THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/73391

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

Parametric Dynamic Survival Models

by
Karla Hemming

A thesis submitted for the degree of Doctor of Philosophy

Department of Statistics
University of Warwick
Coventry
CV4 7AL

June 2000

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor Ewart Shaw for all of his help and guidance over the past 3 years, and the EPSRC for supporting me financially. I wish to thank also my colleagues for their valuable help. Finally thanks to Henry, my family and Pete.

Declaration

I declare that this thesis is my own work, and has not been submitted for a degree at another university.

Summary

A non-proportional hazards model is developed. The model can accommodate right censored, interval censored and double interval censored data sets. There is also an extension of the model to include multiplicative gamma frailties.

The basic model is an extension of the dynamic Bayesian survival model developed by Gamerman (1987), but with some alterations and using a different method of model fitting. The model developed here, the Normal Dynamic Survival Model, models both the log-baseline hazard and covariate effects by a piecewise constant and correlated process, based on some division of the time axis. Neighbouring piecewise constant parameters are related by a simple evolution equation: normal with mean zero and unknown variance to be estimated.

The method of estimation is to use Markov chain Monte Carlo simulations: Gibbs sampling with a Metropolis-Hastings step. For double interval censored data an iterative data augmentation procedure is considered: exploiting the comparative ease at which interval censored observations may be modelled.

The model is applied within a range of well known, and illustrative data sets, with convincing results. In addition the impact of censoring is investigated by a simulation study.

General Notation

D	Data including observations and priors.
$i = 1, ..., n_T$	Subscript for individuals.
$i = 1, ..., n_S$	Subscript for individuals in data set for S.
K	Number of observed covariates.
$k = 1, ..., K$	Subscript for covariates.
n_S	Number of individuals in the initiating study population.
n_T	Number of individuals in the study population.
T	Survival times.
$T = X - S$	Survival time based on observing X and S.
t	Observed value of T.
t_i	Survival or censoring time for individual i.
S	Survival time for the initiating event.
s	Observed value of S.
X	Calendar time of terminating event.
z_k	Covariate value k.

Time Axis Notation

β_j	Dynamic parameter for model.
D_j	Information observed up to time t_j.
$G_S = \{s_0, s_1, ..., s_{N_S}\}$	Time axis division for initiating model.
$G_T = \{t_0, t_1, ..., t_{N_T}\}$	Time axis division for survival model.
$I_j = (s_{j-1}, s_j]$	Interval on the initiating time axis.
$I_j = (t_{j-1}, t_j]$	Interval on the time axis.
$j = 1, ..., N_S$	Subscript for interval points on the initiating time axis.
$j = 1, ..., N_T$	Subscript for interval points on the time axis.
N_S	Number of intervals on initiating time axis.
N_T	Number of intervals on time axis.
s_j	Point on the initiating time axis.
t_j	Point on the time axis.

1 Introduction

Cox's proportional hazards model permits the impact of covariates on survival to be estimated, and using the partial or marginal likelihood approaches as in Cox (1972) and Kalbefleisch and Prentice (1980), estimates for the model parameters have been shown to have the same asymptotic properties as the maximum likelihood estimates (Tsiatis, 1981). Over the last three decades the model has been heavily used: it has formed a large part of the basis of statistical research in survival analysis, and has become probably the most commonly used multivariate survival model in medical applications. Its major drawback is the constraint of proportional hazards: an assumption which is not always appropriate. The consequences of using such a model when the assumption is not appropriate can be high, and for instance may result in the conclusion of a single superior treatment, when infact this may not be the case (Carter *et al.*, 1983). Aware of the constraints of the proportional hazards model, Gore *et al.* (1984) fitted several proportional hazards models in consecutive intervals (being careful to ensure an adequate amount of data within each of the intervals), thus avoiding an overall proportional hazards assumption. The method is not ideal; will only work with large data sets; and is very sensitive to the number and location of the intervals. Carter *et al.* (1983) included treatment multiplied by time as a time dependent covariate, thereby modelling a linear change in the treatment effect (the model

can be extended to model other polynomial changes). Gamerman (1987) used a piecewise constant baseline hazard and covariate effect, relating interval estimates by a simple semi-parametric relationship. Parameter estimates are sequentially updated as additional data are received, using conjugate and linear Bayes approximations, finally sequencing backwards through the intervals to obtain smoothed retrospective estimates. The piecewise constant correlated process (often abbreviated to piecewise correlated process), is an attractive way to model both the baseline hazard and covariate effects. Gray (1992), amongst others, used cubic splines to model both the baseline hazard and covariate effects over time. Apart from the quoted alternatives above, there exist few other survival models which are not based on the proportional hazards assumption.

For interval censoring, double interval censoring and frailty models (non-independent observations), similar constraints exist in that the majority, if not all, of the multivariate models are again based on Cox's proportional hazards model. Piecewise correlated functions have been used to model the baseline hazard for more complicated censoring types (Ghosh and Sinha, 1995), but the method has never been extended to modelling dynamic covariate effects. One of the main reasons for this has been the difficulty in estimating so many model parameters.

In this thesis a Markov chain will be constructed with a posterior distribution which models the data by piecewise correlated functions for both the baseline and covariate effects: thus avoiding the constraints of proportional hazards, and at the same time overcoming the usual difficulty of estimating model parameters. The model will be applied to standard right censored data, and also to additional types of censoring such as interval censoring, double interval censoring, and frailties. For right censored data the method will be an alternative to that proposed by Gamerman (1987). For more complicated data types, a non-proportional hazards model will thus have been developed.

In Chapter 2, many of the standard survival analysis methods are introduced (including both frequentist and Bayesian methods). A discussion is given as to why proportional hazards may sometimes be an unreasonable assumption to make, and some alternative methods are suggested. Recent advances in model fitting techniques are explained, and the method of Markov chain Monte Carlo (MCMC), including Gibbs sampling and Metropolis-Hastings, are described. Additional types of censoring (interval censoring, double interval censoring, and also non-independent observations), the problems which they bring, and methods which exist to deal with them are then discussed.

In Chapter 3, Gamerman's model (Gamerman, 1987) is introduced, and the method for estimating model parameters outlined. It is firstly noted how the paramater which determines the amount of evolution from interval to interval (the evolution variance) must be specified prior to model fitting. Secondly it is observed that the method used by Gamerman to fit the model could not be extended to accommodate interval, double interval censoring, or frailty models. For these reasons, a parametric version of Gamerman's model is introduced, called the Normal Dynamic Survival Model (NDSM). The model has the same basic structure as in Gamerman (1987) but assumes that the evolution of the parameters follows a normal distribution (it also models the evolution variance as a hyper-parameter). It is a parametric model, but only weakly so, and the model remains sufficiently flexible to accommodate a wide range of both proportional and non-proportional hazard functions. The model does however have one disadvantage, and that is no smooth estimates of covariate effects and baseline hazard exist, although survival estimates are smooth. To estimate model parameters, Gibbs sampling with a Metropolis-Hastings step is used. It is noted how Gibbs sampling could not be used in the semi-parametric model developed in Gamerman (1987), as without modelling the evolution by some parametric distribution, the full conditionals could not be computed. To improve the efficiency of the Gibbs sampler, a

reparametrisation is introduced, and the likelihoods are based on a temporal factorisation. The method is illustrated on a set of gastric cancer survival times (Gamerman, 1987). The development of the NDSM, and data applications using this model are new developments.

In Chapter 4, likelihoods are developed under the same Normal Dynamic Survival Model, for extended censoring types (including interval censoring, double interval censoring, and frailties). For interval censoring, the likelihood is fairly straightforward, and uses only a slight extension of the conditional survival functions (developed in Chapter 3 and which were used within the construction of the likelihood for the right censored model). For double interval censoring, it is observed that the exact likelihood can not be derived, and so an approximation is given in its place. For both interval censoring and double interval censoring, it is not possible to create a temporal factorisation of the likelihood and so a factorisation over each observation (individual factorisation) is derived as an alternative. For the right censored frailty model, the likelihood is first given as a temporal factorisation, and secondly as a factorisation over the groups: so that the most efficient factorisations may be used to compute different full conditionals. Likelihoods are similarly developed for the interval censored frailty models (it is also described how the likelihood could be constructed for the double interval censored frailty model). Finally

a small discussion is given on the concept and use of individual frailties. All of the work within Chapter 4 is new work, although some similarities may exist between likelihoods of others, and where this is the case it will be made clear.

In Chapter 5, it is shown how Gibbs sampling can be used to estimate the model parameters using the likelihoods developed in Chapter 4. For double interval censoring it is explained how either the approximate likelihood could be used in a full likelihood analysis using Gibbs sampling; or as an alternative, a method using imputation is outlined, where the data is augmented to interval or right censored data as appropriate. Applying MCMC techniques to survival data is by now a common feature. The techniques within this Chapter are not original, but applying them to the NDSM is. As with the right censored model, a reparmeterisation is used: finding an effective reparmeterisation is quite a common procedure, Gamerman (1998) also suggested it within the context of a dynamic generalised linear model, although work here was completely independent. Aslanidou *et al.* (1998) have constructed full conditionals for the frailty parameters in a similar model, and those full conditionals do bare some similarity to those constructed here, although their model is a proportional hazards model. To examine the impact that the degree of censoring has on the model, data sets will be simulated with increasing degrees of censoring. Survival estimates using the NDSM

will then be compared to the relevant non-parametric technique.

In Chapter 6, the methods are applied to some real data applications: including a set of breast cancer survival times, generating an interval censored data set (Finkelstein, 1986); a group of haemophiliacs infected with HIV and AIDS resulting in an incubation time which is double interval censored (Kim *et al.*, 1993); and some kidney infection times (McGilchrist and Aisbett, 1991) with non-independent observations. The data sets within this Chapter arise frequently in papers within the particular field. Although to knowledge they have never been analysed using a dynamic Bayesian survival model.

In any analysis of data throughout this thesis, assumptions will be made based on the dosage of the treatment, the nature of combination, and the ordering of the treatment, to name but a few. It will also be assumed, unless otherwise stated, random allocation of treatment, controlled clinical trials and independent censoring (Chapter 2, section 2.1). So when interpreting results from fitted models, caution must be exercised, not to come to medical conclusions which are not appropriate to the nature of the data. Reference will often be made to treatment effects, and treatment differences, but may be extended to cover all covariate effects (dependent upon the nature of the

study).

Very general terms and concepts commonly used in Bayesian statistics and survival analysis are not defined, although two very good reference for both basic and in depth concepts and methods are Carlin and Lewis (1998) and Collett (1994), with the latter being particularly relevant to survival analysis and the former for methods involving Markov chain Monte Carlo simulations. When terms are introduced for the first time, they will be highlighted by the use of italics.

All of the standard calculations are carried out using S-PLUS (S-PLUS, 1999). More complicated models were programed in C, on a sun-sparc ultra 250.

2 Survival Analysis

A brief and concise summary of survival analysis is given below. In the rest of this Chapter, and those that follow, the topic will be introduced in greater depth, where terms and phrases used in this initial brief summary will be properly defined.

Survival analysis is the term used to describe the analysis of the time between some defined time origin and a pre-defined failure event. There are several reasons why survival analysis differs from standard statistical data analysis. These include lack of symmetry (we often observe many short survival times and few long observations), restriction to positive outcomes (survival times can not be negative, which rules out using the normal distribution), and lack of observing all end points (censoring). Probably the main reason why survival analysis is so distinct is because of the censoring. Reasons for censoring include an individual withdrawing from the study early, still being alive at the end of the study, being lost to follow up, or dying from another cause. In such cases the survival time is known only to be greater than the time that the individual was last known to be alive, called the right censoring time. Even more complicated types of censoring arise when the failure is known only to have occurred within some interval of time, often due to the detection of the failure via a medical test, giving rise to interval censoring (section

2.8). Survival analysis is used in a complete range of applications, and the term failure does not always have to be associated with death. It could for instance denote the time from manufacture to breakdown of some mechanical component, or the time from prison release to date of reconviction, in the study of reconviction rates. In this thesis the term death is taken to incorporate both death in the usual sense, and failure of the more general kind.

Bayesian analysis of survival data consists of a survival model, with priors and possibly hyper-priors for the parameters (Sweeting, 1987). Estimates for model parameters are evaluated using both the data and the priors. Complications arise when the prior distributions are not conjugate to the likelihood, making the model intractable (section 2.4). Intractable models arise either because a conjugate prior (Carlin and Lewis, 1998) is not appropriate, or because the likelihood is too complicated to have an associated conjugate distribution. Until quite recently many Bayesian models were often either unrealistically constrained to conjugate priors; likelihoods simplified; approximation methods used; or otherwise were subject to difficult numerical integration. However with the development of the Expectation Maximisation (EM) algorithm; imputation methods; Markov chain Monte Carlo simulation techniques (section 2.7); along with the general increase in computational power; survival models have become much more realistic and at the same

time achievable.

Definition 2.1 ***The Survival Time***

The survival time, T, is defined to be the time between the initiating event S, and the time of the terminating event X. So that

$$T = X - S.$$

In birth - death processes, the initiating event time is zero, and the terminating event time is the age at death. For the incubation period of the Acquired Immune Deficiency Syndrome (AIDS), the initiating event time is the time of infection with the Human Immunodeficiency Virus (HIV); and the terminating event time is the time of progression from HIV to AIDS. In clinical trials, the initiating event time is the time of randomisation to treatment; and the terminating time will be the time of death, censoring, or remission (dependent upon the study).

Definition 2.2 ***The Time Axis***

As the majority of survival models to be considered in detail within this thesis are based on the evolution of the parameters over time, it is necessary to create a division of the time axis. The division of the time axis will be defined as:

$$G_T = \{t_0, t_1, ..., t_N\},$$

and intervals within this division will be referred to by the notation $I_j = (t_{j-1}, t_j]$ *for* $j = 1, ..., N$, *and* t_N *should be greater than the last observation time.*

2.1 Censoring and Truncation

A fundamental feature of survival analysis is that the failure time may not always be observed; in addition it is possible that only a subset of survival times are observed due to the nature of sampling involved in the data collection methods. In survival analysis this is known as censoring and truncation.

An observation is said to be *right censored* if it is observed only that the survival time is greater than some observed time point, called the right censoring time. In any analysis of survival data in this thesis, the assumption will be made that the probability of an observation being censored, does not depend on the survival time that would have been observed, had the observation not been censored (conditional on observed covariates), such types of censoring will be refered to as independent censoring.

As right censoring is the most common type of censoring, it is usual to

define a right censoring indicator for each observation:

$$\delta = \begin{cases} 1 & \text{death,} \\ 0 & \text{censored.} \end{cases} \tag{1}$$

An observation may also be *left censored.* Left censoring occurs when it is known only that the survival time is smaller than the left censoring time (when an individual contracts HIV, it is usually only known that the time of infection is prior to the first positive test).

Left truncation arises when an individual is included in the study only if the individual has a survival time which is greater than the so called left truncation time. *Right truncated* data arise when the individual is included in the observed data set only if the event of interest is experienced prior to the chronological time of data ascertainment.

Interval censored data arise when the survival time t is observed only to lie within some interval, commonly referred to as $(R, L]$, but the precise time of death is not known. The observed data therefore consists of $t \in (R, L]$.

