Nonparametric Simultaneous Modelling of Operative Mortality and Long-Term Survival after Coronary Artery Bypass Surgery

by

Michael Chui

B.Sc.(Hon.), Simon Fraser University, 1999

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY
December 2002

APPROVAL

Name: Michael Chui

Degree: Master of Science

Title of project: Nonparametric Simultaneous Modelling of Operative Mortality and Long-Term Survival after Coronary Artery Bypass Surgery

Examining Committee: Dr. Richard Lockhart
Chair

Dr. Charmaine Dean
Senior Supervisor
Simon Fraser University

Dr. Brad McNeney
Simon Fraser University

Dr. John Spinelli
External Examiner
BC Cancer Agency

Date Approved: December 6, 2002

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Nonparametric Simultaneous Modelling of Operative Mortality and Long-Term Survival after Coronary Artery Bypass Surgery

Author: _____
(signature)

Michael Chu
(name)

Dec 6, 2012
(date)

Abstract

Survival analysis is typically concerned with quantifying lifetimes of individuals, and relating them to covariates. Here two endpoints are of interest, a response state and survival. The response state indicates a positive outcome to an invasive treatment and subsequently, survival time is monitored. The segmentation of survival experience into two parts is helpful in the coronary artery bypass application considered which records (i) operative survival, defined as survival 30 days after coronary artery bypass surgery and (ii) long-term mortality, given operative survival. With coronary artery bypass surgery there is a high risk of operative mortality, i.e. death within 30 days of surgery. It is of interest to explore the effects of covariates on both of the outcomes (i) and (ii). In a previous analysis, Ghahramani et. al. (1999) used a fully parametric approach combining the logistic regression model, for analyzing operative mortality, and the Weibull regression model, for analyzing long-term survival. Although the Weibull is a flexible model, and seemed to give a fair fit in that analysis, a more robust approach is applied here using the Cox's proportional hazards (PH) model. We also make linkages with multi-state modelling. The model will be broadly useful in assessing the joint influence of covariates on surviving the application of a severe treatment and, on treatment success, in terms of subsequent survival experience.

Acknowledgments

I would like to thank my supervisor, Dr. Charmaine Dean, for her guidance and advice, and her kindness in supporting me throughout my study.

I would also like to thank Dr. John Spinelli of the BC Cancer Agency for providing data as well as for very helpful discussions.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cardiac bypass surgery is a common surgical treatment for severe blockages of the arteries of the heart. This type of surgery poses a substantial risk of operative mortality, defined as death within 30 days after surgery. In order to develop strategies to improve patient care, an accurate and objective estimate of preoperative risk will be helpful in aiding patients and clinicians in the assessment of the risks of surgical and conservative treatment.

The objective of this study is to simultaneously model operative mortality and long-term survival using data from British Columbia's Cardiac Registry Database. We develop a semi-parametric regression analysis of (i) operative mortality, or death within 30 days, and (ii) long-term survival, or survival after 30 days, given survival to the first 30 days, using current cardiac registry data. We address here specifically the similarity of prognostic factors for operative mortality and long-term survival using a Cox model for long-term survival. We will show that the analysis can be viewed as one arising from multi-state modelling of the outcomes considered.

In the rest of this chapter, background information will be provided on the dataset used in the analysis, followed by a preliminary examination of the effects of covariates and the design of this study.

1.1 British Columbia's Cardiac Registry Database

British Columbia's Cardiac Registry, headed by Dr. Mike Kiely at St. Paul's Hospital, is one of the most comprehensive cardiovascular databases in Canada, containing over 10,000 patient records. It was created in 1989 by the provincial Ministry of Health in response to the report of long waiting times for cardiac surgery. One of the panel's recommendations was the establishment of a registry of cardiac patients to acquire detailed information. In fall 1990, the development of the registry commenced. Data collection started in 1991. The database captures prognostic information on all open heart surgeries performed in the province. Currently, open heart surgery is performed at four hospitals in the Lower Mainland and Victoria. One of the strengths of this database is that it is a population-based registry that captures every cardiac surgery procedure in British Columbia, as opposed to the usual single-hospital, selected-patient databases. The data is derived primarily from the Operative Report form, which is completed by the surgeon immediately after the surgery. The format of the form was revised in 1994 to capture additional information than previously required.

Prognostic information on the form is grouped into a number of different categories. These categories include: demographic information, information on previous cardiac surgery, diagnosis information, co-morbidity or diseases. Based on discussions with clinicians and previous work published on predicting operative mortality, a list of potential prognostic factors was developed. Data on most of these were recorded on the Operative Report form; however for some variables this data was obtained from other sources. Tables 1.1 - 1.3 (reproduced from Ghahramani 1999) list the potential prognostic factors and their definitions.

1.2 Coronary Artery Bypass Data

The vessels that bring blood to the heart are called the coronary arteries. They are somewhat similar to narrow tubes. A fatty substance called plaque can build up in these arteries and make them narrow, so less blood gets to the heart. This is called

Factor	Definition
Demographic variables	Age, gender, place of residence(urban/rural), hospital, year of surgery, average household income
Ejection fraction	Fraction of blood heart pumps out of ventrical when it beats
Urgency of surgery	Can be elective, urgent, or emergency
Left ventricular end diastolic pressure > 15 mg Hg	Indicator for left ventricular end diastolic pressure > 15 mg Hg
Re-operation	Indicator for previous cardiac operation
Unstable angina/recent MI < 6 weeks	indicator for unstable angina/recent MI < 6 weeks Angina is chest pain; MI is myocardial infarction or heart attack
MI > 6 weeks	Indicator for MI > 6 weeks; death of heart muscle due to an occlusion of the coronary artery
Chronic obstructive pulmonary disease	Indicator for chronic obstructive pulmonary disease such as emphysema & chronic bronchitis
Diabetes	Indicator for diabetes
ASA within 5 days	Indicator for aspirin administered within 5 days

Table 1.1: Prognostic factors

Factor	Definition
Pre-op symptomatic arrhythmia	Indicator for pre-op symptomatic arrhythmia; anything other than the normal heart rhythm which is called the sinus rhythm
Pre-op intra-aortic balloon pump	Indicator for pre-op intra-aortic balloon pump; a measure of severity of disease. A balloon is placed in aorta to decrease resistance which the heart must pump against when patient is having refractory angina
Pre-op iv nitroglycerine	Indicator for pre-op iv nitroglycerine; if patient is having angina, intravenous nitroglycerine is administered to relieve the pain and open up the arteries
Pre-op diuretic	Indicator for pre-op diuretic; used if there is excess body fluid and is also a marker for congestive heart failure
Pre-op iv Heparin	Indicator for pre-op iv Heparin; when patient experiences unstable angina and there is partial clotting of the arteries, heparin is administered to open the clot
Pre-op ventilation/intubation	Indicator for pre-op ventilation or intubation; ventilation is needed as the patient cannot breathe on his own due to congestive heart failure or shock, for example
Pre-op coumadin	Indicator for pre-op coumadin; a patient on coumadin has a high risk of stroke

Table 1.2: Prognostic factors

Factor	Definition
Pre-op steroids	Indicator for pre-op steroids
Pre-op thyroid replacement	Indicator for pre-op thyroid replacement
Other endocrine disease	Indicator for other endocrine disease; Example: adrenal disease
Hypertension	Indicator for very high blood pressure
Peripheral vascular disease	Indicator for peripheral vascular disease; hardening of arteries in body other than heart
Cerebrovascular disease	Indicator for cerebrovascular disease; indicator for hardening of arteries in brain
Dialysis/elevated creatinine	Indicator for dialysis/elevated creatinine; measures how well kidneys filter the blood
Congestive heart failure	Indicator for congestive heart failure which occurs when blood backs up to lungs since the heart is unable to pump blood to the rest of body
Pulmonary hypertension	Indicator for pulmonary hypertension; another measure of high blood pressure
Number of diseased vessels	Can be one of main left stenosis, more than 3 diseased vessels, or, 1-2 diseased vessels. Main left stenosis is the most important major artery coming out of the aorta which supplies arteries on the left side of the heart with blood
Malignant disease	Indicator for cancer

Table 1.3: Prognostic factors

Year	Number of CAB procedures	Operative Deaths
1991	1372	30
1992	1571	40
1993	1550	35
1994	1561	40
Total	6064	145

Table 1.4: The isolated coronary artery bypass procedure

coronary artery disease. The isolated coronary artery bypass (CAB) graft surgery is the most common treatment for serious coronary artery disease.

CAB surgery involves using blood vessels from other places in the body to 'bypass' the blockages in the coronary arteries. As a result blood flow is restored to the heart muscle.

In this study, we are interested in identifying risk factors for operative mortality and long-term survival after CAB surgery. From British Columbia's Cardiac Registry database, the dataset which we used consists of the first isolated CAB of all individuals receiving at least one CAB between 1991 to 1994. The total number of individuals in this dataset is 6064. The ages of these patients ranged from 12 to 92 with the median age being 65. The breakdown of the 6064 CAB's by year and number of operative deaths in a particular year is given in Table 1.4.

Figure 1.1 illustrates the Kaplan-Meier survivor function for the CAB patients; note this does not take any covariates into account. The estimated 30-day survival probability is 97.6%. The estimated 1-year and the estimated 3-year survival probabilities are 95.8% and 92.6% respectively. Notice the steep initial descent in the Kaplan-Meier curve which represents operative mortality.

For the CAB data, Tables 1.5 - 1.7 (reproduced from Ghahramani 1999) provide the frequency of each of the potential prognostic categorial variables identified previously. Only individuals with complete covariate vectors are considered for the development of a prognostic model. There were 5066 such cases, which represents about 84% of the CAB dataset.

