEXP, DCOND)
      IF (DDET) 25, 30, 25
C **WRITE OUT RESULTS**
25    WRITE (6,40) N, DDET, JEXP, DCOND
40    FORMAT ('N=', I2, 5X, 'DETERM=', G10.3, '*10**', I2/'INVERSE')
      WRITE (6,50) (( DT(I,J), I = 1,N), J=1, N)
50    FORMAT (1X, 14G10.3)
      STOP
30    WRITE  (6,60)
60    FORMAT ('INVERSION FAILED')
      STOP
      END
```

CONTROL LANGUAGE FOR BMDP:9R PROGRAM

```
/ PROBLEM                    TITLE IS 'PVD DATA'.

/ INPUT                      UNIT = 9.
                             CASES = 303.
                             VARIABLES = 15.
                             TYPE = COVA.
                             SHAPE = SQUARE.
                             FORMAT is '(15F8.3)'.

/ VARIABLE                   NAMES ARE AGE, SEX, ISCH, CLAUD, PVOP,
                             ANGINA, MI,DIAB, HYPT, D1, D2, D3, D4,
                             ADDOP, SURVIVAL .

/ REGRESS                    DEPENDENT IS SURVIVAL.
                             INDEPENDENT ARE 1 to 14.
                             METHOD     = CP.
                             TOLERANCE  = 0.0001.
                             PENALTY    = 2.
                             NUMBER     = 3.
                             ZERO.

/END
```