Doubly censored data arise when the time of the initiating event, S, is observed to lie within an interval, so that $s \in (M, P]$. It is possible that data are censored on both the left and right, hence the term doubly censored. *Double interval censored data* arise when both the initiating and terminating event

are observed up to an interval only.

It is usual within a survival analysis, in addition to observing survival times, to also observe covariates. To make accurate inferences, it must be the case that conditional on all of the observed covariates, all observations are independent. Later within this thesis, non-independent observations will be considered, and the method of accounting for such dependences, called *frailty modelling* will be investigated.

2.2 The Survival and Hazard Functions

T is the random variable representing the survival time of an individual. The survival function, $S(t)$, refers to the probability that an individual has a survival time T which is greater than t:

$$S(t) = P(T > t).$$

The corresponding probability density function for T is f(t), with the distribution function of T given by:

$$F(t) = p(T \leq t) = \int_0^t f(u)du.$$

The hazard rate, $h(t)$, or the instantaneous death rate, is the conditional probability that an individual having survived to time t, will die at time

$t + \delta t$. More formally:

$$h(t) = \lim_{t \to 0} \frac{P(t < T < t + \delta t \mid T > t)}{\delta t}.$$

From the above definitions it is easy to verify the following relationships between the survival function and the hazard rate:

$$h(t) = \frac{f(t)}{S(t)}.$$

The cumulative hazard

$$H(t) = \int_0^t h(s)ds,$$

is such that:

$$H(t) = -\log(S(t)).$$

Unless otherwise stated the continuous distributions will be used to model the survival distribution $S(t)$.

2.3 The Likelihood

The likelihood function will be defined throughout by L. The notation $L(\theta|x)$ will be used to emphasise the likelihood as a function of the parameters and data, but this will often be abbreviated to $L(x)$ (where x denotes the data and θ the parameters). L_i will denote the likelihood contribution for an individual i, and similarly L_j will denote the likelihood contribution from an interval I_j.

From an exact observation, the contribution to the likelihood will consist of $f(t_i)$, from a right censored observation the contribution will be $S(t_i)$. The likelihood accommodating exact and right censored data, may then be written as:

$$L = \prod_{i=1}^{n} S(t_i)^{1-\delta_i} f(t_i)^{\delta_i}.$$

Using relationships between the hazard and density function:

$$\begin{aligned} L &= \prod_{i=1}^{n} S(t_i) h(t_i)^{\delta_i} \\ &= \prod_{i=1}^{n_1} S(t_i) \prod_{i=1}^{n_2} f(t_i). \end{aligned} \tag{2}$$

Here n_1 is the number of right censored observations, n_2 is the number of exact observations, and $n = n_1 + n_2$.

2.4 The Bayesian Approach

A Bayesian analysis uses both observed data and priors to make estimates for model parameters in a given model (classical statistical methods use the data only). The posterior may be obtained from the likelihood and priors by using Bayes' Theorem (Carlin and Lewis, 1998):

$$p(\theta|x) = \frac{L(\theta|x)p(\theta)}{\int L(\theta|x)p(\theta)d\theta}.$$

Up to a constant of proportionality this is:

$$p(\theta|x) \propto L(\theta|x)p(\theta).$$

Priors for the parameters may come from previous similar studies, or they may be based on expert opinions. Where there is no such detailed information then vague priors may be used. Sinha (1997) argues that priors should be based on expert opinions or be based on stage zero studies, however it is also acknowledged in that same paper that this may not always be possible.

Where the prior is conjugate to the likelihood, then the posterior will be tractable. Essentially this means that for any observed data, the likelihood is such that the posterior will belong to the same family as the prior. Examples of practical Bayesian survival analysis are given in Raftery *et al.* (1996) and Dellaportas and Smith (1993). A very simple example is given below, which will form the basis of some calculations later in the thesis.

Definition 2.3 ***The Gamma Distribution***

A random variable is said to have a gamma distribution:

$$\theta \sim G(\alpha, \gamma),$$

when its probability density function takes the form:

$$p(\theta) = \frac{\gamma^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\gamma\theta},$$

with mean and variance:

$$E(\theta) = \frac{\alpha}{\gamma} \text{ and } V(\theta) = \frac{\alpha}{\gamma^2},$$

where $\Gamma(\cdot)$ *is the Gamma function and is defined by* $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$.

Example 2.4 ***Constant Hazard - Gamma Prior***

The survival is modelled using the constant and proportional hazards model:

$$h(t) = \lambda e^{z\lambda},$$

using the notation $\theta = \lambda e^{z\lambda}$. *The chosen conjugate prior for* θ *is the gamma distribution, with suitably chosen parameters:*

$$\theta \sim G(\alpha, \gamma).$$

The survival function for a set of exact and right censored data, with hazard θ *is:*

$$\begin{aligned} S(t) &= \exp(-\int_0^t h(u)du) \\ &= \exp(-\theta t). \end{aligned}$$

The likelihood (using equation 2) is therefore:

$$\begin{aligned} L(T) &= \prod_{i=1}^{n} S(t_i)h(t_i)^{\delta_i} \\ &= \exp(-\theta \sum_{i=1}^{n} t_i)\theta^{\sum_{i=1}^{n}\delta_i}. \end{aligned}$$

Using Bayes' Theorem, the updated distribution is obtained:

$$p(\theta|x) \propto L(\theta|x)p(\theta)$$

$$p(\theta|x) \propto \theta^{\alpha-1}e^{-\gamma\theta}\theta^{\sum_{i=1}^{n}\delta_i}e^{-\theta\sum_{i=1}^{n}t_i}$$

which is proportional to a gamma density, so that:

$$\theta|x \sim G(\alpha', \gamma'),$$

where

$$\alpha' = \alpha + \sum_{i=1}^{n} \delta_i \text{ and } \gamma' = \sum_{i=1}^{n} t_i + \gamma.$$

2.5 Survival Models

There exist many methods for modelling survival data. The methods may be roughly broken down into three groups: non-parametric, semi-parametric, and parametric. Each of these three groups may be sub-divided into Bayesian and classical type approaches.

Probably the most commonly used non-parametric model is the Kaplan-Meier estimate (Kaplan and Meier, 1958). The Kaplan-Meier model is for a single sample only, as it does not incorporate covariates. Strictly speaking the Kaplan-Meier curve is not defined at the exact observation points, and is also not defined after the last observation point (if the last observation is a censored observation).

Proportional hazards is another common method for modelling survival data, and may either be of a semi-parametric or non-parametric form. The semi-

parametric version was first introduced by Cox (Cox, 1972). For an individual with covariate vector z, the hazard is the product of the baseline hazard $\lambda_0(t)$, multiplied by a function of the covariates:

$$h(t) = \lambda_0(t) \exp(z\beta).$$

The baseline hazard represents the hazard of an individual with all covariates at the baseline, i.e. zero. The part of the hazard $e^{z\beta}$ is known as the relative hazard. Cox's proportional hazards model is a semi-parametric model and without some parametric form assumed for the baseline hazard, maximum likelihood estimates for the parameters can not be derived. Cox's approach was to use a conditional likelihood, which depended only on the observed death times. Kalbfleisch and Prentice (1980) claimed that the conditional likelihood needed additional justification, and derived a marginal likelihood (for the case of no ties the marginal and partial likelihoods are equivalent). In Cox (1975) the condition likelihood was renamed the partial likelihood and Efron (1977) showed that inferences based on Cox's partial likelihood are asymptotically equivalent to those based on all of the data.

One method for modelling the baseline hazard in a proportional hazards model, was introduced by Breslow (1974), and is called the *piecewise constant baseline hazard model.* In this model the baseline is modelled as a

series of constants spanning the time axis, so that for $t \in I_j$:

$$\lambda_0(t) = \lambda_j \text{ for } j = 1, ..., N,$$

where the time axis is divided into N intervals. This form of baseline hazard has been used in several applications in the statistical literature. The *piecewise correlated baseline hazard,* has the additional feature that the piecewise baseline hazards are related across intervals. This relationship, which may either be modelled parametrically or non-parametrically, is often referred to as the evolution equation. Usually either the evolution of the baseline hazard is modelled by a gamma distribution (Aslanidou, Dey and Sinha, 1995), or the log of the evolution of the baseline is modelled, this time following a normal distribution (Ghosh and Sinha, 1995). Other commonly used processes for modelling the baseline hazard are Lévy processes (Kalbfleisch, 1978), although the independence assumption involved within this process has been criticised (Arjas and Gasbarra, 1994).

Widely used parametric proportional hazards model include modelling the baseline hazard using the Weibull (Aitkin and Clayton, 1980) or exponential distribution. These are parametric analogues of Cox's proportional hazards model. Although restricted by their parametric nature, these models do have the added advantage of tractable maximum likelihood estimates.

There exist both graphical and non-graphical methods for testing the validity of the assumption of proportional hazards, a detailed account may be found in Anderson *et al.* (1992). Further theoretical work on testing the assumption of proportional hazards may be found in Grambsch and Therneau (1994). Their method is based on using a weighted function of the Schoenfeld residuals (Schoenfeld, 1982), estimating how the covariate effects may change over time. They further develop a chi-squared test statistic with proportional hazards being the null hypothesis. One of the main drawbacks of this method is that no survival prediction is established based on the alternative estimate for the covariate effect. Within Splus a graphical estimate of the covariate effect over time is produced along with confidence intervals, and observed Schoenfeld residuals.

2.6 Non-proportional Hazards

Within a proportional hazards analysis, there is an assumption that the effects of covariates are constant over time. Constant coefficients are essential to the idea of proportional hazards. An extension of the proportional hazards model is to make the relative hazard function a function of both the covariates and time (denoted here by $g(z,t)$):

$$h(t) = \lambda_0(t) \exp(g(z,t)).$$

Various forms of the relative hazard function $g(z,t)$ will be discussed, and those of particular interest are those which result in non-proportionality.

In the treatment of some illnesses by surgery, the initial hazard may be very high, much larger than under any other non-surgical treatment. Once the patient has passed through this initial critical stage, the hazard may decrease far below that of the non-surgical alternatives: and so treatments will not have proportional hazards. Another example of non-proportional hazards occurs in the analysis of breast cancer patients by stage. Here late staged individuals have a comparatively high hazard during the first 10-20 months after diagnosis, but then in relation to early staged cases, the high hazard drops considerably.

Gore *et al.* (1984) fitted a survival model to a data set of breast cancer patients. They found that the effect of the treatment diminished over time, contrary to the usual proportional hazards assumption. As an alternative, a stepwise proportional hazards model was fitted. That is, within several different divisions of the time axis, separate, unrelated proportional hazards models were used. The model identifies the need for dynamic covariate effects, but is limited to large data sets and a small number of intervals (so as to ensure that there is sufficient data in each interval to provide accurate

estimates).

Carter *et al.* (1983) fitted a survival model to a set of gastric cancer data, and examined the possibility of covariates varying over time. In preference to estimating the effects of the covariates independently, over different divisions of the time axis, they instead included as an additional parameter: the covariate as a function of time, and estimated its effect. The method models a linear change in the covariate effect over time, and may be extended to include other polynomial changes. Unfortunately linear order polynomials are seldom appropriate for modelling.

Gamerman (1987) also considered an analysis of this same gastric cancer data set, using a dynamic Bayesian survival model. The dynamic Bayesian survival model, models both the log of the baseline hazard and the covariate effects by a piecewise correlated process, using a non-parametric evolution distribution. Using some conjugate assumptions, parameters are estimated using a *linear Bayes approximation* (section 3.1). The model developed in Gamerman (1987) is considered in detail in the following Chapter.

A further group of models which include non-proportionality are those where the covariate effect is modeled by a spline function. Gray (1992) considered

modelling both the baseline hazard and covariate effects by cublic splines. Although noted that for modelling the covaraite effects, due to instability in the right tails of the distribution, piecewise constant splines were used instead of cubic.

2.7 Methods for Model Fitting

So far most of the discussion has concentrated on models that may be solved using maximum likelihood estimates, conjugate Bayesian methods, or some other relatively standard and straightforward technique. However in the following Chapters and the rest of this one, survival models will be introduced that are increasingly flexible, not only in the types of hazards which they incorporate, but also in the types of data that they can accommodate. In order to estimate parameters for such models, the methods for model fitting need to generally be more powerful: conjugacy and maximum likelihood estimation are no longer adequate. Dellaportas and Smith (1993, page 443) are also of this opinion:

> " ..a major impediment of the routine Bayesian implementation in this large class of models (referring to survival analysis) has certainly been the difficulty of evaluating the integrals required."

An example of a conjugate / tractable analysis was given in section 2.4, unfortunately, most analysises are not this simple: likelihoods are often much

too complicated to have an associated conjugate distribution. Approximate normality could be used, but there are many instances when this will not be appropriate. Numerical integration (Reilly, 1976) is another possible alternative, which although a useful approach, and one which has been used within Bayesian survival analysis (Greive, 1987), can only really be considered in applications of up to around twenty dimensions. The Expectation-Maximisation (EM) algorithm, Dempster, Laird and Rubin (1977), is an iterative extension of the maximum likelihood technique, ideally suited for missing data applications; but is a frequentist approach and does not accomodate prior distributions (although there do exist Bayesian variations on this technique). *Multiple imputation,* again a technique for dealing with missing data, iteratively replaces the missing data values with appropriately simulated values. It is ideal for use when conditional on the missing data the model is in some sense tractable (as is the EM algorithm). Data augmentation (Tanner, 1996) has some similar properties to the EM algorithm: it exploits properties of the likelihood (or posterior) for an agumented data set.

Monte Carlo simulations are useful when trying to find the expected value of some function $g(x)$:

$$E[g(x)] = \int_x g(x)f(x)dx.$$

If this integration cannot be solved analytically, then an approximation may be obtained by repeatedly simulating from $f(x)$, and estimating the expectation, as the mean of the corresponding simulated $g(x)$. Where direct simulations from $f(x)$ are not possible (for example in high dimensions), then simulations may be taken by observing a Markov chain which has $f(x)$ as its equilibrium distribution. This method is known as taking *Markov chain Monte Carlo (MCMC)* simulations and is becoming ever increasingly popular, not only in Bayesian survival analysis but also more generally in Bayesian statistics. This is because to construct a Markov chain which has the distribution $f(x)$ as its posterior, it is only necessary to know $f(x)$ up to a constant of proportionality: thus avoiding the difficult integration problem involved in computing the constant of proportionality within a Bayesian analysis.

Definition 2.5 ***A Markov chain Monte Carlo Simulation***

- *Find a Markov chain with the target posterior as its unique equilibrium distribution.*
- *Simulate from the Markov chain, until the chain is at the equilibrium, and disregard all sampled values prior to this point.*
- *Take repeated samples from the equilibrium distribution to obtain many Monte Carlo samples.*

Clearly the difficulty consists of finding an appropriate Markov chain which has the posterior as its equilibrium distribution. Several possible well established algorithms, which do just this, are described below.