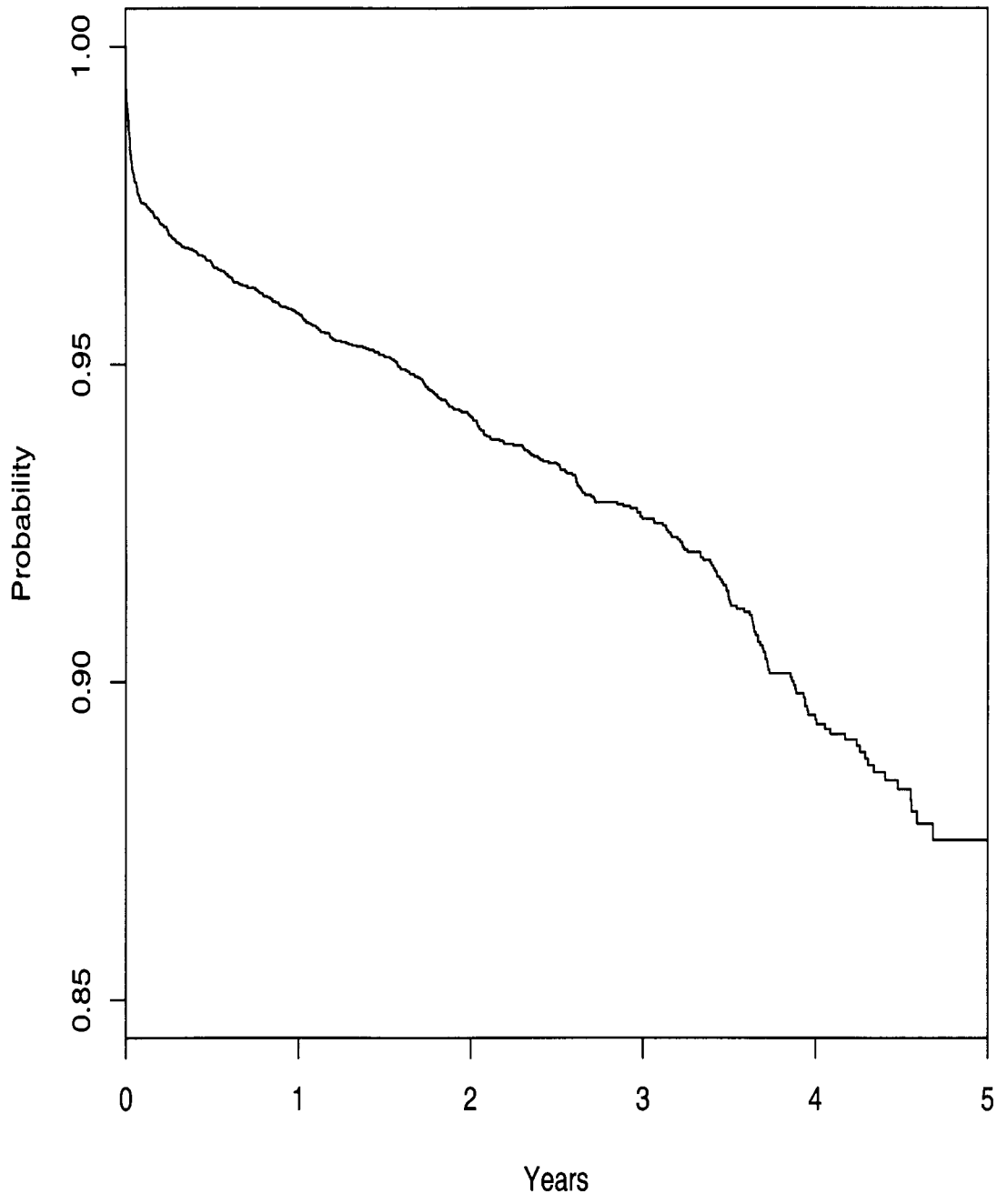

Figure 1.1: Estimated Survivor Function: 1991-1994

Factor	Label	Frequency
Gender	males	4821
	females	1243
Ejection fraction	<35%	512
	35-50%	1581
	>50%	3275
	missing	696
Urgency of surgery	elective	3766
	urgent	2034
	emergency	249
	missing	15
Left ventricular end diastolic pressure	yes	1874
> 15 mg Hg	no	3985
	missing	205
Re-operation	yes	409
	no	5450
	missing	205
Unstable angina/recent MI < 6 weeks	yes	1990
	no	3869
	missing	205
MI > 6 weeks	yes	1990
	no	3869
	missing	205
Chronic obstructive pulmonary disease	yes	567
	no	5292
	missing	205
Diabetes	yes	877
	no	4982
	missing	205
ASA within 5 days	yes	2071
	no	3788
	missing	205
Pre-op symptomatic arrhythmia	yes	272
	no	5587
	missing	205

Table 1.5: Potential risk factors and their frequencies

Factor	Label	Frequency
Pre-op intra-aortic balloon pump	yes	84
	no	5775
	missing	205
Pre-op iv nitroglycerine	yes	1036
	no	4823
	missing	205
Pre-op diuretic	yes	384
	no	5475
	missing	205
Pre-op iv heparin	yes	1854
	no	4005
	missing	205
Pre-op ventilation/intubation	yes	25
	no	5834
	missing	205
Pre-op coumadin	yes	73
	no	5786
	missing	205
Pre-op steroids	yes	73
	no	5786
	missing	205
Pre-op thyroid replacement	yes	135
	no	5724
	missing	205
Other endocrine disease	yes	31
	no	5828
	missing	205
Hypertension	yes	1499
	no	609
	missing	205
Peripheral vascular disease	yes	609
	no	5250
	missing	205
Cerebrovascular disease	yes	405
	no	5454
	missing	205

Table 1.6: Potential risk factors and their frequencies

Factor	Label	Frequency
Dialysis/elevated creatinine	yes	472
	no	5387
	missing	205
Congestive heart failure	yes	432
	no	5427
	missing	205
Pulmonary hypertension	yes	105
	no	5754
	missing	205
Number of diseased vessels	1-2 vessels	1087
	3 vessels	3715
	main left stenosis	1224
	missing	38
Average household income	<40k	908
	40-44k	1790
	45-49k	1862
	>50k	1388
	missing	116
Malignant disease	yes	51
	no	5808
	missing	205
Urban/rural residence	greater Van/Vic	2710
	other	3238
	missing	116
Year of surgery	1991	1372
	1992	1571
	1993	1550
	1994	1571
Institution of surgery	Vancouver	2244
	St. Paul	1786
	Royal Jubilee	1226
	Royal Columbian	808

Table 1.7: Potential risk factors and their frequencies

1.3 Preliminary Examination of Effects of Covariates

To determine which levels of a factor experience higher operative mortality or long-term survival, we examined each covariate's effect on operative mortality and long-term survival without adjusting for the other risk factors in a preliminary analysis. For each level of a factor, the percentage who experienced operative mortality for the group of patients belonging to that level was determined. Also, estimates of 1-year and 3-year survival probabilities were obtained without setting survival to the first 30 days as a condition. Tables A1 - A3 in the Appendix summarize these results.

Based on this preliminary analysis, the list of factors which seem to be singly important for operative mortality are age, gender, ejection fraction, urgency of surgery, re-operation, chronic obstructive pulmonary disease, diabetes, pre-operative symptomatic arrhythmia, pre-operative intra-aortic balloon pump, pre-operative nitroglycerine, preoperative diuretic, pre-operative ventilation or intubation, the need for dialysis or elevated creatinine levels, congestive heart failure and pulmonary hypertension.

In this thesis, we analyze the data set simultaneously considering operative mortality and long-term survival.

1.4 Plan of the Project

The plan of the project is as follows.

In Chapter 2 we develop a semi-parametric proportional hazards model for simultaneously fitting long-term survival and operative mortality. Inference using maximum likelihood theory is discussed.

In Chapter 3 the model is fitted to the British Columbia cardiac registry data and the proportional hazards assumption is examined.

Chapter 4 discusses an alternate approach viewing the model developed as a multi-state model. Summary comments are provided.

Chapter 2

Nonparametric Simultaneous Modelling

2.1 Introduction and Model Assumptions

Traditional survival analysis involves fitting a model to a single response, lifetime. Its application here has the assumption that both operative mortality (or short-term survival) and long-term survival (defined as survival after 30 days) are affected by the same set of covariates. Although this may be reasonable, we wish to establish a more general model, which will view these two outcomes separately, and allow predictions for them.

Natural models for these two outcomes are a logistic regression model for operative mortality and a proportional hazards (PH) model or a parametric regression model for long-term survival.

Ghahramani et. al. (1999) use the Weibull regression model for analyzing long-term survival. Although the Weibull is a flexible model, and seemed to give a fair fit in that analysis, a more robust approach is applied here using the Cox PH model. This provides a semi-parametric analysis where covariates enter the survivor function in a specified parametric way, but the shape of the survivor function is left unspecified and estimated non-parametrically.

Under the proportional hazards assumption the hazard function of T, given lifetime

covariates x, is of the form $h(t|\mathbf{x}) = h_0(t)g(\mathbf{x}, \beta)$, where β is a vector of unknown parameters and $g(.)$ is fully specified. Following Cox (1972), we will focus on the particular model that has

$$h(t|\mathbf{x}) = h_0(t)\exp(\mathbf{x}\beta) \tag{2.1}$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ is a vector of regression coefficients, $h_0(t)$ is a hazard function termed the baseline hazard function and corresponds to the hazard for an individual with $\mathbf{x} = \mathbf{0}$. The model in (2.1) is called the Cox PH model.

The advantage of using the Cox PH model is that it is essentially distribution-free: a partial likelihood function for β is constructed which does not depend upon the underlying lifetime distribution or, equivalently, on the baseline hazard $h_0(t)$. Even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained. This is what is implied when the Cox PH model is described as 'robust'. If the Weibull is in fact appropriate, the results from using the Cox PH model will closely approximate the results for the Weibull.

When in doubt, however, about the suitability of a particular parametric model, the Cox PH model becomes an appealing alternative because of the reliability of results from fitting this model and their reasonably high efficiency. For these reasons the Cox PH analysis is extremely popular, and we adopt it herein for the analysis of long-term survival.

For $i = 1, \ldots, n$ let

$$z_i = \begin{cases} 1 & \text{if the } i\text{th individual died within the first 30 days after surgery} \\ 0 & \text{otherwise} \end{cases}$$

Let $p_i = P(z_i = 1)$ be the probability that the ith individual dies within the first 30 days after surgery, $\mathbf{p} = (p_1, \cdots, p_n)$. Furthermore, let T_i denote the ith lifetime, L_i denote the ith censoring time for the individual who has survived beyond 30 days after surgery and $t_i = \min\{T_i, L_i\}$. Here, lifetime is defined as the interval between date of surgery and death for our application. Then let

$$Y_i = \begin{cases} t_i - 30 & \text{if } z_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

Fitting two separate models, a logistic regression model to the binary response, z_i, and a Cox PH regression model to the lifetimes y_i conditional on $z_i = 0$ gives

$$\begin{aligned} \text{logit}(\mathbf{p}) &= \mathbf{G}\gamma \\ h(y|\mathbf{x}, z = 0) &= h_0(y)\exp(\mathbf{X}^*\beta) \end{aligned} \tag{2.2}$$

for covariate design matrices $\mathbf{G}$ and $\mathbf{X}^*$.

We also wish to consider simultaneous estimation of covariate effects on operative mortality and long-term survival. Where such simultaneous estimation is appropriate, information will be pooled from the model predicting operative mortality with that which predicts long-term survival to estimate the corresponding effects. For example, if the same covariates affect $\text{logit}(\mathbf{p})$ and $h(y|\mathbf{x})$ so that as operative mortality decreases, long-term survival increases, we may have

$$\begin{aligned} \text{logit}(\mathbf{p}) &= \tau\mathbf{X}\beta \\ h(y|\mathbf{x}, z = 0) &= h_0(y)\exp(\mathbf{X}^*\beta) \end{aligned} \tag{2.3}$$

for covariate matrices $\mathbf{X}, \mathbf{X}^*$, and unknown real-valued parameter τ. Note that the covariate matrices $\mathbf{X}, \mathbf{X}^*$ contain the same set of covariates but differ in their number of rows since $\mathbf{X}$ contains data from all cases, while $\mathbf{X}^*$ contains information only from those cases who survived the first 30 days. When $\tau > 0$, the risk of operative mortality increases as long-term survival decreases, and as $\tau \to -\infty$, the risk of operatively mortality diminishes. As $\tau \to \infty$, the risk of operatively mortality becomes certain.

The parameterization in (2.3) however, assumes that the same set of covariates influence both operative mortality and long-term survival. To provide a bit more flexibility to accommodate those covariates whose effects on the two outcomes may differ, model (2.3) is reformulated as

$$\begin{aligned} \text{logit}(\mathbf{p}) &= \tau\mathbf{X}\beta + \mathbf{Z}\gamma \\ h(y|\mathbf{x}, z = 0) &= h_0(y)\exp(\mathbf{X}^*\beta) \end{aligned} \tag{2.4}$$

where γ contains the parameters corresponding to those covariates whose effects on operative mortality and long-term survival differ. In order to avoid over-parameterizing,

while $\mathbf{Z}$ contains a general constant term represented by a column of ones, $\mathbf{X}$ and $\mathbf{X}^*$ do not. What is envisioned is that X and Z should contain different sets of covariates, or, perhaps, just a few in common. The focus of fitting the simultaneous model (2.4) is to explore commonalities which lead to simple structures for covariate effects.

2.2 Likelihood Development and Maximum Likelihood Estimation

The likelihood functions under the three different model formulations (2.2), (2.3) and (2.4) are very similar and only the likelihood function for the model (2.4) will be presented. Suppose that a sample of n individuals yields k distinct observed lifetimes and $n-k$ censoring times. The k observed lifetimes will be denoted by $t_{(1)} < \cdots < t_{(k)}$, and $R_i = R(t_{(i)})$ will be used to represent the risk set at time $t_{(i)}$, that is, the set of individuals alive and uncensored just prior to $t_{(i)}$. Let $\mathbf{z} = (z_1, \cdots, z_n)$ be the vector of indicators for operative mortality. The likelihood $L(\gamma, \beta, \tau)$ is given by

$$L = \prod_{i=1}^{n} f(z_i; \gamma, \beta, \tau) \prod_{i=1}^{k} \left(e^{\mathbf{x}^*_{(i)}\beta} / \sum_{l \in R_i} e^{\mathbf{x}^*_l \beta} \right) \tag{2.5}$$

where $f(.)$ is the probability function of z_i, $\mathbf{x}_{(i)}$ is the regression vector associated with the individual observed to die at $t_{(i)}$. Notice the second term in the likelihood (2.5) is the partial likelihood function of long-term survival suggested by Cox (1995). If there are relatively few ties, a modification of the likelihood (2.5) suggested by Breslow (1974) may be adopted: replace L in (2.5) with

$$L = \prod_{i=1}^{n} f(z_i; \gamma, \beta, \tau) \prod_{i=1}^{k} e^{\mathbf{S}_{(i)}\beta} / \left(\sum_{l \in R_i} e^{\mathbf{x}^*_l \beta} \right)^{d_i} \tag{2.6}$$

where d_i is the number of lifetimes which equal $t_{(i)}$ and $\mathbf{S}_{(i)}$ is the sum of the regression vectors $\mathbf{x}^*$ for these d_i individuals. That is, if D_i represents the set of individuals who die at $t_{(i)}$, then $d_i = |D_i|$ and $\mathbf{S_i} = \sum_{l \in D_i} \mathbf{x}^*_\mathbf{l}$. When there are no ties, all $d_i = 1$ and (2.6) reduces to (2.5).

Let $\mathbf{X_i}$, $\mathbf{X_i^*}$, $\mathbf{Z_i}$ denote the ith row of $\mathbf{X}$, $\mathbf{X^*}$ and $\mathbf{Z}$, respectively. Then, since

$$f(z_i) = p_i^{z_i}(1-p_i)^{(1-z_i)} \tag{2.7}$$

the log of the likelihood becomes:

$$\begin{aligned} \log L &= \sum_{i=1}^{n}\{z_i[\tau\mathbf{X}_i\beta + \mathbf{Z}_i\gamma] - \log(1+\exp(\tau\mathbf{X}_i\beta + \mathbf{Z}_i\gamma))\} \\ &\quad + \sum_{i=1}^{k}\mathbf{S_i}\beta - \sum_{i=1}^{k} d_i \log\left(\sum_{l\in R_i} e^{\mathbf{x^*}_l\beta}\right) \end{aligned} \tag{2.8}$$

The maximum likelihood estimates of β, γ, τ are obtained from (2.8) using a Newton-Raphson algorithm. First and second partial derivative are required for the algorithm.

Let

$$a_i = \exp(\tau\mathbf{X}_i\beta + \mathbf{Z}_i\gamma)$$

The components of the score vector

$$\mathbf{U} = \left(\frac{\partial \log L}{\partial\gamma_1}, \ldots, \frac{\partial \log L}{\partial\gamma_q}, \frac{\partial \log L}{\partial\beta_1}, \ldots, \frac{\partial \log L}{\partial\beta_p}, \frac{\partial \log L}{\partial\tau}\right)$$

and observed Fisher information matrix $\mathbf{I_0} = (I_{rs})$ are the following:

$$
\begin{aligned}
\frac{\partial \log L}{\partial \gamma_r} &= \sum_{i=1}^{n}\left(z_i - \frac{a_i}{1+a_i}\right) Z_{ir} \\
\frac{\partial \log L}{\partial \beta_r} &= \tau \sum_{i=1}^{n}\left(z_i - \frac{a_i}{1+a_i}\right) X_{ir} + \sum_{i=1}^{k}\left(S_{ir} - d_i \sum_{l\in R_i} x_{lr} e^{\mathbf{x}_l\beta} / \sum_{l\in R_i} e^{\mathbf{x}_l\beta}\right) \\
\frac{\partial \log L}{\partial \tau} &= \sum_{i=1}^{n}\left(z_i - \frac{a_i}{1+a_i}\right) \mathbf{X}_i\beta \\
\frac{\partial^2 \log L}{\partial \gamma_r \gamma_s} &= -\sum_{i=1}^{n} \frac{a_i}{(1+a_i)^2} Z_{ir} Z_{is} \\
\frac{\partial^2 \log L}{\partial \gamma_r \beta_s} &= -\tau \sum_{i=1}^{n} \frac{a_i}{(1+a_i)^2} Z_{ir} X_{is} \\
\frac{\partial^2 \log L}{\partial \gamma_r \tau} &= -\sum_{i=1}^{n} (\mathbf{X}_i\beta) \frac{a_i}{(1+a_i)^2} Z_{ir} \\
\frac{\partial^2 \log L}{\partial \beta_r \beta_s} &= -\tau^2 \sum_{i=1}^{n} \frac{a_i}{(1+a_i)^2} X_{ir} X_{is} \\
&\quad -\sum_{i=1}^{k} d_i \left(\frac{\left(\sum_{l\in R_i} e^{\mathbf{x}_l\beta}\right)\left(\sum_{l\in R_i} x_{lr}x_{ls}e^{\mathbf{x}_l\beta}\right) - \left(\sum_{l\in R_i} x_{lr}e^{\mathbf{x}_l\beta}\right)\left(\sum_{l\in R_i} x_{ls}e^{\mathbf{x}_l\beta}\right)}{\left(\sum_{l\in R_i} e^{\mathbf{x}_l\beta}\right)^2} \right) \\
\frac{\partial^2 \log L}{\partial \beta_r \tau} &= -\sum_{i=1}^{n}\left(z_i - \frac{a_i}{1+a_i}\right) X_{ir} - \tau \sum_{i=1}^{n} \mathbf{X}_i\beta \frac{a_i}{(1+a_i)^2} X_{ir} \\
\frac{\partial^2 \log L}{\partial \tau^2} &= -\sum_{i=1}^{n} (\mathbf{X}_i\beta)^2 \frac{a_i}{(1+a_i)^2}
\end{aligned}
$$

The partitioned form of the observed information matrix $\mathbf{I}_0$ is

$$
-\begin{bmatrix}
\frac{\partial^2 \log L}{\partial \gamma \partial \gamma'} & \frac{\partial^2 \log L}{\partial \gamma \partial \beta'} & \frac{\partial^2 \log L}{\partial \gamma \partial \tau} \\
\frac{\partial^2 \log L}{\partial \beta \partial \gamma'} & \frac{\partial^2 \log L}{\partial \beta \partial \beta'} & \frac{\partial^2 \log L}{\partial \beta \partial \tau} \\
\frac{\partial^2 \log L}{\partial \tau \partial \gamma'} & \frac{\partial^2 \log L}{\partial \tau \partial \beta'} & \frac{\partial^2 \log L}{\partial \tau^2}
\end{bmatrix}
$$

The maximum likelihood equations $\frac{\partial \log L}{\partial \gamma_r} = 0$, $\frac{\partial \log L}{\partial \beta_r} = 0$ and $\frac{\partial \log L}{\partial \tau} = 0$ are then solved by using the Newton-Raphson updating algorithm. Let ξ be the full parameter vector

$\xi = (\gamma_1, \ldots, \gamma_q, \beta_1, \ldots, \beta_p, \tau)'$, and $\xi^{(n)}$ indicate the current value of the estimate of ξ. Then $\xi^{(n)}$ is updated to $\xi^{(n+1)}$ using

$$\xi^{(n+1)} = \xi^{(n)} + [\mathbf{I}_0(\xi^{(n)})]^{-1}\mathbf{U}(\xi^{(n)}) \quad n = 1, 2, 3, \ldots$$

This updating continues to convergence.

We use the standard large sample likelihood ratio test to test the hypothesis H_0 : $\beta_i = 0$ versus the alternative $H_0 : \beta_i \neq 0$ or $H_1 : \gamma_i = 0$ versus $H_1 : \gamma_i \neq 0$. In order to evaluate whether is reasonable to assume that $\tau = 1$, a hypothesis of interest, a profile plot for τ should be examined.

We are also interested in the effects of the covariates via odds ratios and relative risks for clinical purposes. Odd ratios of predictors of operative mortality are found by exponentiating the parameter estimates in the logistic regression model. For the Cox PH regression model, the estimated relative risk of the jth risk factor relative to the baseline is $\exp(\hat{\beta}_j)$ as described below.

The hazard function of Y given covariate vector $\mathbf{x}$ is given by

$$h(y|\mathbf{x}) = h_0(t)\exp(\mathbf{x}\beta)$$

We can write the relative risk (RR) of $\mathbf{x}_1^*$ versus $\mathbf{x}_2^*$ as $h(y|\mathbf{x}_1^*)$ divided by $h(y|\mathbf{x}_2^*)$ or

$$\begin{aligned} RR &= \frac{h_0(t)\exp(\mathbf{x}_1^*\beta)}{h_0(t)\exp(\mathbf{x}_2^*\beta)} \\ &= \frac{h_0(t)\exp(\sum_{i=1}^p \beta_i x_{1i}^*)}{h_0(t)\exp(\sum_{i=1}^p \beta_i x_{2i}^*)} \\ &= \exp\left(\sum_{i=1}^p \beta_i(x_{1i}^* - x_{2i}^*)\right) \end{aligned}$$

Given all covariates are fixed except x_j, the relative risk becomes

$$RR = \exp\left(\beta_j(x_{1j}^* - x_{2j}^*)\right)$$

For a categorical variable, x_j, when $x_j = 1$ versus when $x_j = 0$, or for a continuous variable when x_j is increased by unity, the relative risk is given by

$$RR = \exp\left(\beta_j\right).$$

Chapter 3

Application to the B.C. Cardiac Registries Data

3.1 Model Fitting

In order to make the model parsimonious, due to the large list of potential risk factors, a number of variable selection procedures were implemented into the model fitting process. These included forward selection, backward elimination, simultaneously dropping several variables, and stepwise procedures. From the non-simultaneous model (2.2), Table 3.1 lists the covariates which were identified by any of the model fitting procedures as important in predicting operative mortality and long-term survival, defined as 'Significant predictors'.

Examination of the estimates for the year of surgery effect indicated fewer cases of operative mortality and longer long-term survival in 1993 than those corresponding to 1991, 1992 and 1994. Since there was no clinical evidence to support inclusion of this of this variable for future predictions, we decided to exclude year of surgery as a covariate.

For the simultaneous model (2.4), the estimate of τ is 1.1 with a standard error of 0.14. The ninety-five per cent confidence interval for τ based on the large sample normal approximation to the distribution of $\hat{\tau}$ is (0.8,1.4). Since there is no evidence against the hypothesis that $\tau = 1$, we refit the model with τ set at unity.

Operative mortality	Long-term survival
Age	Age
Gender	Ejection fraction
Ejection fraction	Urgency of surgery
Urgency of surgery	Re-operation
Re-operation	Diabetes
Pre-op iv nitroglycerine	Pre-op diuretic use
Pre-op diuretic use	Peripheral vascular disease
Pre-op ventilation/intubation	Dialysis/elevated creatinine
Pulmonary hypertension	Congestive heart failure
No. of diseased vessels	No. of diseased vessels
Urban/rural residence	Year of surgery
Congestive heart disease	
Peripheral vascular disease	

Table 3.1: Significant predictors

Interaction terms with sex and age were also considered as well as a quadratic term with age. None of these was significant.

The parameter estimates corresponding to the fitted model with τ equal to one are reported in Table 3.2. The estimated odds ratio for operative mortality corresponding to the fitted model are reported in Table 3.3. The estimated relative risks for long-term mortality corresponding to the fitted model are reported in Table 3.4.

Although all the variables listed in Table 3.2 contribute to an understanding of risks in this analysis, note that gender, urgency of surgery, pre-operative ventilation and rural residence are significant predictors of operative mortality, but not long-term survival while peripheral vascular disease is a significant predictor of long-term survival but not operative mortality. Interpretation of covariates that affect both operative mortality and long-term survival similarly is straightforward. For example, having an ejection fraction $\leq 35\%$ increases the risks corresponding to both operative mortality and long-term survival.

Factors which pose extremely high risk for operative mortality (OR > 2) are emergency surgery, pre-operative ventilation/intubation and ejection fraction of less

Factor	Label	$\hat{\gamma}$	SE($\hat{\gamma}$)
Female gender		0.51	0.21
Urgency of surgery	elective		
	urgent	0.41	0.21
	emergency	1.18	0.39
Pre-op ventilation/intubation		1.93	0.63
Peripheral vascular disease		-0.88	0.32
Rural residence (baseline in urban residence)		0.58	0.21
Factor	Label	$\hat{\beta}$	SE($\hat{\beta}$)
Age		0.06	0.01
Ejection fraction	>50%		
	35-50%	0.28	0.12
	<35%	0.94	0.15
Re-operation		0.59	0.15
Diabetes		0.26	0.13
Pre-op diuretic		0.42	0.15
Peripheral vascular disease		0.61	0.15
Cerebrovascular disease		0.32	0.16
Dialysis/elevated creatinine		0.32	0.14
Congestive heart failure		0.55	0.14
Number of diseased vessels	1-2		
	≥ 3	0.31	0.18
	main left stenosis	0.58	0.19

Table 3.2: Parameter estimates and their standard errors for the simultaneous fit with $\tau = 1$.