The Metropolis and Hastings samplers, Metropolis *et al.* (1953) and Hastings (1970), iteratively propose possible parameter values, each of which are in turn accepted or rejected. Roberts and Smith (1994) and Tierney (1994) describe the conditions under which the sampling algorithms will converge to the posterior. All of these sampling algorithms have the same basic structure but variations exist in terms of the proposals and acceptance probabilities. The algorithms were first introduced by Metropolis *et al.* (1953), but after some alterations by Hastings (1970), they are often referred to as the Metropolis-Hasting algorithms. As mentioned there exists various forms of these samplers, and the one that has been used in this thesis is the Hastings sampler. The reason why this sampler has been chosen is that the values that are proposed are dependent on the chain's current position (the proposal is a sensibly chosen distribution with mean at the current value and appropriately chosen variance). So in some sense the proposal may be thought of as dynamic with respect to the chain. The choice of the proposal distribution is important, and proposals which are close to the posterior distribution will result in gains in efficiency.

The Hastings Sampler

- Sample a value θ' from the proposal $q(\theta'|\theta)$ conditional on the current value θ.

- Accept the sampled value with probability $\alpha(\theta', \theta)$, where:

$$\alpha(\theta', \theta) = \min\Big(1, \frac{p(\theta'|\theta)q(\theta|\theta')}{p(\theta|\theta')q(\theta'|\theta)}\Big),$$

where $p(\cdot)$ represents the posterior; $q(\cdot)$ the proposal density function; θ represents the current value of the parameter; and θ' the proposed value.

Another Markov chain which will allow samples to be taken from the posterior distribution is the Gibbs sampler. The Gibbs sampler, not without reason, has become probably the most used form of MCMC in Bayesian survival analysis. Its popularity is based on its ease of handling multivariate parameters. Suppose that the object of the analysis is to provide estimates for N parameters:

$$p(\theta_1, ..., \theta_N).$$

Gibbs sampling repeatedly samples from the full conditionals:

$$[\theta_i|\cdot] \quad = p(\theta_i|\theta_1, ..., \theta_{i-1}, \theta_i, ..., \theta_N),$$

replacing the current estimate θ_i by the sampled value. Writing $\theta = (\theta_1, ..., \theta_N)$, the Gibbs sampler takes the following form:

Definition 2.6 ***The Gibbs Sampler***

1. *A starting value* $\theta^0 = (\theta_1^0, ..., \theta_N^0)$ *is chosen for the complete multivariate parameter space.*

2. θ_1 *is sampled from its full conditional, conditional on the initial estimates for all of the other parameters.*

3. θ^0 *is updated in place 1 with the sampled value for* θ_1 *to obtain* $(\theta_1^1, \theta_2^0 ..., \theta_N^0)$.

4. *The process is repeated for each parameter, and many iterations are carried out*

Gemen and Gemen (1984) showed that, under weak conditions, repeated samples as described above converge in distribution to the marginal distribution of the parameters. Gelfand and Smith (1990) amongst others have applied Gibbs sampling to general Bayesian statistics.

Gibbs sampling requires samples to be taken at each iteration from the full conditional distribution. In practice the full conditionals will rarely be well known distributions (which would allow sampling to be via standard techniques). Where the full conditional is not of standard from, then some form of sampling technique must be used. Devroye (1986) provides an excellent account of many of the current sampling techniques: examples

include sampling-resampling (Gelfand and Smith, 1990) and rejection sampling. Most of these sampling techniques are very dependent upon the distribution to be sampled from (the full conditional in this application). Because Gibbs sampling involves generating samples from many different densities, a method is needed that not only provides a sample from the correct distribution, but which is also efficient and generally applicable. Two of the most frequently used methods in Gibbs sampling, which do just this are adaptive rejection sampling, and a Metropolis-Hastings step.

Gilks and Wild (1992) proposed the method of adaptive rejection sampling for cases where the full conditional is log-concave (Devroye, 1986), and an extension was given in Gilks *et al.* (1996) for cases where the full conditional is almost log concave. The technique uses the fact that a log-concave distribution may be bounded above and below by piecewise constant hulls, allowing a rejection sample to be easily implemented (without having to know the location of the modes of the distribution). Dellaportas and Smith (1993) applied Gibbs sampling, with adaptive rejection sampling, to a Bayesian survival problem. Although the method is time consuming to programme in comparison to the Metropolis-Hastings step (below), it does have the advantage of always providing a sample from the current full conditional.

A Metropolis-Hastings step could also be used to sample from the full conditional, see Tierney (1991) and Gelman and Rubin (1993). A value is sampled from the full conditional by sampling a possible value from a proposal and accepting it with an acceptance / rejection criteria. Being easy to implement, the method does not necessarily provide a sample from the full conditional until the chain has reached equilibrium, although good choices of proposals can improve this. Sargent (1998) considered such a technique in a survival analysis application.

After implementing an MCMC simulation, checks should be made to ensure that the chain has reached the equilibrium distribution, and that only those values sampled after reaching equilibrium are used in any subsequent analysis (some chains may converge very slowly). Only in very limited circumstances do there exist exact checks on convergence. More generally a range of convergence diagnostics are used. These convergence diagnostics are an aid only and do not provide a definitive answer as to whether convergence has occurred or not. Indeed many of the convergence diagnostics themselves rely on a range of assumptions and approximations. There exists a complete range of published and unpublished material on convergence diagnostics, and an excellent review of the current methods is provided by Cowles and Carlin (1996). These methods range from the theoretical to very

intuitive: for example Arjas and Gasbarra, (1994) are confident that their model has converged when the resulting survival curves for a large data set are similar to the Kaplan-Meier curves. In this thesis the convergence diagnostics provided by the package CODA (Best *et al.*, 1993) are used to assess convergence, and to determine the required length of the "burn in".

Further methods which are related to MCMC include the *MCEM* algorithm (Wei and Tanner, 1990) where the E step in the EM algorithm is replaced by Monte Carlo simulations, and *data augmentation* (Tanner and Wong, 1987) which is an iterative method for finding the posterior distribution (rather than just the mode as the MCEM algorithm does), data augmentation may be used in conjunction with Gibbs sampling (Tanner, 1996).

Now that some advanced methods of model fitting have been introduced, it is possible to consider some of the current methods for dealing with more complicated censoring types within a survival analysis.

2.8 Interval Censoring

Interval censored data, initially defined in section 2.1, arise when the survival time is observed to lie within an interval of time, called the censoring interval. When monitoring the occurrence of some disease which may be identified

only by a medical test, then the observed failure time will be known to lie between the last negative examination and the first positive examination, resulting in interval censored observations. An example of an interval censored data set is given in example 2.7.

Grouped data are related to interval censored data: all that is known about the survival time of an individual, is that their survival time lies within some interval. However with grouped data, at every censoring or survival time, it is possible to determine a risk set: it is known exactly how many patients are at risk and how many have died since the last observation point. Risk sets are known with grouped data as the censoring intervals do not overlap: this enables the data to be ranked. Prentice and Gloecker (1978) give a detailed account of an extension of Cox's proportional hazards model for grouped data. Turnbull (1976) extends the Kaplan-Meier estimate to accommodate grouped data.

As interval censored data can not be ranked, the usual survival methods, such as Cox's proportional hazards, or the Kaplain-Meier curve are not immediately adaptable. Peto (1973) developed life table techniques for interval censored data. Turnbull (1976) constructed a Non-Parametric Maximum Likelihood Estimate (NPMLE), based on a "self consistency" algo-

rithm. Frydman (1994) and Alioum and Commenges (1996) later extended and amended Turnbull's method to be applicable for both interval censoring and truncation. Pan and Chapell (1998) show that Turnbull's estimate can underestimate survival at early times for left truncation (due to small risk sets). Under a non decreasing hazards assumption, and using a gradient projection algorithm, they provide an alternative estimate which they claim has a superior performance (in terms of both bias and variance).

Finkelstein (1986), extended Cox's proportional hazards model for both right censoring and interval censoring, based on an extension of Turnbull's self consistency algorithm. Unlike Cox's original model, in Finkelstein's extension, the baseline needs to be estimated. Satten (1996) has developed a proportional hazards model which not only does not require the baseline to be estimated, but has the additional feature that the model reduces to the usual proportional hazards assumption as the length of the censoring intervals shrink to zero (unlike Finkelstein's model). More recently, Goggins *et al.* (1998) developed an EM algorithm (where the E step is replaced by MCMC simulations), again for the analysis of interval censored data under the assumption of proportional hazards. Finkelstein and Wolfe (1985) develop a semi-parametric regression survival model, using maximum likelihood techniques.

Ghosh and Shina (1995) have developed a semi-parametric model for interval censored data, based on an piecewise correlated baseline hazard process (no covariates were included): using a posterior likelihood approach they advocate that the method may be used to check the assumption of proportional hazards. Sinha (1997) modelled the baseline hazard by a discrete version of a beta Lévy process, and used Markov chain Monte Carlo simulations to estimate the parameters. More generally Sinha and Dey (1996) give a review of Bayesian survival analysis, and include interval censoring. Pan (2000) uses a form of data augmentation to model interval censored survival data by Cox's proportional hazards model.

There are alternative methods for dealing with interval censored data, such as substituting interval midpoints as exact survival times, or by using some other estimate of what the true survival time may have been. Using midpoints or right end points may give biased results, especially when the censoring intervals are large. Estimates for the precision of the estimates will also be overestimated, as the uncertainty associated with the substituted values will not be accounted for. Using a Weibull based accelerated failure time model, Odell *et al.* (1992) compared a midpoint analysis with that based on maximum likelihood estimates of the real interval censored data. Their

conclusions were that that the maximum likelihood method generally gave a better fit, especially where hazards were not constant, censoring intervals were long, and there was a large percentage of interval censored data.

With interval censored data, it may be less clear (compared with right censored data) whether the censoring mechanism is independent. If for example the interval censoring mechanism is generated by patients visiting a doctor, it may be possible that the onset of symptoms may make it more likely that a patient will either keep an appointment or make an earlier one. Similarly if a patient feels healthy they may be more inclined to miss an appointment. Extra care should be taken to be sure that the interval censored data are censored in an independent way. Farrington and Gay (1999) offer an approach for dealing with interval censored data where the censoring is informative. But the approach does rely on detailed information of all visits to the clinician, which will often be unavailable. As previously mentioned, in this thesis the assumption of an uninformative censoring mechanism will always be made.

Example 2.7 ***Breast Cancer and Cosmetic Outcome***

Breast cancer patients monitored every 4 to 6 months for cosmetic deterioration following the treatment of either chemotherapy or chemotherapy combined with radiotherapy, generate a set of interval censored data which

were studied in Finkelstein and Wolfe (1985) . Standard independence assumptions of the censoring mechanism were assumed. Justification for these assumptions were as follows:

> *"According to the medical investigator, the cosmetic deterioration did not affect the patients' return to the clinic, and thus necessary assumptions of independence of the censoring and failure distributions are satisfied".*

This data set may be found in table 7 (page 198), and will be examined in more detail in Chapter 6.

2.9 Double Interval Censoring

Double interval censoring, explained within this thesis in section 2.1, occurs when each of the initiating event and terminating event are observed only to lie within some interval. In effect the data are interval censored on both the left and right. For individual i the interval for the initiating event is denoted by $(M_i, P_i]$, and the censoring interval for the terminating event is $(R_i, L_i]$. This is a generalisation of where the data are *doubly censored:* implying that the data are censored on the left, usually interval censored; and censored also on the right, but usually right censored.

One of the reasons why data analysis that accommodates doubly censored

data has become increasing interesting, is due to the study of HIV and AIDS. Infection with HIV can only be ascertained by a screening test. If a series of such screening tests were available, along with the corresponding negative and positive outcomes, then the date of HIV infection could be identified to lie within the interval of the most recent negative and first positive test dates. The Center for Disease Control, has defined a set of AIDS defining conditions (including wasting, dementia and Kaposi's sarcoma): so that an individual moves from being HIV positive to having AIDS, when they have one or more of these conditions. The surveillance definition has changed over time, with more recent additions including invasive cervical cancer and CD4+ counts below 200 units (CD4+ cells are depleted as the HIV spreads through the body). Some of the defining conditions are in some sense subjective (wasting), and others such as Kaposi's sarcoma are evaluated by medical screening: progression to AIDS is much less clearly defined than time of HIV infection, and the time of progression of AIDS varies from being tabulated as exact, to interval censored. Where data are tabulated on the right as interval censored, then that is how they will be analysed; similarly where they are tabulated as exact or right censored then that is also how they will be analysed. A fuller discussion of the AIDS virus is given in Chapter 6, section 6.3.

After using this example as an illustration, it is noted that doubly censored

data sets do not frequently arise within the analysis of the AIDS incubation distribution. It is only in special circumstances that the time of HIV infection can be determined to lie within some censoring interval as described above. Usually it is observed only that the time of HIV infection occurred before the patient tested positive for HIV (left censoring). This type of information is not very informative for prediction of the AIDS incubation distribution.

There are generally two circumstances where such censoring occurs in AIDS studies. The first is where partners of HIV infected individuals are monitored over time by regular checks for HIV. The second type of data situation, which is the type of application which will be used in this thesis, is illustrated via an example below.

Example 2.8 ***Doubly Interval Censored AIDS Data***

An example of doubly interval censored data (Brookmeyer and Goedert, 1989), concerns the study of haemophiliacs receiving replacement blood clotting factor at three different treatment centres during the 1980's. Blood samples were stored and subsequently tested for HIV infection, so that the date of infection can be determined up to an interval of time. The subjects were followed up and times at which they progressed to AIDS are also recorded. This data set is interval censored on the left and exact or right censored on the right.

Doubly censored data is not limited to AIDS cases, but may also arise in completely different, medical or non-medical applications. One possible application is in the analysis of the survival times of patients diagnosed with breast cancer after having taken part in a screening programme. If the time of developing a tumour is known to lie between two screening visits, then the initiating event will be interval censored. If the women are subsequently followed until death, the time of the terminating event will be exact or right censored. Problems do arise when considering tumours detected at non screening visits (lack of uninformative censoring) and questions as to whether the tumour was missed at the last screen, or whether it really is a new growth.

In the AIDS example described within example 2.8, it is clear that the censoring mechanism for the infection time is independent of the survival event of interest. This is because the serum samples were stored for reasons completely independent of the HIV virus (any lost serum samples were due to events such as freezer malfunctions). With other AIDS data sets, patients may be more likely to be make frequent visits to be tested for HIV infection when they have reason to suspect that they may be infected, and so the independence of the censoring mechanism can no longer be assumed. It is much less certain, whether the censoring interval for the time of AIDS progression

is independently censored. These patients were followed up after it had been identified that they were HIV positive, and visits to the clinician are likely to become more frequent when CD4+ cell counts are low. This problem is acknowledged, but not considered further.

Brookmeyer and Goedert (1989) have developed a model which accommodates interval censoring on the left and right censoring on the right. For the time of HIV infection, the log-baseline hazard is modelled by a piecewise-constant hazard, with constant covariate effects. The Weibull model is used to model the incubation distribution. The decision to use the Weibull distribution was based on previous studies cited in their paper (Hessol *et al.*, 1987), which suggested that the hazard for developing AIDS increased continuously after diagnosis (although Bacchetti (1990) estimates that the hazard will flatten out after around seven years). The model is a two stage model, which first estimates the parameters in the infection model, and subsequently estimates the parameters in the incubation model. The model used is a continuous model, and for an individual with left censoring interval $(M, P]$, and AIDS diagnosis time t, has a likelihood contribution:

$$\int_M^P f_1(s) f_2(t-s) ds,$$

where $f_1(s)$ is the distribution of the HIV infection times, and $f_2(t)$ is the distribution function for the AIDS incubation times. The Newton-Rapson

algorithm was used to find the maximum likelihood estimates.