Factor	Label	OR	95% CI
Female gender		1.67	(1.11,2.51)
Urgency of surgery	elective	1.00	
	urgent	1.51	(1.01, 2.27)
	emergency	3.26	(1.53, 6.96)
Pre-op ventilation/intubation		6.92	(2.03, 23.60)
Peripheral vascular disease		0.77	(0.39, 1.52)
Rural residence (baseline in urban residence)		1.79	(1.19, 2.69)
Age		1.06	(1.05, 1.08)
Ejection fraction	>50%	1.00	
	35-50%	1.33	(1.05, 1.68)
	<35%	2.57	(1.93, 3.42)
Re-operation		1.80	(1.33, 2.43)
Diabetes		1.29	(1.00, 1.66)
Pre-op diuretic		1.52	(1.14, 2.04)
Cerebrovascular disease		1.37	(1.01, 1.87)
Dialysis/elevated creatinine		1.38	(1.05, 1.81)
Congestive heart failure		1.73	(1.31, 2.30)
Number of diseased vessels	1-2	1.00	
	≥ 3	1.36	(0.97, 1.93)
	main left stenosis	1.78	(1.22, 2.60)

Table 3.3: Odd ratios for operative mortality and their confidence intervals for the simultaneous fit with $\tau = 1$.

Factor	Label	RR	95% CI
Age		1.06	(1.05, 1.08)
Ejection fraction	>50%	1.00	
	35-50%	1.33	(1.05, 1.68)
	<35%	2.57	(1.93, 3.42)
Re-operation		1.80	(1.33, 2.43)
Diabetes		1.29	(1.00, 1.66)
Pre-op diuretic		1.52	(1.14, 2.04)
Peripheral vascular disease		1.84	(0.38, 2.45)
Cerebrovascular disease		1.37	(1.01, 1.87)
Dialysis/elevated creatinine		1.38	(1.05, 1.81)
Congestive heart failure		1.73	(1.31, 2.30)
Number of diseased vessels	1-2	1.00	
	≥ 3	1.36	(0.97, 1.93)
	main left stenosis	1.78	(1.22, 2.60)

Table 3.4: Relative risks for long-term survival and their confidence intervals for the simultaneous fit with $\tau = 1$.

than 35%. The factor which poses extremely high risk for long-term survival (RR > 2) is ejection fraction of less than 35%. Factors which pose high risk ($1.5 \leq$ OR/RR ≤ 2) for both operative mortality and long-term survival are re-operation, pre-operative diuretic, congestive heart failure and main left stenosis. In addition, high risk factors for operative mortality ($1.5 \leq$ OR ≤ 2) are also female gender, urgent surgery and rural residence.

Comparing the results from this analysis with those from Ghahramani et. al. (1999) who used a fully parametric approach combining the logistic regression model for analyzing operative mortality, and the Weibull regression model for analyzing long-term survival, we note some striking similarities. First, the set of all the significant factors for long-term survival in this analysis is parallel to those from Ghahramani's analysis as shown in Table 3.5 and the estimated effects and their standard errors are very close, with the exception of the covariate *Pre-op ventilation/intubation.* In the analysis above, this variable is not significant for predicting long-term survival. Though the confidence interval for this effect was quite wide in Ghahramani's analysis,

Factor	Label	Chui's RR	Ghahramani's RR
Age		1.06	1.06
Ejection fraction	>50%	1.00	1.00
	35-50%	1.33	1.32
	<35%	2.57	2.56
Re-operation		1.80	1.79
Diabetes		1.29	1.29
Pre-op diuretic		1.52	1.53
Pre-op ventilation/intubation		not sig.	2.90
Peripheral vascular disease		1.84	1.83
Cerebrovascular disease		1.37	1.37
Dialysis/elevated creatinine		1.38	1.37
Congestive heart failure		1.73	1.71
Number of diseased vessels	1-2	1.00	1.00
	≥ 3	1.36	1.35
	main left stenosis	1.78	1.74

Table 3.5: Comparison of relative risks for long-term survival between Chui's results and Ghahramani's results.

(1.19, 7.05), and substantially wider than any of the other confidence intervals for relative risks for long-term survival in that analysis, nevertheless it was significant. The clinical significance of having this variable as a predictor of operative mortality is certainly clear as this is another indicator of the urgency of the surgery. However, having it as a predictor of long-term survival, given survival 30 days after surgery and subsequent discharge, is not as clear.

Considering the analysis of operative mortality reported in Table 3.6, again we identify a close parallel between the results of the two analyses. Differences occur for the covariates *Emergency surgery* and again for *Pre-op ventilation/intubation.* The estimated relative risk for Emergency surgery is lower in this analysis than in Ghahramani et. al.'s while the estimated relative risk for Pre-op ventilation/intubation is much higher. Note that these two variables are correlated and this may play a part in explaining the differences in the analyses.

Factor	Label	Chui's OR	Ghahramani's OR
Female gender		1.67	1.67
Urgency of surgery	elective	1.00	1.00
	urgent	1.51	1.53
	emergency	3.26	3.92
Pre-op ventilation/intubation		6.92	2.90
Peripheral vascular disease		0.77	0.77
Rural residence		1.79	1.77
Age		1.06	1.06
Ejection fraction	>50%	1.00	1.00
	35-50%	1.33	1.32
	<35%	2.57	2.56
Re-operation		1.80	1.79
Diabetes		1.29	1.29
Pre-op diuretic		1.52	1.53
Cerebrovascular disease		1.37	1.37
Dialysis/elevated creatinine		1.38	1.37
Congestive heart failure		1.73	1.71
Number of diseased vessels	1-2	1.00	1.00
	≥ 3	1.36	1.35
	main left stenosis	1.78	1.74

Table 3.6: Comparison of odd ratios for operative mortality between Chui's results and Ghahramani's results.

3.2 Checking the Proportional Hazards Assumption

The cumulative hazard function, $H(t|\mathbf{x})$, is related to the hazard function, $h(t|\mathbf{x})$, by

$$\begin{aligned} H(t|\mathbf{x}) &= \int_0^t h(u|\mathbf{x})\mathrm{d}u \\ &= \int_0^t h_0(u)\mathrm{d}u \exp(\mathbf{x}\beta) \\ &= H_0(t)\exp(\mathbf{x}\beta) \end{aligned}$$

and the survival function $S(t|\mathbf{x})$, is related to $H(t|\mathbf{x})$, by

$$\begin{aligned} S(t|\mathbf{x}) &= \exp(-H(t|\mathbf{x})) \\ &= \exp(-H_0(u)\exp(\mathbf{x}\beta)) \\ &= \exp(-H_0(u))^{\exp(\mathbf{x}\beta)} \\ &= S_0(t)^{\exp(\mathbf{x}\beta)} \end{aligned}$$

The PH assumption implies that $S(t|\mathbf{x}) = S_0(t)^{\exp(\mathbf{x}\beta)}$, i.e. the survival curves for different values of $\mathbf{x}$ are powers of one another. We can use this observation to perform rudimentary checks of the PH assumption through inspection of Kaplan-Meier survival curve estimates or Cox PH model survival curve estimates for various levels of a covariate, since the crossing of these survival curves indicates departure from the assumption. The PH assumption also implies that $H(t|\mathbf{x}) = H_0(t)\exp(\mathbf{x}\beta)$, i.e. that the cumulative hazard curves obey a proportionality assumption. Here again, crossing curves indicate violations of the PH assumption.

Since $H(t|\mathbf{x}) = -\log S(t|\mathbf{x})$, we can use the logarithmic transformation of the Kaplan-Meier estimate or Cox PH model estimate for this assessment. The PH assumption further implies that

$$\log H(t) = \log H_0(t) + \mathbf{x}\beta;$$

or we can rewrite the PH model as

$$\log[-\log S(t)] = \log[-\log S_0(t)] + \mathbf{x}\beta.$$

Therefore, under the PH assumption, plots of $\log[-\log \hat{S}_i(t)]$, or equivalently, plots of $\log \hat{H}_i(t)$, are roughly parallel for different values of $\mathbf{x}$.

Violations of the PH assumption may also be assessed by comparing survival estimates based on the Cox PH model with estimates computed independently, as for example, with the Kaplan-Meier estimates. Departures between the two provide evidence against the PH assumption. In addition, Schoenfeld residuals (Schoenfeld, 1980) may be plotted versus time. Briefly, for a dichotomous covariate x, these are either 1 - $\mathrm{E}(x_i|R_i)$ or 0 - $\mathrm{E}(x_i|R_i)$, where R_i denotes the risk set at the ith failure time and the expectation of the covariate value is computed under the Cox model. If the PH assumption holds, the residuals should be approximately in two horizontal bands, while time trends indicate a departure from PH.

Figure 3.1 displays Kaplan-Meier survivor functions corresponding to the factors found significant for the analysis of long-term survival, while Figure 3.2 shows plots of the log cumulative hazard based on the Kaplan-Meier survivor function. In general, there is no strong evidence of departure from proportional hazards seen in these plots, though some hints of departure may be seen in Figure 3.1 for age, ejection fraction and no.of diseased vessels, and in Figure 3.2 for ejection fraction, re-operation, diabetes, pre-op diuretic and no. of diseased vessels. The curves for diabetes in fact look coincident, reflecting the fact that this variable is marginally significant. The curves for no. of diseased vessels also reflect the lack of significance between the effects of ≥ 3 diseased vessels and main left stenosis (mls). The variables age, ejection fraction, re-operation, diabetes, pre-op diuretic and no. of diseased vessels are investigated further in Figures 3.3 - 3.8 which compare survivor function estimates based on the Cox model with the Kaplan-Meier estimates. The figures show no striking dis-similarities between the survivor function estimates.

Figures 3.9 and 3.10 show the scaled Schoenfeld residuals and the corresponding LOWESS-smoothed curves (and 95% confidence intervals) of all significant factors in the analysis of long-term survival. In general, the LOWESS-smoothed curves center around $\hat{\beta}$ in Table 3.2 as horizontal lines with the exceptions of diabetes and peripheral vascular disease. The LOWESS-smoothed curve for diabetes shows a general increasing trend, while the LOWESS-smoothed curve for peripheral vascular disease

appears to be a U-shaped. However, note the large confidence interval in the smoothed estimates especially in the regions where the trends are observed. In the significant test, the p-values of all the significant factors are greater than 0.5. In summary, there does not seem to be strong evidence of departures from PH evident in the analysis of long-term survival.

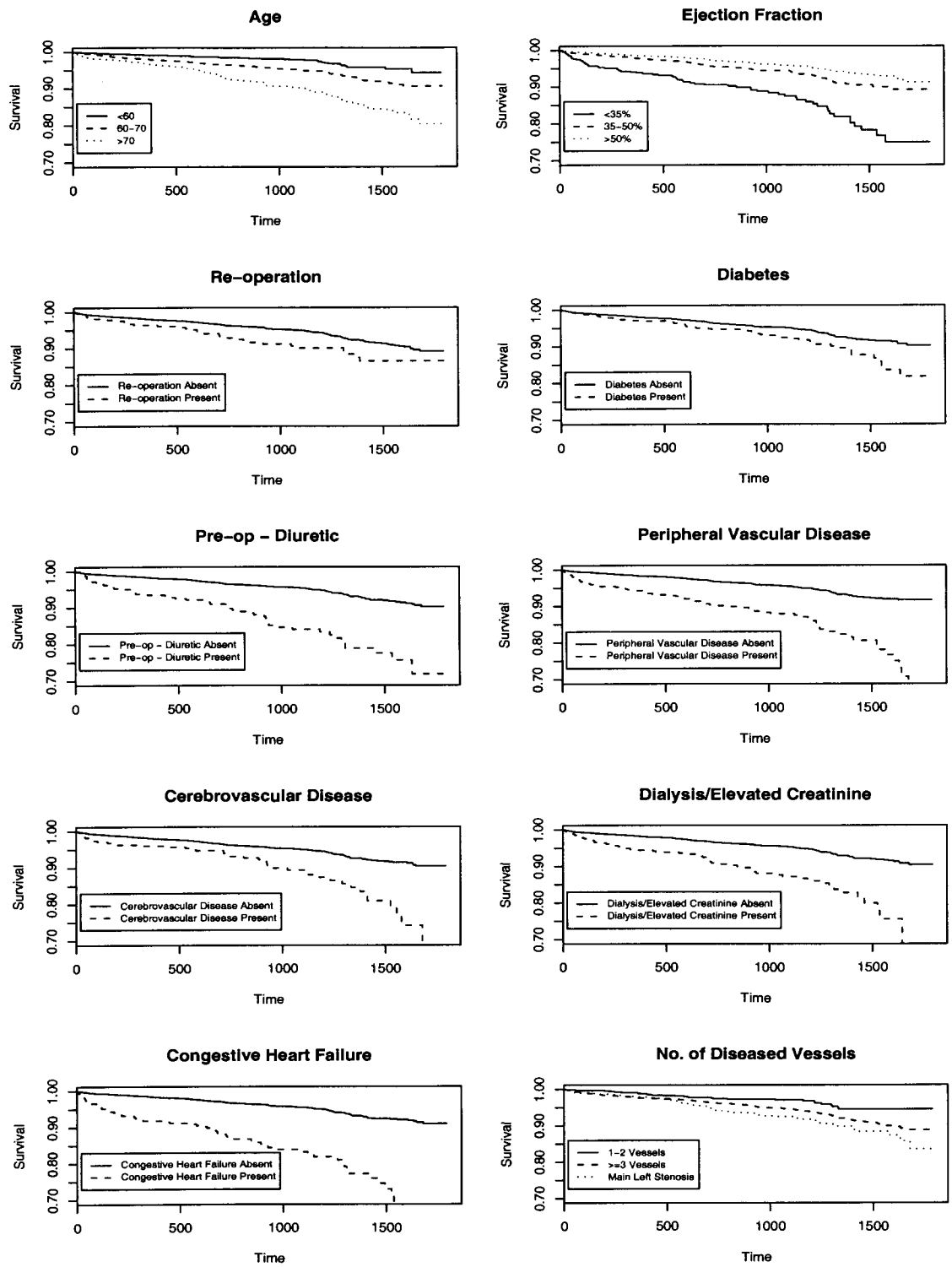

Figure 3.1: Kaplan-Meier survival curve estimates

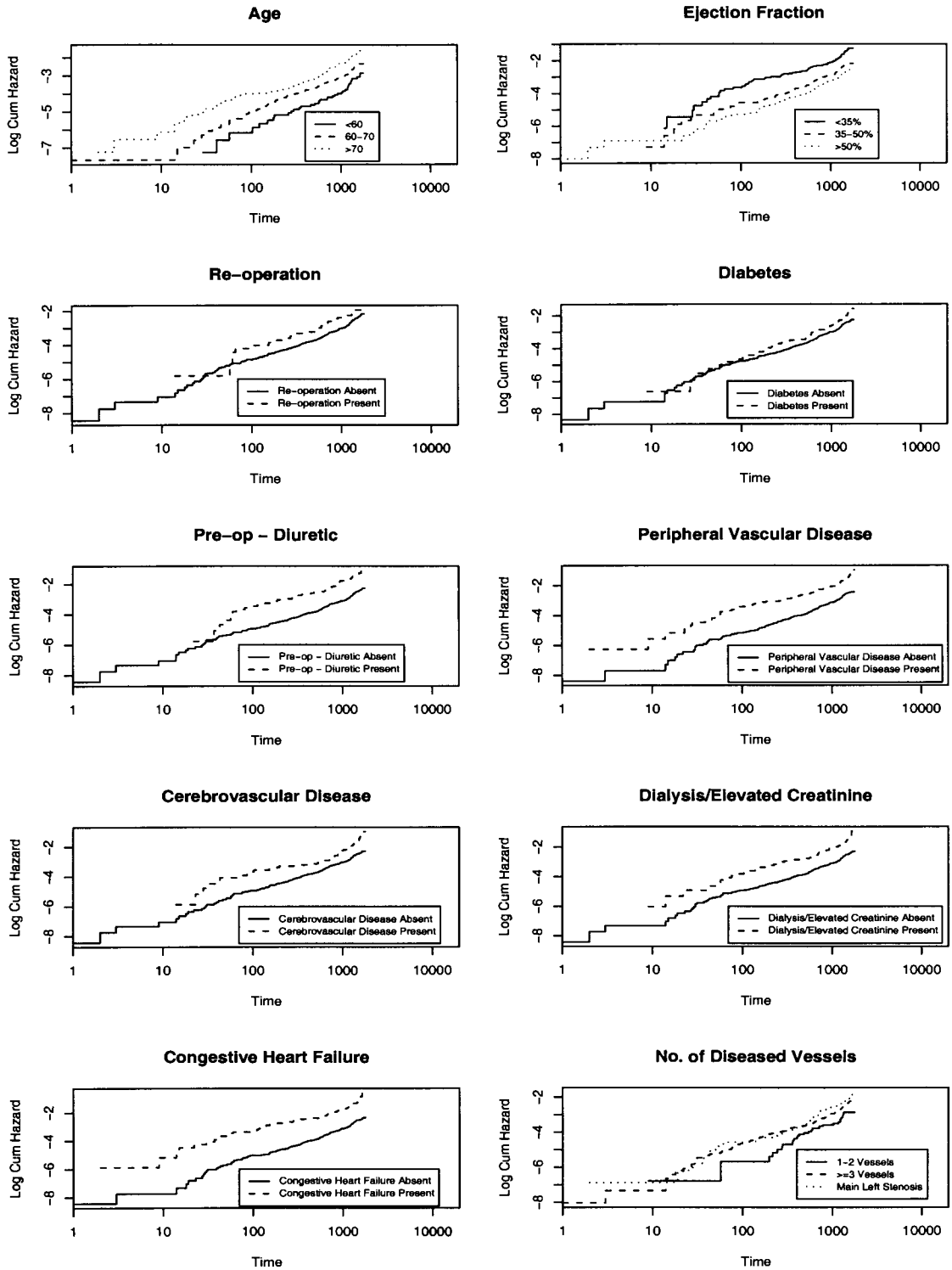

Figure 3.2: Log cumulative hazard functions for Kaplan-Meier estimate

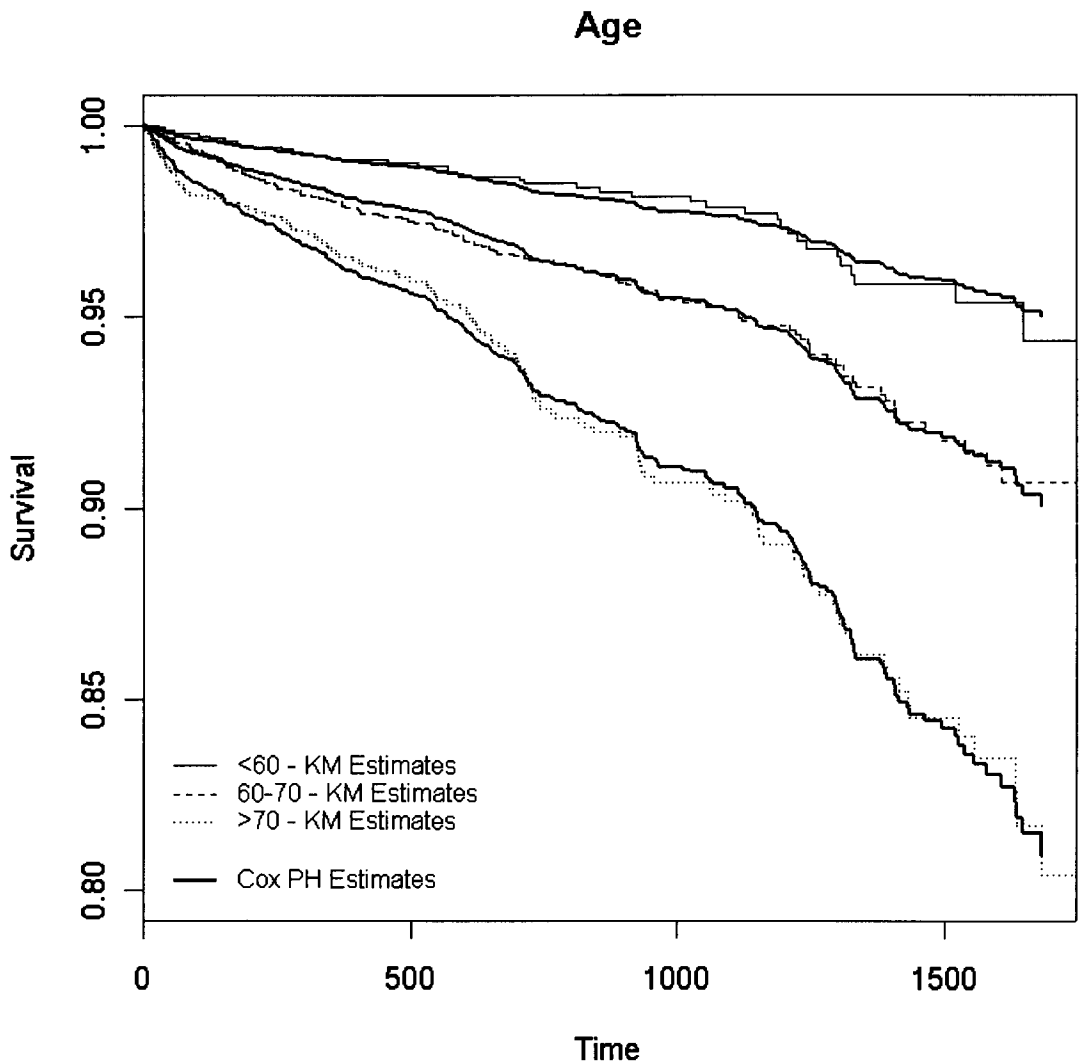

Figure 3.3: Kaplan-Meier vs Cox PH model survival curve estimates for the levels of age

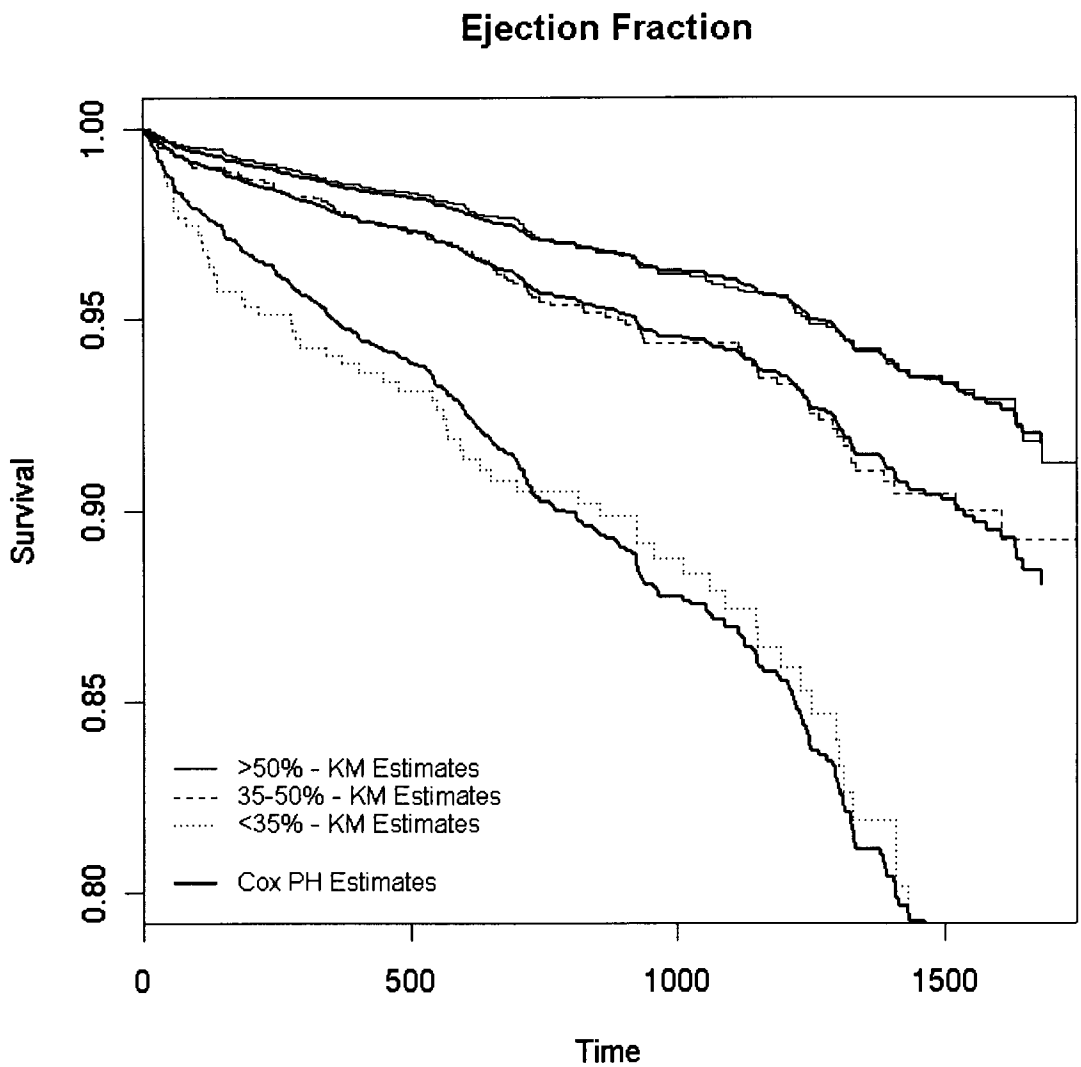

Figure 3.4: Kaplan-Meier vs Cox PH model survival curve estimates for the levels of ejection fraction