DeGruttola and Lagakos (1989) develop an extension of Turnbull's model (using the "self consistency algorithm") for double interval censoring. It is a non-parametric model, which does not include covariates, and has some similar properties to the Kaplan-Meier curve (non unique after the last censoring time). The model may be used to check any future parametric assumptions. Again assuming that an individual has initiating censoring interval $(M, P]$ and terminating censoring interval $(R, L]$, the likelihood contribution for such an individual is:

$$\sum_{s=M}^{P} p_1(S = s)p_2(R - s < T \leq L - s).$$

Where $p_1(.)$ is the discrete distribution modelling the infection times S; $p_2(.)$ is the discrete distribution modelling the incubation times T; and s runs over the discrete set of admissible initiating event times within the interval $(M, P]$.

Kim *et al.* (1993) furthered the work of the above two cited references, by developing a proportional hazards model for double interval censored data, again based on an iterative method using the self consistency algorithm (Turnbull, 1976). As already noted under interval censoring, Turnbull's self consistency algorithm is highly dependent on the starting value (Sinha, 1997), and this will also be true when applying the method to double

interval censoring. Bacchetti (1990) comments that incubation estimates are either based on strong parametric assumptions, or if they are non-parametric then they may be unstable, and so proposes a method which does not use individual level data, but rather is based on aggregate data, and is similar to the method of back calculation (Brookmeyer and Gail, 1994).

Law and Brookmeyer (1992) show how wrong inferences can be drawn when using a midpoint imputation approach, as did Odell *et al.* (1992) for interval censored data. DeGruttola and Lagakos (1989) note that midpoint imputation will only be appropriate where the density for infection is uniform in chronological time (imputation can also underestimate the standard error of the coefficients).

Other data sets which arise in the study of the incubation period of the AIDS virus, are transfusion related data sets. Such data sets arise when individuals who have undergone blood transfusions, later develop AIDS. The time of HIV infection can often be linked to the date of the transfusion. Although such data sets will not necessarily be double interval censored, they do fall under the wide heading of estimating the incubation distribution. However such data sets will often be right truncated, and so will lead to biased results towards shorter incubation periods if the truncation is not accounted

for. What is more, the estimate of the incubation distribution may not be representative of the population as a whole, as transfusion patients are often elderly and frail, again biasing estimates towards shorter incubation periods (although similarly a cohort of haemophiliacs cannot be assumed to be representative of the population as a whole). Truncation is not considered further within this thesis, although applications may be found within Kalbfleisch and Lawless (1996) and Finkelstein *et al.* (1993).

2.10 Frailty

Suppose we are interested in modelling the survival time of breast cancer patients where the study population consists of some genetically related women (mothers and daughters, sisters etc.). It is now well known that some cancers have a genetics basis: with this knowledge it is expected that some of the survival times (which could include both time to onset of the disease and the time of diagnosis till death) may be related.

Within any standard survival analysis, a fundamental assumption is that conditional on observed covariates, the survival times are independent. In some studies the data may be grouped: in the above example the groups represent family relationships; in other examples the groups may represent such things as geographical location, genetic traits, or treatment centres. Within

groups, it may be the case that the survival times are positively related to each other. By conditioning on the groups, as with conditioning on covariates, the survival times become independent. Dummy covariates could be assigned, but the model will almost certainly be required to make general predictions for the whole population: interest focuses on the marginal distribution, and not the conditional.

In another example, suppose the effects of two treatments are to be compared across several different hospitals. An observational study is carried out, and two covariates defined, one for treatment indicator, and the second for treatment centre. A survival model, possibly Cox's model (assuming the proportional hazards assumption is valid), could be fitted to the data, and evaluations could then be made as to which treatment, within each different treatment center, produces the better survival rates. This method of analysis, will not tell us which treatment is best across all treatment centres. On the other hand, if we were to ignore the treatment center, and include as the only covariate the treatment indicator, then the survival times conditional on covariates, may no longer be independent. Again as interest lies in the marginal distributions, standard survival models are not appropriate.

One solution to this problem is to assign a grouping effect, called a frailty,

to each cluster or group. Estimates for each grouping effect are made, and the unconditional distribution, that is the marginal survival distribution, is used to make general population survival estimates. Frailty, or heterogeneity, models were first introduced into demography by Vaupel *et al.* (1979). They have subsequently evolved in both frequentist and Bayesian statistics. There are many variations of the frailty model (some of which will be discussed), and the basic, conditional proportional hazards model, with gamma frailties is introduced below. But first an example of a data set which has appeared in several papers with a frailty interpretation, and which will also be referred to several times in this thesis.

Example 2.9 ***Kidney Patient Infection Times***

38 patients are monitored for kidney infection whilst undergoing dialysis. Infection times for each individual are recorded, as the time from insertion of a catheter until infection. After the infection has healed, another catheter is inserted and the time until infection is recorded again. For each patient up to two survival times were recorded. The object of the study is to compare the survival times across a variety of covariates. As observations for each individual are expected to be related, this could be accounted for by using frailties. This is the approach that has indeed been taken first by McGilchrist and Aisbett (1991) and later by Asalnidou et al. *(1998), amongst others. This data set may be found in table 10 (page 205) of this thesis, and again*

will be discussed in greater depth in Chapter 6.

The Gamma Frailty Model

- The conditional proportional hazard at time t, for individual i in group g, with frailty u_g acting multiplicatively, is:

$$h(t|z_{ig}, u_g) = \lambda_0(t) \exp(z_{ig}\beta) u_g,$$

  that is the product of the baseline hazard, the relative hazard, and the multiplicative frailty u_g.

- The frailties follow independent gamma distributions, with unit mean (ensures identifiability) and variance μ^{-1}:

$$u_g \sim G(\mu, \mu).$$

- Conditional on the frailty, u_g, the model is defined to be a proportional hazards model, although this relationship does not hold marginally. The marginal hazard is:

$$h(t|z_i) = \frac{\mu e^{z_i\beta} \lambda_0(t)}{\mu + \Lambda_0(t) e^{z_i\beta}},$$

  where $\Lambda_0(t)$ is the integrated baseline hazard.

The marginal hazard will be derived in full, for the Normal Dynamic Survival Model in Chapter 4, section 4.4: the method there is a generalisation of the

proportional hazards frailty model which is being described here, and so the marginal hazard is not derived here, merely given.

Marginally for the gamma frailty model, there will be a convergence of hazards, the rate of which will depend on the parameter of the gamma distribution μ (Clayton, 1991). It may be argued that long term survivors will all have approximately the same hazard, irrespective of treatment and other covariates, indicating a convergence of hazards. However this will not always be the case, and especially may be untrue where the data are not long term observations. To avoid marginal convergence of hazards Hougaard (1986) considered using the positive stable distribution: a one parameter distribution, characterised by its Laplace transform. The positive stable frailty distribution is the only distribution which models a proportional hazards model both conditionally and marginally. Sinha (1998) points out that the positive stable distribution assumes that the frailty paramater has an infinite variance, and so implicitly assumes the possibility of an infinite number of infections: it should therefore be avoided in data applications like the kidney infection data, where there are just two infections per group. After saying this, it is noted that Qiou (1997) compares the gamma frailty with the positive stable frailty and conclude that for the kidney infection data the positive stable frailty model performs better, in terms of predictive distributions.

Another popular method for modelling the frailties is to use the log-normal distribution, as in Gray (1994) and McGilchrist and Aisbett (1991).

Frailties are unobservable random variables, and there exist only limited ways of checking the appropriateness of model assumptions. Hougaard (1986), demonstrated how important the choice of the frailty distribution is, and how different choices can lead to remarkably differing results. Sargent (1998) comments how most choices for frailty distributions are restricted so as to obtain closed form full conditionals. A comparison was made in the same paper between various different distributions for the frailties, and in that particular application little difference was observed between any of the models. Paik *et al.* (1994), modeling the baseline hazard by a piecewise constant process, discuss the impact of the chosen frailty distribution on the marginal, and consider the case of a time dependent, rather than constant frailty effect.

In the above example of the basic frailty model, the baseline hazards was left unspecified, as in Cox's proportional hazards model. Depending on the type of model fitting technique that is to be used, some parametric assumptions for the distribution of the baseline may have to be made. Clayton (1991) amongst others used a Lévy process, where increments in the cumulative baseline were assumed to be independent gamma variables. As previ-

ously remarked for the non frailty model, increments in neighbouring intervals could be related. An alternative parametric distribution is the Weibull proportional hazards model, used for example by Aalen (1988), again the appropriateness of this distribution will depend on the nature of the data, as it will only model (conditional) hazards which are monotonically increasing or decreasing.

Since Clayton (1991) first applied Bayesian methods to frailty models, by extending the freqentist approach of Clayton and Cuzick (1985), Bayesian frailty models have received a fair amount of interest. A general review is given by Sinha and Dey (1996). Sinha (1998) develops a gamma frailty model, with a piecewise correlated baseline hazard for right censored data using the posterior likelihood approach. Aslanidou and Dey (1996) compare the gamma frailty model with a piecewise correlated baseline (gamma) hazard against a model with the Weibull baseline hazard (also with gamma frailties). Using Gibbs sampling with Metropolis-Hastings, again for the kidney data set, they found that the piecewise correlated baseline hazard method provided a much better fit.

2.11 Summary

A broad discussion of survival analysis has now been given. Both general and more complicated types of data have been discussed. It has been noted that one of the major constraints of survival models for more complicated data types, is the assumption of proportional hazards. Various forms of model fitting techniques have also been considered, with the intention of being used with the development of a non-proportional hazards model for complicated types of censoring (interval censoring, double interval censoring, and frailty applications).

3 Dynamic Survival Models

It has already been noted that misleading, or incorrect, conclusions could be drawn when fitting a proportional hazards model, where the assumption of proportional hazards is not appropriate. On the other hand, in every day applications of medical statistics, it is important to be able to carry out a multivariate analysis. These were probably the reasons which lead Gamerman (1987) to develop a multivariate, non-proportional hazards, survival model. Gamerman's model is a very effective way of avoiding the assumptions involved within the usual multivariate models. But for reasons which will be explained later, it is not possible to incorporate both interval and double interval censored data sets into this model. It is for this reason that after introducing Gamerman's model, the model will be changed slightly, and a different method of model fitting used, in a way which will make the model more easily adaptable to an extended range of data types. Comparisons will be made with the results obtained using this new approach with those obtained by Gamerman (1987). In the following Chapters, this adapted model will then be applied in a range of interval censored and double interval censored data sets.

Gamerman (Gamerman 1991, and Gamerman 1987) called his model a dynamic Bayesian model for survival data: both the log-baseline hazard and

covariate effects are modelled by the piecewise correlated process. In this thesis all of the non-proportional hazards models based on the piecewise correlated process will fall under the broad heading of Dynamic Survival Models (DSM). The adapted model is based on a normal parameterisation, and will be called the Normal Dynamic Survival Model (NDSM).

3.1 The Dynamic Bayesian Model for Survival Data

In this first section the original model, developed by Gamerman (1987), will be introduced, and the method of model fitting used in Gamerman (1987) outlined. It will be shown how the method relies on being able to establish a risk set at each point on the time axis. Details will not be given as to why this restricts the model to right censored data (this will be left until the model is extended in the following section); here the restriction will be merely noted. The model will then be discussed, and any weak points of the model highlighted, with a view to possibly being rectified when the model is re-defined in section 3.2. Some of the notation used in this first section is slightly different to that used by Gamerman (1987), but it is on the other hand consistent with the notation used within this thesis.

Gamerman's model is an adaptation of a set of models known as the Dynamic Linear Models (West and Harrison, 1997). These time series models

allow the covariate effects to vary over time. The model fitting techniques are based on a Bayesian methodology, which sequentially updates priors into posteriors, as observations are received. Within a proportional hazards survival model the impact of introducing a dynamic covariate effect is to create a non-proportional hazards model. Again using the notation z_k for $k = 1, ..., K$ to represent covariates, and a division of the time axis with intervals referred to by the notation $I_j = (t_{j-1}, t_j]$.

Definition 3.1 ***The Dynamic Survival Model***

The model is defined via its hazard function, which at time t is defined as:

$$h(t) = \exp(\beta_0(t) + \sum_{k=1}^{K} z_k \beta_k(t)),$$

where

$$\beta_0(t) = \beta_{0j} \text{ for } t \in I_j,$$

$$\beta_k(t) = \beta_{kj} \text{ for } t \in I_j.$$

Parameters are related across intervals by an evolution equation:

$$\beta_{0j} = \beta_{0j-1} + w_{0j}, \qquad \text{where } w_{0j} \sim [0, W_0],$$

$$\beta_{kj} = \beta_{kj-1} + w_{kj}, \qquad \text{where } w_{kj} \sim [0, W_k],$$

with priors $\beta_{00} \sim [m_0, C_0]$, *and* $\beta_{k0} \sim [m_k, C_k]$, *for* $k = 1, ..., K$. *Information about survival times, censoring indicators, and covariates is denoted by* D,

and $D_{i,j}$ denotes information from all observations in intervals $1, ..., j$, and the first i observations in interval I_{j+1}.

For notational convenience the hazard may also be written in the abbreviated form:

$$h(t) = \exp(z\beta(t)),$$

where

$$z\beta(t) = (1, z_1, ..., z_k)(\beta_0(t), \beta_1(t), ..., \beta_k(t))^T.$$

Within interval I_j, for a given covariate, the hazard will be constant:

$$h(t) = \lambda_j \text{ for } t \in I_j,$$

where $\lambda_j = \exp(\beta_{0j} + \sum_{k=1}^{K} \beta_{kj} z_k)$. Similarly within any interval the covariate effect will also be constant. The model will approximate a continuous function, the true yet unknown hazard, by a piecewise constant one, and by including a sufficient number of intervals on the time axis, the model should provide a good approximation to all hazard functions. The greater the number of divisions on the time axis, the greater the potential is to create a very good approximation, although increasing the number of intervals increases the number of parameters to be estimated. A balance must be drawn.

The Linear Bayes Method

The method of estimating parameters described here and used by Gamerman (1987), is based on a linear Bayes approximation (West and Harrison, 1997). This estimate is a linear estimate of the parameter, whose value minimises a loss function (where the loss function will represent the accuracy of the estimate). The estimate may be compared to the non-Bayesian estimation techniques of minimum variance and least squares.

At the start of each interval a prior is defined for the parameters of the current interval (where the current interval is the first interval then this prior relates to initial prior beliefs, otherwise it is based on parameter estimates from the previous interval). As information is received within the interval, the priors are updated into posteriors. Since the parameters are dynamic, at each change of interval, the posterior for the current interval is transformed into a prior for the upcoming interval. Sequencing through all observations in all intervals results in the estimation of $\theta_j|D_{j-1}$ (θ_j is a general representation for some parameter at time t_j; and D_{j-1} represents observed information, up to and prior to, time t_{j-1}). Once information from all N intervals has been incorporated, it is then necessary to sequence back through the intervals and update the estimates from the form $\theta_j|D_{j-1}$ to $\theta_j|D_N$, these are often referred to as the smoothed or retrospective estimates.