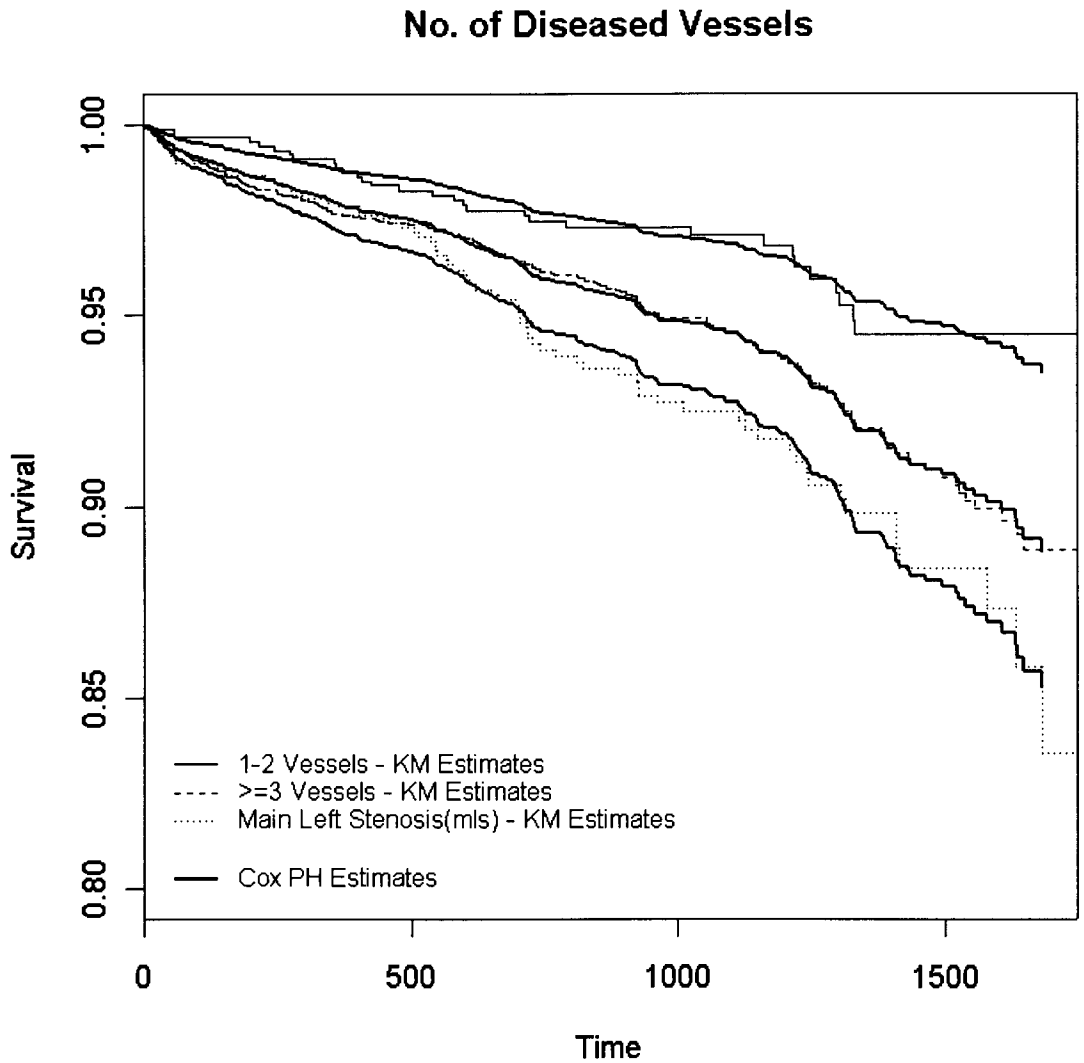

Figure 3.5: Kaplan-Meier vs Cox PH model survival curve estimates for the levels of no. of diseased vessels

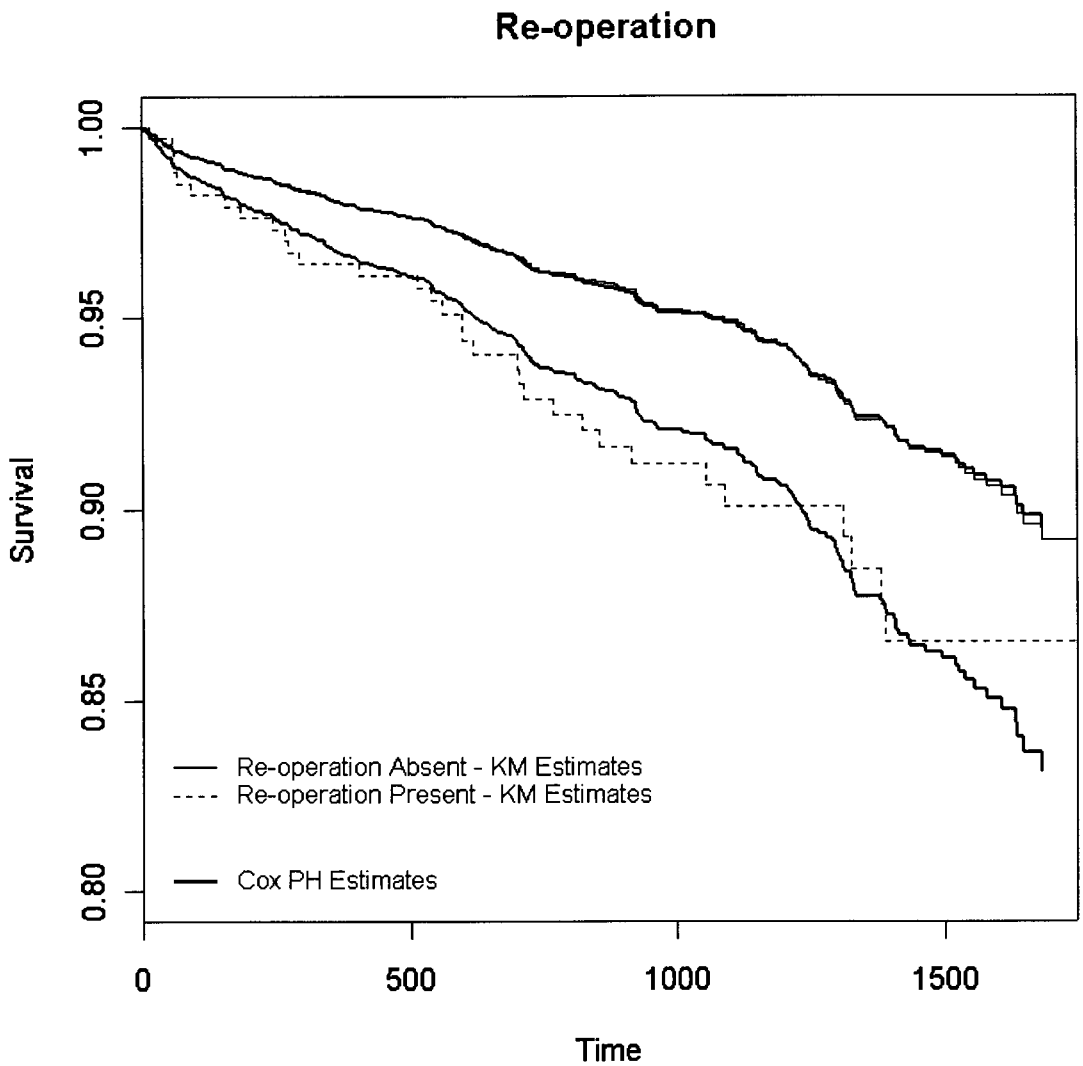

Figure 3.6: Kaplan-Meier vs Cox PH model survival curve estimates for re-operation present and absent

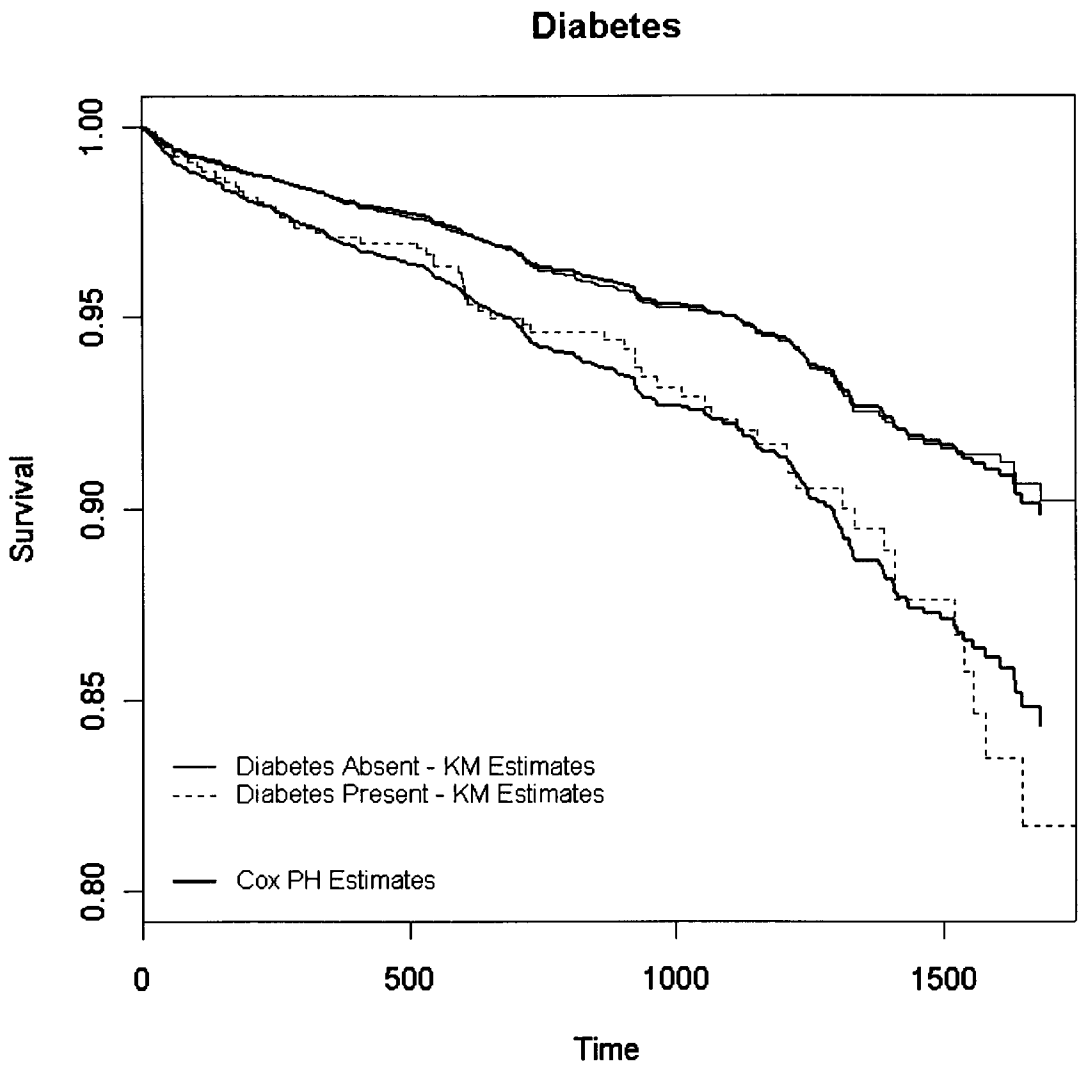

Figure 3.7: Kaplan-Meier vs Cox PH model survival curve estimates for diabetes present and absent

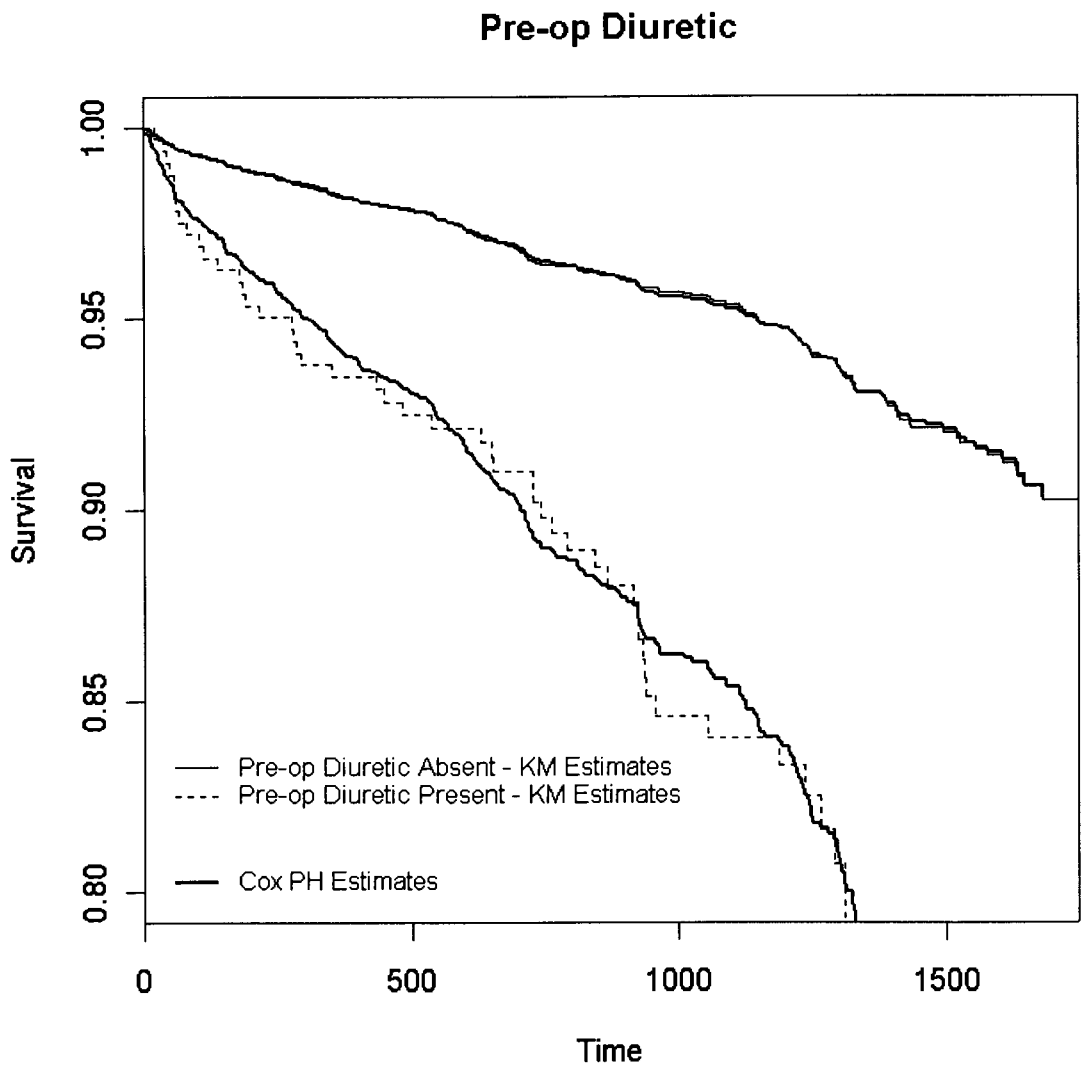

Figure 3.8: Kaplan-Meier vs Cox PH model survival curve estimates for pre-op diuretic present and absent

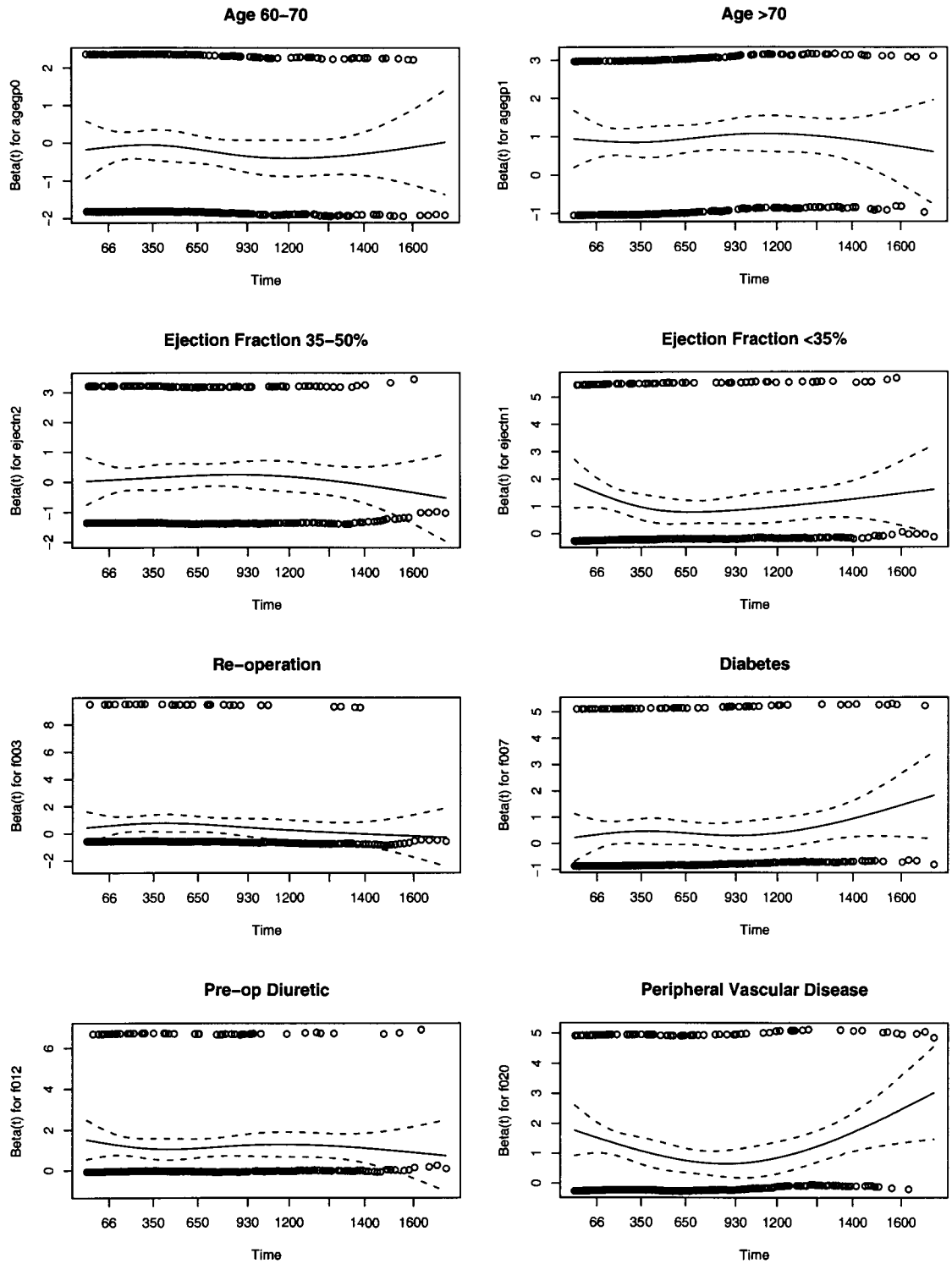

Figure 3.9: Schoenfeld residuals

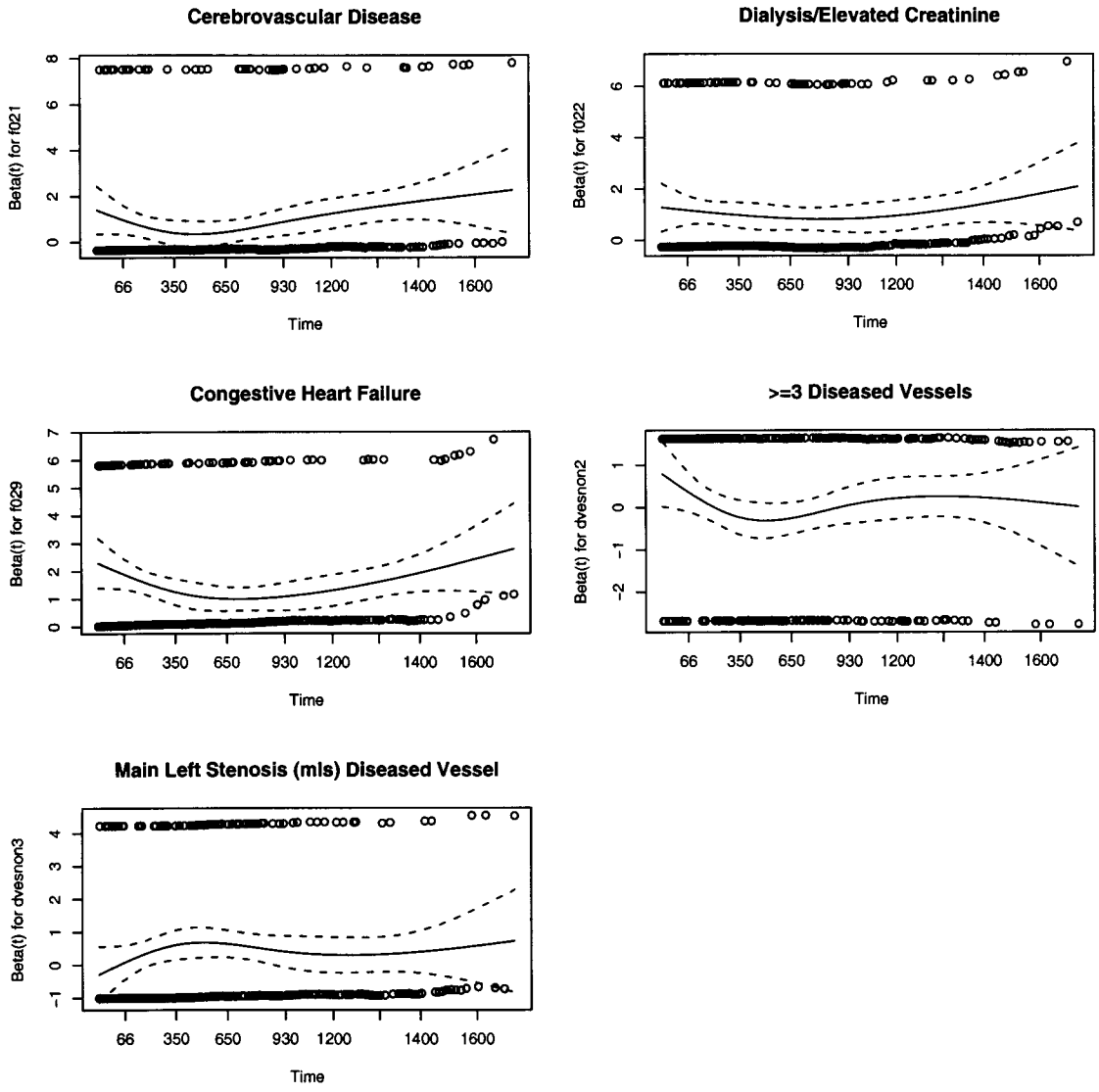

Figure 3.10: Schoenfeld residuals

Chapter 4

Connection with Multi-State Modelling

In situations where treatment is harsh, and survival is measured by time from treatment and analyzed using traditional methods such as the proportional hazards model, it is usual to observe steep initial descent of the estimated survivor curve followed by a less sharp decline. Sometimes the survivor curve flattens after such steep initial descent indicating that, for those individuals who survive treatment it is quite successful in curing disease. In these situations it is often of interest to quantify the effects of covariates on (i) surviving the treatment experience and (ii) long-term survival. In coronary artery bypass surgery, for example, the most common open heart surgical procedure, there is a high risk of operative mortality, defined as death within 30 days after surgery. However, for those individuals who survive surgery, the operation is typically quite successful with high promise of long-term survival.