The number of observations at risk (i.e. still alive) at the beginning of interval I_j is denoted by r_j. Every observation within the risk set for a particular interval, is given a modified survival and censoring indicator, for that interval. This will represent whether an observation was observed (i.e. through failure or right censoring), or whether that observation survived through the interval, and so with respect to the interval under question, is right censored at the end of the interval. For interval I_j, the modified censoring indicator, and survival time are:

$$\begin{aligned} \delta_{ij} &= \delta_i \text{ if } t_i \leq t_j \\ &= 0 \text{ if } t_i > t_j, \end{aligned} \tag{3}$$

$$\begin{aligned} t_{ij} &= t_i \text{ if } t_i \leq t_j \\ &= t_j \text{ if } t_i > t_j, \end{aligned} \tag{4}$$

where t_i represents the survival or censoring time for individual i, and the standard censoring indicator δ_i was defined on page 12.

Method Outline

So as to keep the notation simple, in this following section β_j will be used to refer to the vector $(\beta_{0j}, \beta_{1j}, ..., \beta_{kj})$.

1. Given information $D_{i-1,j-1}$ (i.e. all information from the first I_{j-1} intervals and the first $i-1$ observations in interval I_j), suppose the

current estimate for β_j is:

$$\beta_j | D_{i-1,j-1} \sim [m_{i-1,j-1}, C_{i-1,j-1}].$$

Note $\beta_0 | D_0 \sim [m_0, C_0]$, and $D_0 = D_{0,0}$ represents prior information at time zero.

2. Let λ_{ij}, be the hazard for individual i in interval I_j, where:

$$\lambda_{ij} = \exp(z_i \beta_j).$$

3. The joint distribution for β_j and $\log \lambda_{ij}$, conditional on $D_{i-1,j-1}$,is:

$$\begin{bmatrix} \beta_j \\ \log \lambda_{ij} \end{bmatrix} \Big| D_{i-1,j-1} \sim \left[\begin{pmatrix} m_{i-1,j-1} \\ f_{ij} \end{pmatrix} \begin{pmatrix} C_{i-1,j-1} & s_{ij} \\ s_{ij}^T & q_{ij} \end{pmatrix} \right]$$

where

$$f_{ij} = E[z_i \beta_j | D_{i-1,j-1}] = z_i m_{i-1,j-1},$$

$$q_{ij} = \mathrm{Var}[z_i \beta_j | D_{i-1,j-1}] = z_i C_{i-1,j-1} z_i^T,$$

$$s_{ij} = \mathrm{Cov}[z_i \beta_j, \beta_j | D_{i-1,j-1}] = C_{i-1,j-1} z_i^T.$$

4. It is assumed that $\lambda_{ij} | D_{i-1,j-1}$, has a gamma distribution, which will provide a partially tractable analysis:

$$(\lambda_{ij} | D_{i-1,j-1}) \sim G(\alpha_{ij}, \gamma_{ij}).$$

To ensure consistency between the first two moments under the gamma distribution and those from the joint distribution:

$$\alpha_{ij} = q_{ij}^{-1},$$

$$\gamma_{ij} = q_{ij}^{-1} \exp(-f_{ij}).$$

5. On receiving information from observation i in interval I_j, the distribution for λ_{ij} is updated by Bayes rule as:

$$p(\lambda_{ij}|D_{i,j-1}) \propto p(\lambda_{ij}|D_{i-1,j-1})L(t_{ij}|D_{j-1}),$$

where $L(t_{ij}|D_{j-1})$ represents the likelihood contribution from observation i in interval I_j, at time t_{ij}, conditional on survival up to the beginning of the interval:

$$L(t_{ij}|D_{j-1}) = \exp(-e^{x_{ij}\beta_j}(t_{ij} - t_{j-1}))e^{x_{ij}\beta_j\delta_{ij}}.$$

Likelihood contributions similar to the one above will be derived under a more general setting later in this Chapter (section 3.2.1), and of course for this specific model the complete derivation may be found in Gamerman (1987).

Following the method used in the conjugate gamma example (section 2.4), the updated distribution is:

$$\lambda_{ij}|D_{i,j-1} \sim G(\alpha_{ij} + \delta_{ij}, \gamma_{ij} + t_{ij} - t_{j-1}).$$

6. The distribution for β_j must then be updated. Writing the updated distribution as:

$$(\beta_j|D_{i,j-1}) \sim [m_{i,j-1}, C_{i,j-1}],$$

where

$$m_{i,j-1} = E[\beta_j|D_{i,j-1}]$$

and

$$C_{i,j-1} = \text{Var}[\beta_j|D_{i,j-1}],$$

updating is obtained using a linear Bayes approximation: full details of which may be found in West and Harrison (1997) and Gamerman (1987). Gamerman (1987), showed that the updated moments for the distribution of β_j are:

$$m_{i,j-1} = m_{i-1,j-1} + \frac{s_{ij}}{q_{ij}} \log\left(\frac{1 + q_{ij}\delta_{ij}}{1 + q_{ij}(t_{i,j-1} - t_{j-1})\exp(f_{ij})}\right),$$

and

$$C_{i,j-1} = C_{i-1,j-1} - \frac{\delta_{ij} s_{ij} s_{ij}^T}{1 + q_{ij}}.$$

This method is used to sequentially update paramater estimates for interval I_j as observations are received. Steps 1 to 6 are repeated until all information from interval I_j has been incorporated.

7. After receiving all information within interval I_j, the current posterior for β_j, is:

$$\beta_j|D_{r_j,j-1} \sim [m_{r_j,j-1}C_{r_j,j-1}],$$

which is equivalent to:

$$\beta_j|D_j \sim [m_j, C_j].$$

As there are no more observations within interval I_j, the current posterior is updated into a prior for the next interval:

$$\beta_{j+1}|D_j \sim [a_{j+1}, R_{j+1}],$$

where $a_{j+1} = m_j$ and $R_{j+1} = C_j + W$ (following the definition of the evolution of the parameters across intervals).

8. Once all the data has been sequentially imputed, the next step is to find the retrospective estimates: that is the estimates based on the complete data set. This is not shown here, but may be found in Gamerman (1987), and is a very similar procedure to that developed in West and Harrison (1997) for the Dynamic Linear Model.

Other forms of the basic evolution $\beta_j = \beta_{j-1} + w_j$ were considered in Gamerman (1987), and consisted of the form $\beta_j = G(\beta_{j-1}) + w_j$ for some function $G(\cdot)$. Gamerman (1987) used this form of evolution to include linear growth of the treatment effect, within the analysis of the gastric cancer data set (section 3.3). This is helpful when interest lies in modelling the rate of change of the treatment effect, and also more generally increases the flexibility of the model.

Apart from the initial priors (which can be left vague), two of the main uncertainties within the model, are the time axis division and the evolution variances. Both of these have a fairly large impact on the inferences obtained, yet both must be specified by the user. Kalbfleisch and Prentice (1973) argued that the division of the time axis should be determined independently of the data, although Breslow (1974) used a division based on the observations. Both of these papers use a piecewise correlated baseline hazard only, although the arguments should equally apply to dynamic covariates. Gamerman (1987) determined a "best" division by comparing Bayes factors under different divisions, and found for his application, that one of the divisions which worked well was a division based on the observed death times. Arjas and Gasbarra (1994) included the time axis division as an unknown quantity within their model, and estimated along with the other model parameters, where and how many divisions on the time axis there should be. This approach certainly avoids many of the uncertainties involved in choosing the time axis, but may be costly in terms of computational time.

Another very important feature of the Dynamic Survival Model, is the evolution variance. This parameter can have a very large effect on the estimated parameters: with a small evolution variance, the parameters will be restricted

to small changes over time; the opposite effect will occur if the evolution variance is large. The linear Bayes approximation method provides no way of estimating this parameter. Values could be chosen by comparing Bayes factors, although this would time consuming, and it would be more appealing if this parameter could be estimated within the model. It may also be realistic to model an evolution variance which changes over time (and indeed theoretically the above model can accommodate this). This would be especially useful if expert knowledge suggested a time when the series moved from being fairly static to volatile (or vise versa). Gamerman (1987) used dynamic evolution variances which are a function of the length of the interval:

$$W_{0j} = (t_j - t_{j-1})W_0.$$

A dynamic evolution variance becomes more important when the intervals of the time axis are of varying length, although where the intervals are sufficiently small, then it may be reasonable to assume a constant evolution variance (Sinha, 1998).

3.2 Parametric Dynamic Survival Models

In section 3.1, a non-proportional, semi-parametric survival model was described, that was first introduced in Gamerman (1987). Perhaps the major shortcoming of the model, is that in its current form it can not easily be

adapted to accommodate alternative types of censoring, such as interval and double interval censoring. This is because when data are interval censored, it is not possible to create a rank ordering. This in turn means that it is not possible to identify a risk set at each point on the time axis. Without this, parameter estimates can not be sequentially updated as under the linear Bayes method. It will be explained in Chapter 4 (section 4.4), why the model is also not immediately adaptable to frailty models.

With this in mind the Dynamic Survival Model, as introduced by Gamerman (1987), is reanalysed, this time using Markov chain Monte Carlo simulations. By using MCMC, the likelihood will not be restricted to a factorisation over the time axis, and instead the more standard likelihood factorised over each individual contribution may be used. The estimates obtained using the approach of Gamerman (1987) were smooth, although this is not possible with the method proposed here.

The hazard function is as given in section 3.1 and in Gamerman (1987), although in contrast to Gamerman, in this approach the evolution distributions must be modelled parametrically. Without this parametric assumption, the full conditionals for the Gibbs sampler would not exist. The evolution variances on the other hand, need not be known in advance, and may be

treated as hyper-parameters to be estimated along with the other parameters.

Definition 3.2 ***The Normal Dynamic Survival Model (NDSM)***

Introducing the notation $T \sim NDSM(\beta, N, W)$ *(where* β *represents the dynamic set of parameters,* N *the number of divisions on the time axis, and* W *the evolution variances), the Normal Dynamic Survival Model is defined via the following hazard function:*

$$h(t) = \exp(\beta_0(t) + \sum_{k=1}^{K} z_k \beta_k(t)),$$

where

$$\beta_0(t) = \beta_{0j} \text{ for } t \in I_j,$$

$$\beta_k(t) = \beta_{kj} \text{ for } t \in I_j.$$

Evolution of parameters takes the form:

$$\beta_{0j} = \beta_{0j-1} + w_{0j} \qquad \text{where } w_{0j} \sim N[0, W_0].$$

$$\beta_{kj} = \beta_{kj-1} + w_{kj} \qquad \text{where } w_{kj} \sim N[0, W_k],$$

The final stage in the hierarchy is to model the evolution variances, with hyper-priors:

$$W_0^{-1} \sim G(\zeta_0, \eta_0) \text{ and } W_k^{-1} \sim G(\zeta_k, \eta_k).$$

Priors $\beta_{00} \sim N[m_0, C_0]$, $\beta_{k0} \sim N[m_k, C_k]$ for $k = 1, .., K$, and ζ_0, η_0, ζ_k, η_k must be specified in order to complete this Bayesian model. Dynamic covariates can readily be incorporated into the model.

The model could be extended to incorporate a dynamic evolution variance, although this would be at the expense of increasing the number of parameters to be estimated. As an effective alternative, where the intervals are not of equal length, then the evolution variance can easily be modelled as a function of the length of the interval.

Although the model provides a flexible approach, it is a parametric model, and so some distribution assumptions have to be made. In this thesis the normal and gamma distributions have been chosen to model the evolution and inverse of the variance of the evolution respectively. The method proposed to estimate the model parameters, will be able to accommodate any appropriate choice of distribution. Indeed for the multiplicative baseline model:

$$h(t) = \beta'_{0j} \exp(\sum_{k=1}^{K} z_k \beta_{kj}) \text{ for } t \in I_j,$$

an alternative, positive distribution would be required for modelling the evolution of the baseline, for instance the gamma distribution.

Motivation for choosing the normal distribution is mainly underlined by the

work by West and Harrison (1997). In Dynamic Linear Models, the normal distribution should, for the most part, provide an adequate representation of the evolution. If there exists a sudden change in the effect, intervention plays an important role (these are after all Bayesian models). In survival analysis, it may also be possible to incorporate intervention and model monitoring, although these would be more complicated due to censoring, and residuals not being so well defined (this approach is not investigated in this thesis). The t-distribution has wider tails, and could be useful in medical applications, where patients may follow periods of stability, intermingled with periods of being unwell. There clearly exist many possibilities for modelling the evolution. For simplicity, and also because for the most part normality is expected to be more than adequate, the normal distribution has been chosen, but this is not to say that in other applications some other distribution could not be used.

The gamma distribution has been chosen to model the inverse of the evolution variance. This will provide full conditional distributions for these hyper-parameters which are also gamma, thus allowing sampling within the Gibbs sampler to be of a standard form. Furthermore, the gamma distribution is very flexible in that it will model a variety of shapes. After saying this the method is by no means constrained to the gamma distribution, and

any other reasonable distribution could be used at little extra computational expense (simply by inserting an additional Metropolis-Hastings step).

For the Normal Dynamic Survival Model, two different interpretations of the likelihood will be considered, both of which are defined below. The first is the temporal factorisation, as derived in Gamerman (1987), and defined within definition 3.3; and the second is what has been labelled the "individual" factorisation (definition 3.4). The temporal factorisation turns out to be the more efficient likelihood when carrying out a Gibbs sample (section 3.2.6). The individual factorisation is also introduced, as it will be this likelihood which must be used in later Chapters of this thesis, where under more complicated censoring mechanisms it will no longer be possible to create a temporal factorisation. The reason for not being able to construct a temporal factorisation, is exactly why the linear Bayes method could not be used: both rely on identifying risk sets, and establishing rank orders.

Definition 3.3 ***The Temporal Factorisation***

A temporal factorisation of the likelihood consists of contributions from each of the separate intervals on the time axis. Within each interval contribution, there is a factorisation over all of the individuals who are still alive at the beginning of the interval. The individual contribution within a particular interval will either consist of the probability that the individual died (or was

right censored) within the interval, given that they were alive at the beginning of the interval; or the probability that they survived through the interval, again conditional on survival at the beginning of the interval. The temporal factorisation is dependent on the ranking of the data.

Definition 3.4 ***The Individual Factorisation***

An individual factorisation of the likelihood, factorises the likelihood into contributions from every individual. Each individual contribution is then factorised over the time axis.

Before either of the likelihoods are derived, several basic conditional survival functions are developed. Further, more complicated, conditional survival functions will be developed throughout this thesis, as and when required.

3.2.1 Conditional Survival Functions (1)

Considering the general case of $T \sim NDSM(\beta, N, W)$, with $j = 1, ..., N$, the probability of surviving past some t, for $t \in I_j$, conditional on survival the beginning of the time interval (i.e. at t_{j-1}) is:

$$\begin{aligned} S(t|T_{j-1}) &= p(T > t|T > t_{j-1}) \\ &= \exp(-e^{z\beta_j}(t - t_{j-1})), \end{aligned} \tag{5}$$

where

$$T_j = \text{Event}[T > t_j].$$

So as to emphasise the model parameter β, this conditional probability will be often written as:

$$S(t|T_{j-1}) = C(t, t_{j-1}, \beta). \tag{6}$$

Proof

Firstly we observe:

$$\begin{aligned} S(t|v) &= \tfrac{S(t)}{S(v)} \\ &= \exp(-\textstyle\int_v^t h(u)du). \end{aligned}$$

Then:

$$\begin{aligned} S(t|T_{j-1}) &= \exp(-\textstyle\int_0^t h(u)du + \int_0^{t_{j-1}} h(u)du) \\ &= \exp(-\textstyle\int_{t_{j-1}}^t h(u)du) \\ &= \exp(-e^{x\beta_j}(t - t_{j-1})). \end{aligned}$$

3.2.2 The Temporal Factorisation

Let r_j represent all individuals who are yet to be observed at the beginning of interval I_j; r_j may be thought of as a risk set at time t_j. Modified censoring indicators and survival times (equations 3 and 4, page 58) were introduced under the linear Bayes approximation and will again be used within this factorisation. Although the linear Bayes method will not be used here, the likelihood is in fact the same as that required under the linear Bayes method: both relying on the temporal factorisation (although within this thesis, the full likelihood was not derived under the linear Bayes method). The form of

the temporal likelihood differs to that derived in Gamerman (1987), as the models are slightly different.

The likelihood is factorised over the N intervals:

$$L = \prod_{j=1}^{N} L_j,$$

where the likelihood contribution in interval I_j is:

$$L_j = \prod_{i=1}^{r_j} S(t_{ij}|T > t_{j-1}) h(t_{ij}|T > t_{j-1})^{\delta_{ij}}.$$

The likelihood factorises each interval contribution by considering all those who survive through the interval, and all those who fail within the interval, all conditional on survival at the beginning of the interval. The conditional survival function, $S(t|T)$ was derived in section 3.2.1 (page 70); and $h(t|T)$ is simply $h(t)$; the full form of the temporal factorisation is:

$$L = \prod_{j=1}^{N} \prod_{i=1}^{r_j} \exp(-e^{x_{ij}\beta_j}(t_{ij} - t_{j-1})) e^{x_i \beta_j \delta_{ij}}. \tag{7}$$

Before the individual factorisation is developed, it is necessary to introduce two further sets of conditional survival functions.

3.2.3 Conditional Survival Functions (2)

The probability of survival past some point t_m on the time axis, conditional on survival at time t_g on the time axis is:

$$\begin{aligned} p(T > t_m | T > t_g) &= S(t_m | T_g) \\ &= \exp(-\sum_{j=g+1}^{m} e^{x\beta_j}(t_j - t_{j-1})). \end{aligned}$$

Throughout the rest of this thesis the above probability will be referred to by the notation:

$$A(a, b, \beta) = \exp(-\sum_{j=a}^{b} e^{x\beta_j}(t_j - t_{j-1})). \tag{8}$$

So that:

$$S(t_m | T_g) = A(g+1, m, \beta).$$

Proof

Following the method used in the derivation of the first conditional survival function:

$$\begin{aligned} p(T > t_g | T_m) &= \exp(-\int_{t_g}^{t_m} h(u)du) \\ &= \exp(-\sum_{j=g+1}^{m} \int_{t_{j-1}}^{t_j} e^{x\beta(u)}du) \\ &= \exp(-\sum_{j=g+1}^{m} e^{x\beta_j}(t_j - t_{j-1})). \end{aligned}$$

Alternatively the probability may be derived from a factorisation over the time axis:

$$\begin{aligned} p(T > t_g|T_m) &= \prod_{j=g+1}^{m} S(t_j|T_{j-1}) \\ &= \prod_{j=g+1}^{m} \exp(-e^{z\beta_j}(t_j - t_{j-1})) \\ &= \exp(-\sum_{j=g+1}^{m} e^{z\beta_j}(t_j - t_{j-1})). \end{aligned}$$

3.2.4 Conditional Survival Functions (2B)

Using the conditional survival functions which were developed in the previous two sections, it is now straightforward to calculate $f(t)$ for any t. Suppose that $t \in I_{m+1}$:

$$\begin{aligned} f(t) &= S(t)h(t) \\ &= p(T > t_m)p(T > t|T > t_m)h(t). \end{aligned}$$

The first conditional survival function (equation 5), is the second part of this probability; and the second conditional survival function (equation 8), allows us to compute the first part of the probability:

$$\begin{aligned} f(t) &= p(T > t_m)p(T > t|T > t_m)h(t) \\ &= A(1, m, \beta)C(t, t_m, \beta)h(t) \\ &= \exp(-\sum_{j=1}^{m} e^{z\beta_j(t_j - t_{j-1})}) \exp(-e^{z\beta_{m+1}}(t - t_m))e^{z\beta_{m+1}}. \end{aligned} \tag{9}$$

3.2.5 The Individual Factorisation

It has already been noted (and will be demonstrated in section 3.2.6) that the temporal factorisation is more efficient with respect to Gibbs sampling; and so the individual factorisation will not actually be used within any analysis of right censored data. The individual likelihood is still derived however, for consistency with other censoring types which will be discussed in later Chapters. Furthermore the individual factorisation is often the clearest way to think about the likelihood, and does not require specification of the modified censoring and survival times (as does the temporal factorisation).

For individual i, with observation time t_i, $t_{m_i} \in G_T$ (recall that G_T denotes the set of division points on the time axis), is found such that:

$$t_{m_i} < t_i \leq t_{m_{i+1}}.$$

That is, it is identified within which interval every observation lies, with the notation that $t_i \in I_{m_{i+1}}$.

The individual likelihood, is taken as the product of each of the n individual contributions:

$$L = \prod_{i=1}^{n} S(t_i)h(t_i)^{\delta_i}.$$

Each individual contribution is the product of conditional survival functions (this consists of the product of conditional survival functions over each complete interval of survival and the conditional survival function over the partial interval in which failure or censoring occurs):

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m_i} S(t_j|T_{j-1})S(t_i|T_{m_i})h(t_i|t_i \in I_{m_{i+1}})^{\delta_i}.$$

Using the notation introduced in section 3.2.3:

$$L = \prod_{i=1}^{n} A(1, m_i, \beta)C(t_i, t_{m_i}, \beta)e^{x_i\beta_{m_{i+1}}\delta_i}.$$

Written out in full:

$$L = \prod_{i=1}^{n} \exp(-\sum_{j=1}^{m_i} e^{x_i\beta_j}(t_j - t_{j-1})) \exp(-e^{x_i\beta_{m_{i+1}}}(t_i - t_{m_i}))e^{x_i\beta_{m_{i+1}}\delta_i}. \quad (10)$$

3.2.6 MCMC Implementation

As previously mentioned MCMC will be used as the method of model fitting. As this will involve estimating a set of multivariate parameters, Gibbs sampling will be used. Some of the full conditionals turn out to be of standard form, but for the majority a Metropolis-Hastings step is inserted. Before evaluating the full conditionals, a reparameterisation is introduced, and various other methods for improving the efficiency of the Gibbs sampler are discussed.

Improving Efficiency

The ultimate object of the analysis is to provide estimates for:

$$\beta_1, ..., \beta_N,$$

where each β_j $(j = 1, ..., N)$, represents the value of the parameter within I_j (in the current application this is either the baseline or covariate effect parameter). When the parameters are highly correlated Gibbs sampling may run into problems, which may slow convergence; and in some circumstances the chain may fail to converge at all. Suppose that the Markov chain happens to be in a position where one of the parameters is assigned to a poor value. The chain attempts to sample a second parameter, which we suppose is highly correlated with the first. In such circumstances the full conditional will give very little weight to values of the parameter which under the marginal distribution would be highly weighted. The chain will move very slowly, if at all, to the target distribution.

To avoid such problems, one method is to consider a reparameterisation to a set of parameters which are not so highly correlated. For the application considered in this thesis, there turns out to be a very easy and effective reparameterisation. Instead of sampling the paramater effects directly, it is

noted that these effects can be derived entirely from:

$$\beta_1, w_2, ..., w_N.$$

This is true by noting that any paramater β_j $(j = 1, ..., N)$, may be re-written as:

$$\beta_j = \beta_1 + \sum_{j'=2}^{j} w_{j'}.$$

The parameters, w_j, which represent evolution, will not be as highly correlated as the untransformed parameters β_j. We may think of the correlation between two parameters β_j and β_{j+1} as having two parts: the first the correlation which exists due to the baseline of the parameter; and the second due to the increment of the parameter. Using the above reparameterisation the first part of the correlation will be eliminated. For the Dynamic Survival Model experience proved that this parameterisation provided an immense improvement.

Gamerman (1998) compared alternative Gibbs sampling schemes for a related model (the Dynamic Generalised Linear Model). Amongst the schemes considered in that paper, was a reparameterisation (similar to the one described above); multi-move samplers (where the parameters are blocked in non-correlated groups); and also a Metropolis-Hastings step for a set of adjusted full conditionals (Gamerman, 1998), which are always normal. Under

the Metropolis-Hastings step, proposals based on a normal distribution with mean at the current value of the parameter and variance at the current estimate of the evolution variance were used. Gamerman found that the method based on the adjusted full conditionals worked best, but also stressed the importance of the reparameterisation. Unfortunately survival models do not have such nice properties, as generalised linear models, and the full conditionals are far from normal.

It is also observed that in some applications of the piecewise correlated baseline hazard (with no dynamic covariate effect), a reparmeterisation was not used within the Gibbs sampler (Aslanidou *et al.*, 1995), with the chain apparently converging nontheless. In all of the applications considered in this thesis, the parameterisation was essential, and without it the chain very clearly failed to converge.

The reparmeterisation described above does not solve any of the potential problems which may arise due to the high correlation which may exist between the parameter effects and the corresponding evolution variance. This is a problem which could potentially arise in any model with a hyper-prior. There exists what could possibly be a serious problem: the Markov chain may get stuck in a place of very high or very small evolution variance; this

will result in the acceptance of either very smooth or dispersed estimates for the covariate effects over time. Clayton (1991) and Tanner and Wong (1987), suggested that this problem may be improved if the Markov chain is sampled in blocks: all of the uncorrelated parameters are blocked together and sampled as in a usual Gibbs sampler, and after a number of iterations the corresponding correlated parameters (in this instance the evolution variance) are sampled. The method may avoid the high correlation problem, but is very costly in terms of disregarded samples. Raftery and Lewis (1992) discuss this problem, and acknowledging that it can be a very fundamental problem, suggest using simultaneous updating: although it has been shown, by Gamerman (1998), that when used in conjunction with a Metropolis-Hastings step, the method can lead to very low acceptance rates. Experience showed that applying the Gibbs sampling without resorting to sampling in blocks, worked well in the applications required in this thesis.

The Full Conditionals

Using the reparmeterisation described above, the first set of parameters which must be estimated, and which will be referred to as the main paramaters, are:

$$\beta_{01}, w_{02}, ..., w_{0N},$$

and

$$\beta_{k1}, w_{k2}, ..., w_{kn.},$$

for $k = 1, ..., K$. The second set of paramaters are the hyper-paramaters: $W_0, W_1, ..., W_k$.

The full conditionals for the main parameters, reduce to the likelihood multiplied by any initial prior for the parameter of concern. Suppose that $K = 1$, then the full conditional for β_{11} is:

$$\begin{aligned}[\beta_{11}|\cdot] &= p(\beta_{11}|\beta_{01}, ..., \beta_{0N}, w_{12}, ..., w_{1N}, W_0, W_1) \\ &\propto Lp(\beta_{11}).\end{aligned}$$

When using the temporal factorisation, the full conditionals for the main parameters reduce to functions of the data from that particular interval only, contributions from other intervals are constants of proportionality:

$$\begin{aligned}[\beta_{11}|\cdot] &\propto \prod_{j=1}^{N} L_j p(\beta_{11}) \\ &\propto L_1 p(\beta_{11}),\end{aligned}$$

where L_j was derived at equation 7 (page 72). Under the individual factorisation, each part of the factorisation could potentially contribute to each of the full conditionals. When using a Gibbs sampler, at each separate iteration, the full conditionals must be re-calculated. Even apparently small gains in efficiency, as exists between the individual and temporal factorisation, can

lead to considerable decreases in computational time. This is why the temporal factorisation of the likelihood has been labelled here as the more efficient factorisation. It is therefore the factorisation which will be used to compute the full conditionals under the right censored model.

The full conditionals derived here are all based on the standard model outlined in definition 3.2. Aslanidou and Dey (1996) computed some similar full conditionals, for a model which although models proportional hazards, is not so dissimilar to the model being developed here.

It has been mentioned that there exist variations on this model, in terms of the parametric distributions assumed in modelling the evolution. For alternative distributions similar forms of the full conditionals may be derived; and whenever the full conditional is not of standard form then a Metropolis-Hastings step can be inserted (section 3.2.6). The full conditionals are:

- For β_{01}, the baseline effect parameter:

$$[\beta_{01}|\cdot] \propto L_1 p(\beta_{01})$$

$$\propto L_1 \exp\Big(-\frac{(\beta_{01}-\beta_{00})^2}{2C_0}\Big),$$

where L_j $(j = 1, ..., N)$ was defined at equation 7 (page 72).

- For the evolution of the baseline effect:

$$[w_{0j}|\cdot] \propto L_j p(w_{0j}) \text{ for } j = 2, ..., N$$

$p(w_{0j})$ was defined within the model definition page 66, so that:

$$\propto L_j \exp\Big(\frac{-w_{0j}^2}{2W_0}\Big).$$

- For the evolution variance of the log-baseline effect:

$$\log[W_0|\cdot] \propto \sum_{j=2}^{N} \log p(w_{0j}|W_0) + \log p(W_0)$$

$$\propto \Big(\frac{N-1}{2} + \zeta_0 - 1\Big) \log(W_0^{-1}) - W_0^{-1}\Big(\sum_{j=2}^{N} \frac{w_{0j}^2}{2} + \eta_0\Big),$$

This is proportional to a gamma density and so:

$$W_0^{-1} \sim G(\zeta_0', \eta_0') \tag{11}$$

where

$$\zeta_0' = \zeta_0 + \frac{N-1}{2} \text{ and } \eta_0' = \eta_0 + \sum_{j=2}^{N} \frac{w_{0j}^2}{2}.$$

- For the parameter β_{k1}, for $k = 1, ..., K$:

$$[\beta_{k1}|\cdot] \propto L_1 p(\beta_{k1})$$

$$\propto L_1 \exp\Big(\frac{-(\beta_{k1} - m_k)^2}{2C_k}\Big).$$

- For the evolution of the covariate effect parameters, for $k = 1, ..., K$:

$$[w_{kj}|\cdot] \propto L_j p(w_{kj}) \text{ for } j = 2, ..., N$$

$$\propto L_j \exp\Big(-\frac{w_{kj}^2}{2W_k}\Big).$$

- For the evolution variance of the covariate effects, again $k = 1, ..., K$:

$$\log[W_k|\cdot] \propto \sum_{j=2}^{N} \log p(w_{kj}|W_k) + \log p(W_k)$$

$$\propto \Big(\frac{N-1}{2} + \zeta_k - 1\Big)\log(W_k^{-1}) - W_k^{-1}\Big(\sum_{j=2}^{N}\frac{w_{kj}^2}{2} + \eta_k\Big),$$

so that:

$$W_k^{-1} \sim G(\zeta_k', \eta_k'), \tag{12}$$

where

$$\zeta_k' = \zeta_k + \frac{N-1}{2} \text{ and } \eta_k' = \eta_k + \sum_{j=2}^{N}\frac{w_{kj}^2}{2}.$$

The Metropolis-Hasting Step

With the parameterisations and distributions specified it is necessary to insert a Metropolis-Hastings step to sample from some of the full conditionals. Within this application, normal proposals were chosen, with mean set to the current value of the parameter and variance chosen to ensure acceptance rates of around 50 percent (this avoids the danger of accepting all values and of accepting almost none, both of which can be indicative of lack of convergence).