Ghahramani et. al. (2001) discussed parametric methods of investigating joint effects of covariate for two outcomes. Lifetimes were modelled using the Weibull model and operative mortality was incorporated as a binary endpoint. Methods for jointly assessing the effects of covariates were employed. This project extends such methods by considering the use of a non-parametric modelling approach for long-term survival. It retains the ideas of Ghahramani et. al. for incorporating covariates but provides a more flexible semi-parametric method for investigating their effects.

The approach used can also be viewed as arising from multi-state modelling. Multi-state models have been useful in other situations where several endpoints are jointly studied and transitions between these investigated. For example, recently, Chevret, Leporrier and Chastang (2000) adopt a multi-state model for incorporating tumour response in randomized phase III cancer trials. Tumour response is indicated by tumour shrinkage and attainment of response is rationalized as a meaningful surrogate endpoint worth investigation. In the study of outcomes after cardiac surgery this is certainly also the case. Although, as mentioned earlier in the project, the two endpoints could be studied separately, using the multi-state modelling approach allows investigation of simple structures for covariate effects on both outcomes. Exploring commonalities in covariate effects in the two outcomes is beneficial for grouping the full picture of such effects with regard coronary artery bypass surgery, and other similar harsh treatments.

The process consists of two compartments. The first, a discrete-time process, monitors transitions out of the treatment state, State 1, into State 2, Treatment Success or State 3, Death. The second compartment is a continuous time Markov model monitoring transitions from State 2 to State 3. Figure 4.1 illustrates the complete process. For transitions from State 1, let $p = p_{12}$ be the probability of a $1 \rightarrow 2$ transition, or the probability of surviving treatment. The intensity governing $2 \rightarrow 3$ transitions is $\lambda(t)$. Let $Y_i(t)$ record the state occupied by individual i at time t, $t \in \{0, t \geq 30 \text{ days}\}$, $i = 1, \cdots, n$, and f_{12i} be an indicator variable for a $1 \rightarrow 2$ transition for an individual i. For an individual who has survived beyond 30 days after surgery, let T_i denote lifetime, L_i denote censoring time, $t_i = \min\{T_i, L_i\}$; and

$$u_i = t_i - 30, \quad \text{if} \quad f_{12i} = 1.$$

Probabilities governing transitions from State 2 are denoted by

$$P_{2k}(u(1), u(2)) = \Pr\{Y(u(2) + 30) = k | Y(u(1) + 30) = 2\}, \quad k = 2, 3$$

where $u(1)$ and $u(2)$ represent two different time points. To incorporate the effects of covariates, initially assume no connectivity between covariate effects influencing p_{12} and those influencing P_{2k}.

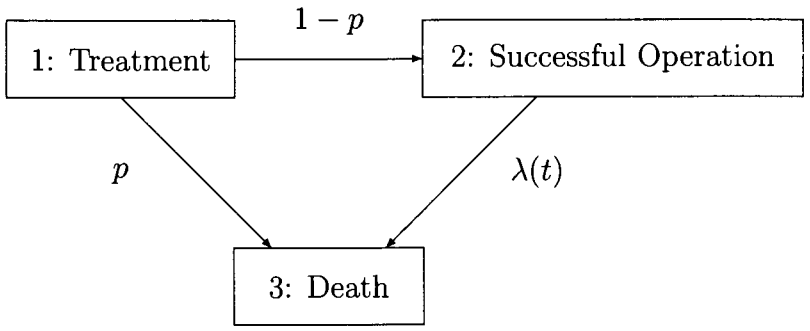

Figure 4.1: Three-State Model for Coronary Artery Bypass Surgery

Assuming a logistic linear model for p_{12} and a proportional intensity model for λ we have

$$\begin{aligned} \text{logit}(\mathbf{p}_{12}) &= \mathbf{G}\gamma \\ \lambda(u) &= \lambda_0(u)\exp(\mathbf{x}'\beta) \end{aligned} \tag{4.1}$$

and the likelihood function becomes

$$\begin{aligned} L &= \left\{\prod_{i=1}^{n}\frac{(e^{g_i\gamma})^{f_{12i}}}{(1+e^{g_i\gamma})}\right\}\left\{\prod_{u_i\in D}\lambda_0(u_i)e^{\mathbf{x}_i'\beta}\right\} \\ &\quad \prod_{u_i\in\{D\cup C\}}\exp\left\{-\int_0^{\infty}\text{Ind}(Y_i(u_i+30)=2)\lambda_0(u_i)e^{\mathbf{x}_i'\beta}du\right\} \end{aligned}$$

where Ind(A) is the indicator variable for event A, and $\mathbf{x}_i$ is the regression vector for the individual observed to fail at u_i, D is the set of times at which $2 \longrightarrow 3$ transitions are observed and C is the set of times at which transitions out of State 2 are censored. Note that $\lambda(t) = \lambda(u+30) = \lambda_0(u+30)\exp(\mathbf{x}'\beta)$.

For mounting a parametric approach, $\lambda_0(u)$ would have a fully specified parametric form. The partial likelihood approach may also be employed and yields the likelihood

$$L = \left\{\prod_{i=1}^{n}\frac{(e^{g_i\gamma})^{f_{12i}}}{(1+e^{g_i\gamma})}\right\}\left\{\prod_{u_i\in D}\frac{\exp(\mathbf{x}_i'\beta)}{\sum_j \text{Ind}(Y_j(u_i+30)=2)\exp(\mathbf{x}_j'\gamma)}\right\}$$

This is the identical likelihood as developed in Chapter 2, and hence that approach can be viewed as arising from a multi-state model.

The model used above, (4.1), is analogous to (2.2) in Chapter 2. Note that models which attempt to connect the relationship between the effects of covariates on $1 \rightarrow 2$ and $2 \rightarrow 3$ transitions, such as (2.3) and (2.4) may similarly be employed. Viewing the analysis as arising from a multi-state model may be convenient when considering extensions of the analysis where other intermediary endpoints between states 2 and 3 are of interest. Extensions of this sort are more easily managed in the multi-state modelling framework.

Appendix

Factor	Label	% Operative Mortality	$\hat{S}(1)$	$\hat{S}(3)$
Age	0-49	0.7	0.99	0.99
	50-54	1.5	0.98	0.97
	55-59	1.0	0.98	0.95
	60-64	1.9	0.96	0.93
	65-69	2.5	0.96	0.94
	70-74	3.7	0.95	0.90
	75-79	3.4	0.93	0.85
	> 80	7.0	0.87	0.75
Gender	males	1.9	0.96	0.93
	females	4.1	0.95	0.91
Ejection fraction	<35%	6.4	0.88	0.82
	35-50%	2.5	0.96	0.92
	>50%	1.4	0.97	0.95
Urgency of surgery	elective	1.4	0.97	0.93
	urgent	3.4	0.95	0.92
	emergency	8.8	0.93	0.89
Left ventricular end diastolic pressure	yes	3.6	0.94	0.90
> 15 mg Hg	no	1.9	0.97	0.94
Re-operation	yes	5.6	0.91	0.85
	no	2.2	0.96	0.93
Unstable angina/recent MI < 6 weeks	yes	3.2	0.95	0.92
	no	1.5	0.97	0.94

Table A1: Preliminary examination of potential risk factors

Factor	Label	% Operative Mortality	$\hat{S}(1)$	$\hat{S}(3)$
MI > 6 weeks	yes	2.2	0.96	0.92
	no	2.6	0.96	0.93
Chronic obstructive pulmonary disease	yes	3.9	0.93	0.87
	no	2.3	0.96	0.93
Diabetes	yes	3.8	0.94	0.89
	no	2.2	0.96	0.93
ASA within 5 days	yes	2.3	0.96	0.93
	no	2.6	0.96	0.92
Pre-op symptomatic arrhythmia	yes	6.6	0.93	0.85
	no	2.3	0.96	0.93
Pre-op intra-aortic balloon pump	yes	15.5	0.84	0.82
	no	2.3	0.96	0.93
Pre-op iv nitroglycerine	yes	5.6	0.93	0.89
	no	1.8	0.96	0.93
Pre-op diuretic	yes	7.3	0.87	0.78
	no	2.1	0.97	0.94
Pre-op iv heparin	yes	4.0	0.84	0.91
	no	1.7	0.97	0.93
Pre-op ventilation/intubation	yes	32.0	0.65	0.65
	no	2.3	0.96	0.93
Pre-op coumadin	yes	2.7	0.97	0.94
	no	2.5	0.96	0.93
Pre-op steroids	yes	4.1	0.94	0.92
	no	2.5	0.96	0.93
Pre-op thyroid replacement	yes	2.2	0.96	0.92
	no	2.5	0.96	0.93
Other endocrine disease	yes	3.2	0.97	0.92
	no	2.5	0.96	0.93
Hypertension	yes	3.4	0.95	0.90
	no	2.2	0.96	0.93

Table A2: Preliminary examination of potential risk factors

Factor	Label	% Operative Mortality	$\hat{S}(1)$	$\hat{S}(3)$
Peripheral vascular disease	yes	3.1	0.92	0.85
	no	2.4	0.96	0.93
Cerebrovascular disease	yes	4.0	0.92	0.86
	no	2.4	0.96	0.93
Dialysis/elevated creatinine	yes	4.4	0.91	0.94
	no	2.3	0.96	0.93
Congestive heart failure	yes	8.3	0.85	0.77
	no	2.0	0.97	0.94
Pulmonary hypertension	yes	11.4	0.82	0.70
	no	2.3	0.96	0.73
Number of diseased vessels	1-2 vessels	1.0	0.98	0.96
	3 vessels	2.4	0.95	0.92
	main left stenosis	3.3	0.95	0.90
Average household income	<40k	2.9	0.95	0.92
	40-44k	2.7	0.95	0.92
	45-49k	2.3	0.96	0.93
	>50k	1.7	0.96	0.93
Malignant disease	yes	3.9	0.94	0.94
	no	2.5	0.96	0.93
Urban/rural residence	greater Van/Vic	1.8	0.96	0.93
	other	2.8	0.95	0.92
Year of surgery	1991	2.2	0.96	0.92
	1992	2.5	0.95	0.92
	1993	2.3	0.97	0.94
	1994	2.5	0.96	0.94
Institution of surgery	Vancouver	2.8	0.96	0.92
	St. Paul	2.2	0.95	0.93
	Royal Jubilee	2.4	0.97	0.93
	Royal Columbian	1.6	0.96	0.93

Table A3: Preliminary examination of potential risk factors

Bibliography

[1] Breslow, N. E. (1974) Covariance analysis of censored survival data. *Biometrics*, **30**, 89-99.

[2] Cox, D. R. (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, **34**, 187-202.

[3] Cox, D. R. (1975) Partial likelihood. *Biometrika*, **62**, 269-276.

[4] Ghahramani, M. (1998) *Simultaneous Modelling of Long and Short Term Survival after Coronary Artery Bypass Graft Surgery.* M.Sc. Project, Simon Fraser University.

[5] Ghahramani, M., Dean, C.B. and Spinelli J.J. (2001) Simultaneous modelling of operative mortality and long-term survival after coronary artery bypass surgery. *Statistics in Medicine*, **20**, 1931-1945.

[6] Hess K. (1995) Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Statistics in Medicine*, **14**, 1707-1723.

[7] Lawless, J.F. (1982) *Statistical Models and Methods for Lifetime Data.* New York: John Wiley and Sons.

[8] Schoenfeld, D. (1980) Chi-squared goodness-of-fit tests for the proportional hazards regresion model. *Biometrika*, **67**, 145-153.