3.2.7 Convergence Diagnostics

After each set of simulations a range of convergence diagnostics were carried out, some very intuitive, and others more formal. Within most of the data applications non-parametric techniques were investigated prior to fitting the NDSM, so as the resulting estimates could be compared with the more standard techniques. Then an initial Gibbs sampler was performed, based on a small number of iterations (say 1000); and then the Raftery and Lewis (1992) diagnostic carried out in CODA (Best *et al.*, 1997), obtaining an initial estimate of how many iterations would be required and how many should be disregarded as "burn in". This statistic typically returned a very low burn in number and a moderate number of iterations (usually less than 10,000). To be cautious the burn in was increased to at least one tenth of the total number of iterations. Finally simulated values were plotted and convergence visually inspected. Occasionally when the chain did not converge (perhaps because of coding error), this method proved to be effective.

Although results quoted are based on a single run of the Gibbs sampler, within the investigation and development of the models, many more runs were considered (from a variety of starting points) and it was checked that all converged to the same posterior.

3.3 Gastric Cancer Data Analysis

Gamerman (1987) used the Dynamic Survival Model to estimate the survival for a group of gastric cancer patients (the data set was initially analysed by Carter *et al.* (1983), and was introduced in this thesis in section 2.6). This data set will also be modelled by the Normal Dynamic Survival Model, firstly for consistency and comparison with Gamerman; and also because the data set illustrates very nicely the advantages of the dynamic modelling technique.

To compare the effect of radiotherapy and the combined treatment of radiotherapy and chemotherapy, in the treatment of gastric cancer, 90 patients were randomly allocated into two equal sized treatment groups, in this controlled clinical trial. Within the analysis in this thesis, the covariate value 0 is assigned to the treatment group of chemotherapy and 1 to the combined treatment of chemotherapy and radiation. The data set and Kaplan-Meier curve may be found in the appendix (table A.1 and figure 27, page 192), with the data being recorded in days. Survival estimated under the proportional hazards assumption using Cox's method (figure 28, page 193) is also presented as it shows the extent of the incorrect inferences which would be drawn from fitting a proportional hazards model. In this particular data set the incorrect conclusion of no treatment difference would be drawn from Cox's model. The Kaplan-Meier curve indicates a clear lack of propor-

tionality, as does the Grambsch and Therneau (1994) test of proportionality illustrated in figure 29 (page 194), which has a chi-squared value of 11.1 on 1 degree of fredom. The plot produced by the Grambsch and Therneau (1994) function (cox.zph) in Splus, shows the estimated treatment effect to decrease over time. From the Kaplan-Meier curve the indication is that survival under chemotherapy is better up until 1000 days, when the treatment curves cross and the combined treatment appears to become more effective. From a medical point of view there does not appear to be a single superior treatment.

In Chapter 2 it was described how Carter *et al.* (1983) fitted Cox's proportional hazards model to the data, but included two covariates: the first being the treatment indicator and the second the treatment indicator multiplied by time (allowing for a linear change in the treatment effect). Using the linear Bayes approximation and the Dynamic Survival Model (section 3.1), Gamerman (1987) came to similar conclusions to the analysis of Carter *et al.* (1983), showing that the treatment effect varied over time. Within the analysis, Gamerman (1987) compared the log-likelihoods for three models: the first being the exponential proportional hazards model (with Chemotherapy as the baseline parameter); the second, an exponential baseline with dynamic treatment effect; and the third a non-exponential, proportional hazards model. He observed little difference in the log-likelihoods between the

first and third model, concluding that the baseline was exponential. He further observed a difference between the log-likelihoods under models 1 and 2, and hence modelled the data via a non-proportional hazards model with exponential baseline. After deciding on the model, and fixing the evolution variances, various divisions of the time axes were compared. The time axis which performed best in terms of log-likelihoods, were a division based on 30 points: $(20, 40, .., 200, 250, 300, ..., 600, 700, ..., 1800)$; and a division based on the observed death times.

Analysis Using the NDSM

Vague priors were used for all parameters. For the treatment effect the chosen prior was $N[0, 1000]$; for the log-baseline hazard $N[-3, 1000]$; and for the evolution variances $G(0.001, 0.001)$.

So as to minimise the computational time involved within the estimation of the paramaters, it was desirable to keep the number of intervals to a minimum. As Gamerman (1987) found that two divisions worked particularly well, these divisions were compared with a shorter one: (100,200,...,1800). The divisions will be labelled division 1 to 3, from shortest to largest. The observed log-likelihoods for the three divisions after 100,000 iterations are presented in table 1, and a comparision of the numerical output under all

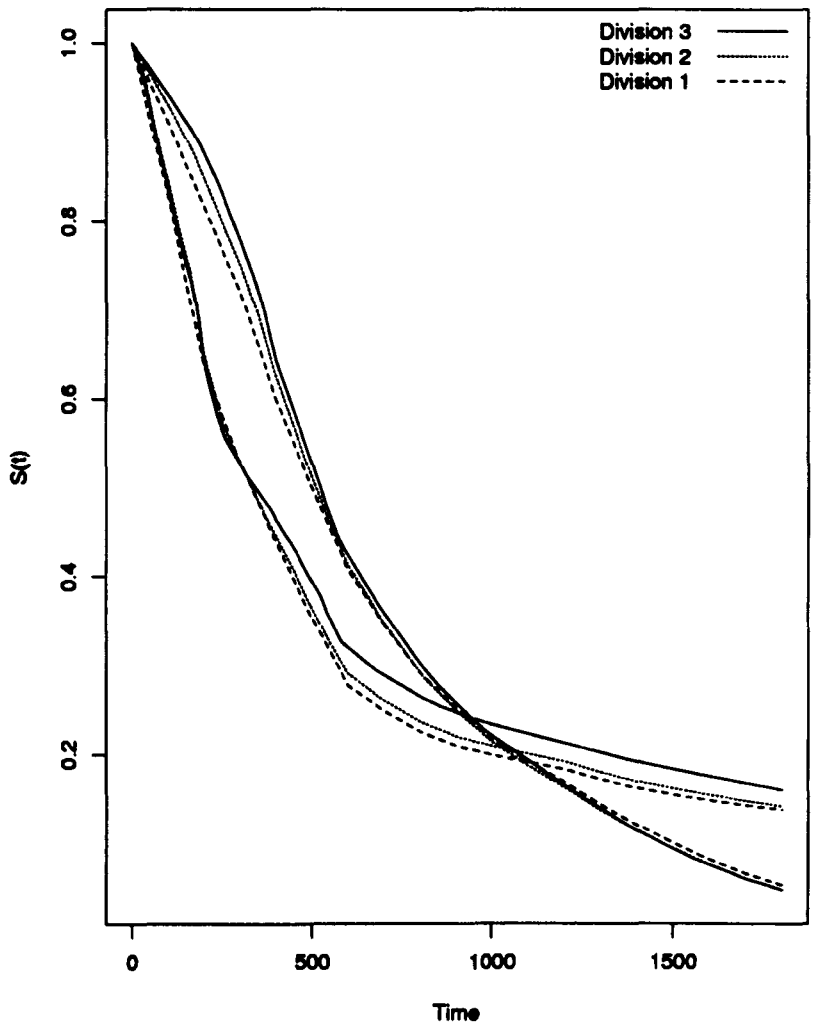

Figure 1: Gastric Cancer: A comparision of the estimated survival using 3 divisions of the time axis. Survival under chemotherapy is better than under radiation, up until 1000 days when the effect is reversed.

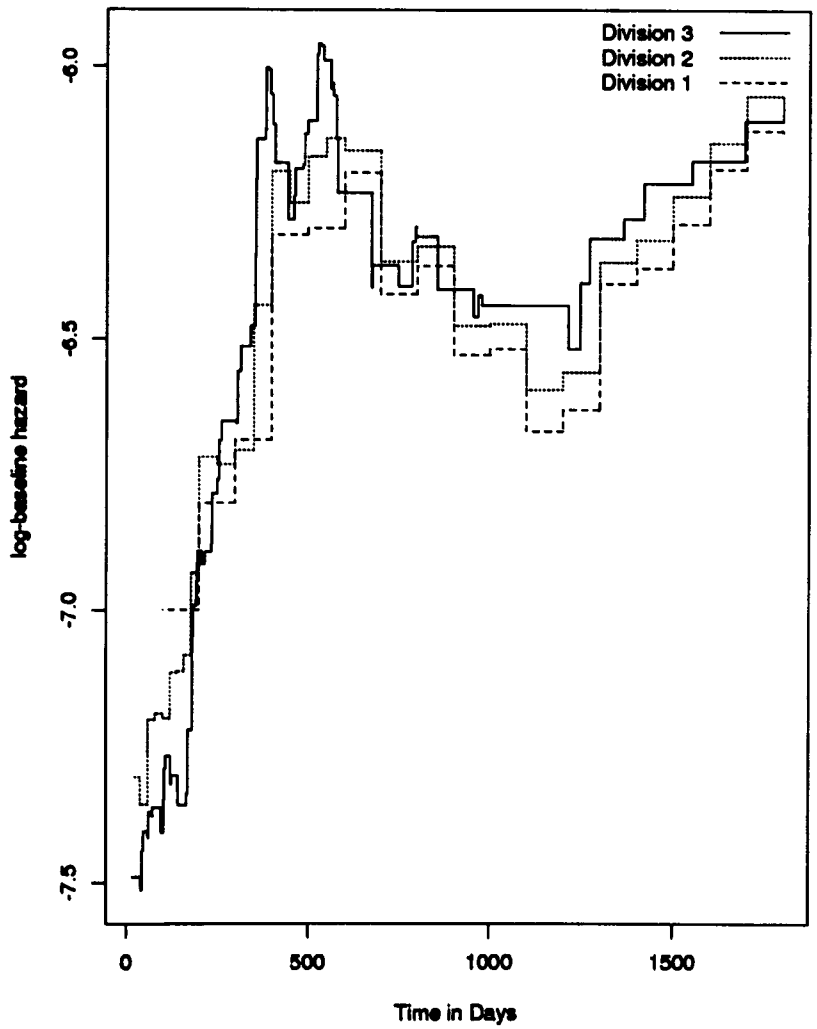

Figure 2: Gastric Cancer: A comparision of the estimated log-baseline hazard using 3 divisions of the time axis. The estimated log-baseline hazard decreases rapidly over the first 500 days, and then remains quite steady.

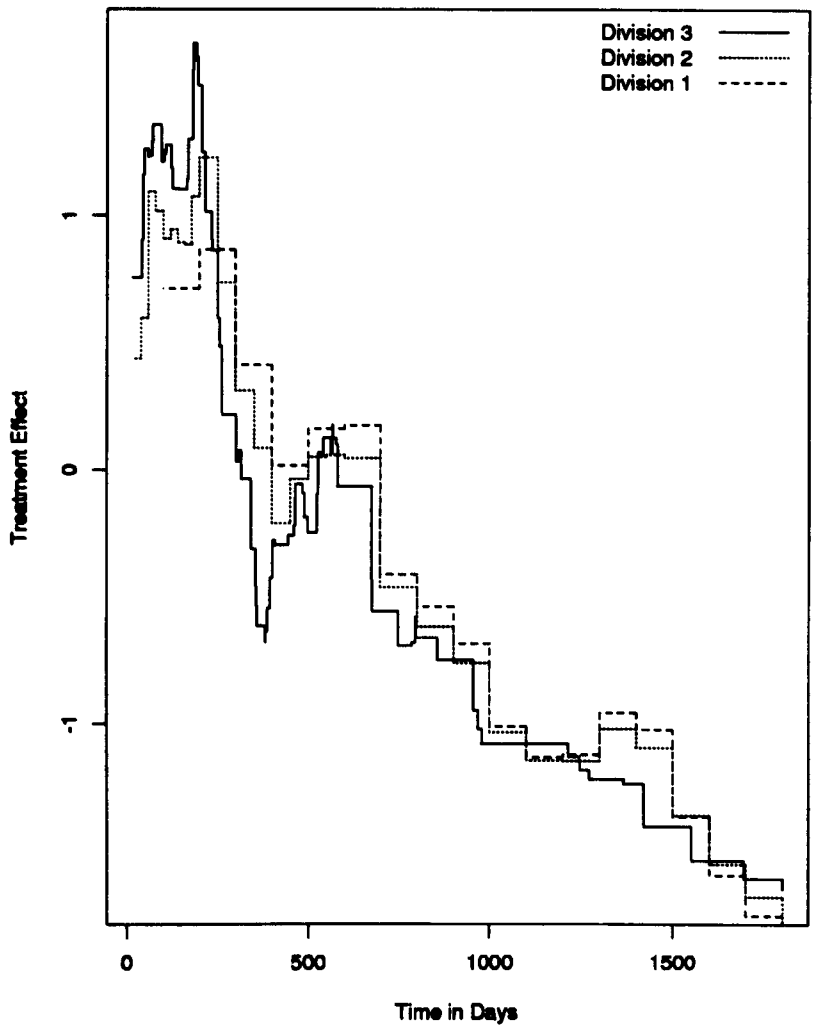

Figure 3: Gastric Cancer: A comparision of the estimated treatment effect using 3 divisions of the time axis. The treatment effect decreases over time. Chemotherapy is initially the most benificial treatment, but after 1000 days the combined treatment takes over.

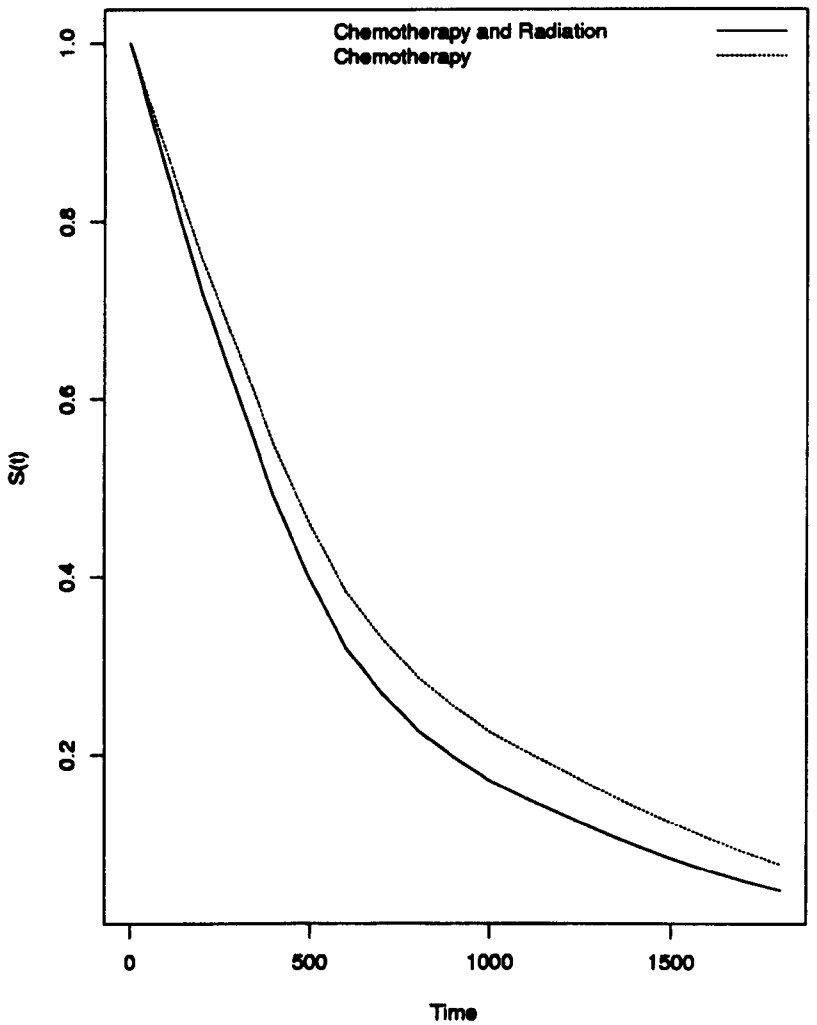

Figure 4: Gastric Cancer: Estimated survival using the Normal Dynamic Survival Model, constrained to return proportional hazards.

Division	log-likelihood
Division 1	-585.23
Division 2	-581.08
Division 3	-581.00

Table 1: Gastric Cancer: Estimated log-likelihoods under various different divisions of the time axis. Division one refers to the shortest division and division three to the longest (see text for exact definitions).

models is presented in table 6 (page 196). The longer time axis divisions appear to provide a slightly better fit, in terms of a smaller log-likelihood, although the difference between survival and covariate estimates was slight (figures 1 to 3). The shorter division was used for the rest of the analysis, as the savings in computational time far outweigh the gains in the reduction of the log-likelihoods.

Again to keep computational time to a minimum, it was desirable to carry out the minimum number of iterations necessary, whilst of course insuring convergence. For the gastric cancer data analysis, very good estimates for the parameters already exist (Gamerman, 1987). The similarity between estimates obtained after just 1000 or even 500 simulations was remarkable. The results obtained after carrying out 1,000, 10,000 and 100,000 iterations

Number of Iterations	log-likelihood
1,000	-587.90
10,000	-585.67
100,000	-585.23

Table 2: Gastric Cancer: Estimated log-likelihoods under different numbers of iterations carried out in the Gibbs sampler. All have a ten percent burn in.

are compared (disregarding the first 1,000 (or first 100) values, and saving every subsequent sampled value). The numerical results obtained under all sets of simulations were extremely similar, and this is reflected in comparisons between the log-likelihoods (table 2).

The estimated survival curve (comparing the three time axis divisions) is given in figure 1, the hazards in figure 2, and the treatment effect in figure 3. Full numerical results are given in the appendix (table 5, page 195). The estimated treatment parameter compares very well with that estimated by Gamerman (1987). Final estimates for the evolution variances were 0.155345 and 0.463127 (variances 0.02 and 0.45) for the baseline and treatment effect respectively. Gamerman (1987) argued that evolution variances other than zero for the baseline performed unfavourably in terms of their Bayes factors;

zero variance for the baseline is within two standard deviations of the mean, although it appears, from the plot of the baseline hazard (figure 2), that there is a sharp decrease in this parameter over early days. The simulated values of a selection of parameters, plotted against the iteration number are presented in figure 30 (page 197). The Raftery and Lewis (1992) diagnostic typically recommend that between 3000 to 7000 iterations should be carried out, although visual inspection after 1000 iterations showed that the chain appeared to have converged.

The results obtained by constraining the NDSM to a proportional hazards model were compared with the estimates obtained using Cox's proportional hazards model in Splus. In Splus the treatment effect is estimated to be 0.164 (standard error 0.225), and the NDSM (constrained to return proportional hazards) estimated the treatment effect to be 0.172 (standard error 0.223) . The plot of the estimated survival curve, using the NDSM, may be found in figure 4.

3.4 Summary

Within this Chapter the model developed by Gamerman (1987), has been adapted slightly, and model parameters estimated using the method of MCMC. Care was taken to ensure convergence of the Gibbs sampler: a reparameteri-

sation was used, and various converegence diagnostics employed. The model developed was subsequently applied to the same set of gastric cancer survival times as analysised in Gamerman (1987), with convincing results.

4 Flexible Dynamic Survival Models

The focus of this Chapter will be on extending the Normal Dynamic Survival Model (Chapter 3, definition 3.2), to accommodate a wider range of data types. The types of data which will be incorporated, are interval censoring, double interval censoring, and non-independent observations (frailties). The need for non-proportional hazard models has been clearly demonstrated throughout this thesis. So far however much of this discussion has concentrated on right censored observations only, although there is no reason why all of this should not equally apply to the data types which will be considered in this Chapter. In Chapter 2 existing methods for modelling data more complicated than right censoring were discussed. Some of these models were non-parametric, some semi-parametric, and some parametric. None of the models however were capable of accommodating a semi-parametric but non-proportional hazards model. By extending the Normal Dynamic Survival Model, a method will exist whereby such data types may be modelled using multivariate techniques, whilst not being constrained by the proportional hazards assumption. An additional advantage of a model which is not constrained by proportional hazards, is that the model itself could be used as a test for proportional hazards (constant covariate effects would be indicative of proportional hazards).

The piecewise correlated process is not a new concept for modelling the baseline hazard in such types of data applications (Ghosh and Sinha, 1995), but taking the process one step further, and modelling one or more dynamic covariate effects is. In a frailty model, using a piecewise correlated baseline hazard function, Sinha (1998) discussed extending the model to include dynamic covariate effects: but commented that it would be unreasonable to maximise the likelihood with so many parameters in their data application (90 observations). Using MCMC will avoid the maximisation problem (although care should still be taken when estimating so many parameters and convergence should be checked). Furthermore the approach worked well for the right censored model, and it seems that this should not be much different for more complicated censoring types. Again any potential high correlation problems will be reduced by using a reparametisation.

It was demonstrated in Chapter 3, section 3.2.6, that the temporal factorisation is more efficient when it comes to Gibbs sampling than the individual factorisation is. Unfortunately it is not possible to create a temporal factorisation for interval censored and double interval censored data, a point already discussed in section 3.2. The main reason for extending the Dynamic Survival Model, by using MCMC and parametric modelling, was after all the inability to use the linear Bayes method within data sets where it was

not possible to identify a rank ordering: where it is not possible to identify a rank ordering, it is also not possible to create a temporal factorisation.

To accommodate numerous types of censoring, either all of the observations should be written as double and interval censored, or the likelihood should be factorised over contributions from all of the different types of censoring; it will be shown within each section how to do this. In this Chapter the emphasis will be on introducing the models and deriving the likelihoods, and various methods of model fitting techniques (including MCMC and missing data methods) will be explored in the following Chapter.

4.1 Conditional Survival Functions (3)

Consider the probability of failure in the interval $(R_i, L_i]$, where $R_i \in I_{r_{i+1}}$ and $L_i \in I_{l_{i+1}}$:

$$p(R_i < T \leq L_i) \text{ for } R_i \in I_{r_{i+1}}, L_i \in I_{l_{i+1}}.$$

The probability is evaluated using the notation and conditional survival functions which were developed in Chapter 3 (equation 9, page 74).

$$\begin{aligned} p(R_i < T \leq L_i) &= p(T > R_i) - p(T > L_i) \\ &= \exp(-\sum_{j=1}^{r_i} e^{x_i\beta_j}(t_j - t_{j-1}))\exp(-e^{x_i\beta_{r_i+1}}(R_i - t_{r_i})) \\ &- \exp(-\sum_{j=1}^{l_i} e^{x_i\beta_j}(t_j - t_{j-1}))\exp(-e^{x_i\beta_{l_i+1}}(L_i - t_{l_i})) \\ &= A(1, r_i, \beta)C(R_i, t_{r_i}, \beta) - A(1, l_i, \beta)C(L_i, t_{l_i}, \beta). \end{aligned} \tag{13}$$

4.2 Interval Censoring

Interval censoring was introduced in Chapter 2, section 2.8 of this thesis. To recap, interval censored observations take the form $(R, L]$, where it is observed only that failure occurred within this censoring interval, but not precisely when.

When considering data with interval censoring, right censoring, and exact data, option one is to consider contributions from all right censored observations; followed by contributions from all exact observations; and then finally from all interval censored observations:

$$L = \prod_{i=1}^{n_1} S(t_i) \prod_{i=1}^{n_2} f(t_i) \prod_{i=1}^{n_3} p(R_i < T \leq L_i),$$

where n_1 and n_2 are the number of right censored and exact observations respectively, and n_3 is the number of interval censored observations. Alternatively all observations could be written as interval censored. If an observation is observed right censored at R_i, then it may be thought of as interval censored at $(R_i, \infty]$. Similarly an exact observation R_i, may be thought of as interval censored at $(R_i, R_i]$.

Since the Normal Dynamic Survival Model is being used to model the data, a division of the time axis must be created. As under the right censored model, the division of the time axis for a set of interval censored data will be denoted by G_T, and points in G_T will be indexed by t_j, for $j = 1, ..., N_T$.

4.2.1 The Likelihood

For each observation $(R_i, L_i]$, it is necessary to find points on the time axis t_{r_i} and t_{l_i}, such that:

$$t_{r_i} < R_i \leq t_{r_{i+1}},$$

and

$$t_{l_i} < L_i \leq t_{l_{i+1}}.$$

Assuming that all observations are interval censored then the likelihood taken as a factorisation over all observations is:

$$\begin{aligned} L(T) &= \prod_{i=1}^{n_T} p(R_i < T \leq L_i) \\ &= \prod_{i=1}^{n_T} \{p(T > R_i) - p(T > L_i)\}. \end{aligned}$$

Using the interval probability developed in equation 13 (page 100):

$$L = \prod_{i=1}^{n_T} \{A(1, r_i, \beta)C(R_i, t_{r_i}, \beta) - A(1, l_i, \beta)C(L_i, t_{l_i}, \beta)\}. \qquad (14)$$

This likelihood is not written out in full, as it would look too cumbersome, but the analytical forms of the relevant functions are easily extracted from equation 13.

Ghosh and Sinha (1995) developed a likelihood for interval censored data, under the piecewise correlated baseline hazard. There exist several differences between the likelihoods developed under their model and the one developed here: firstly their likelihood construction is based on grouped data; and secondly the model developed here includes dynamic covariate effects.

4.3 Double Interval Censoring

Where data are double interval censored, then both the initiating event time and terminating event time, are observed to lie within an interval. The intervals, for the calendar time of the occurrence of the initiating and terminating

event, are defined as $(M, P]$ and $(R, L]$ respectively. S is defined to be the calendar time of the initiating event, X the calendar time of the terminating event, and T the survival time, or the difference between the initiating and terminating event, which will usually be referred to in this thesis as the incubation period.

Both S and T are modelled using the Normal Dynamic Survival distributions, using notation previously introduced: $S \sim NDSM(\lambda, N_S, V)$ and $T \sim NDSM(\beta, N_T, W)$. For double interval censoring, $G_S = \{s_1, ..., s_{N_S}\}$ will denote the time axis division for the initiating event, with the subscript s_j referring to points in this division, and intervals within this division will be referred to by $\bar{I}_j$. Similarly $G_T = \{t_1, ..., t_{N_T}\}$ will denote the time axis division for the incubation or survival time, and t_j will once again refer to points within this division, with I_j referring to intervals within the time axis division G_T. For the data sets used in this application some of the initiating event times will be right censored, resulting in possibly fewer observations, n_T, of terminating event times, than initiating event times n_S. Because the nature of a double interval censored model is slightly more complicated than the other censoring types considered, a brief outline of the main points involved within any analysis of this type is given.

Summary

- The object is to make inferences about $T = X - S$. These inferences will depend on the model parameters for S, so firstly inferences must be made about S.

- As each initiating event is observed up to an interval only, making inferences for S involves estimating the survival of an interval censored data set. Using the likelihood for interval censored data developed in the previous section, estimates are obtained for the parameters of the distribution for S.

- Using inferences about the distribution of S across the interval $(M, P]$, it will be possible to make inferences about the incubation distribution for T.

Where the data set contains several types of censoring, the likelihood should be factorised over all censoring types (as described for the interval censored approach in section 4.2). Letting n_1 represent the number of right censored observations $(t_i, \delta_i = 0)$; n_2 the number of exact observations $(t_i, \delta_i = 1)$; n_3 the number of interval censored observations $((R_i - s_i, L_i - s_i])$; and n_4 the

number of double interval censored observations ($(M_i, P_i]$ and $(R_i, L_i]$):

$$L = \prod_{i=1}^{n_1} S(t_i) \prod_{i=1}^{n_2} f(t_i) \times \prod_{i=1}^{n_3} p(R_i - S < T \leq L_i - S | S = s_i) \times \prod_{i=1}^{n_4} p(R_i - S < T \leq L_i - S | s_i \in (M_i, P_i]).$$

4.3.1 The Initiating Event

The likelihood for the incubation times T, in a double interval censored model, as previously stated, will depend on the parameters in the model for S. The first step is to construct the likelihood for the initiating event times S, and estimate all parameters within the initiating model. Modelling the initiating event consists of modelling an interval censored data set. The construction of the likelihood is therefore exactly as in section 4.2.1, although it is redefined here so as to highlight the differences in the notation when the interval censored data set consists of initiating observations. Firstly it is identified within which interval each observation lies:

$$M_i \in \bar{I}_{m_i+1}$$

$$P_i \in \bar{I}_{p_i+1}.$$

Using this notation, and the likelihood for interval censored data as was defined in equation 14 (page 102), the likelihood for the initiating event is:

$$L(S) = \prod_{i=1}^{n_S} \{A(1, m_i, \lambda) C(M_i, s_{m_i}, \lambda) - A(1, p_i, \lambda) C(P_i, s_{p_i}, \lambda)\}. \qquad (15)$$

4.3.2 The Terminating Event

The initiating time and incubation period, S and T, are both being modelled by continuous distributions. The correct likelihood for T is therefore obtained by integrating over all possible values for the initiating event:

$$L(T) = \prod_{i=1}^{n_T} \int_{s \in (M_i, P_i]} f(s|M_i < S \leq P_i) p(R_i - s < T \leq L_i - s) ds.$$

Because the model is dynamic, this integration is written as a summation of integrals across the time axis:

$$L(T) = \prod_{i=1}^{n_T} \int_{s \in (M_i, t_{m_i+1}]} g(s) ds + \sum_{j=m_i+1}^{p_i-1} \int_{s \in \bar{I}_{j+1}} g(s) ds + \int_{s \in (t_{p_i}, P_i]} g(s) ds,$$

where

$$g(s) = f(s|M_i < S \leq P_i) p(R_i - s < T \leq L_i - s).$$

For $s \in (M_i, P_i]$, and if $s \in \bar{I}_{j+1}$, then it is straightforward to evaluate the first part of this function. First of all it is noted that:

$$f(s|M_i < S \leq P_i) = \frac{f(s)}{p(M_i < S \leq P_i)}.$$

Using the conditional survival functions which were developed in this and earlier Chapters (equation 9, page 74):

$$f(s) = A(1, j, \lambda) C(s, s_j, \lambda) e^{x_i \lambda_{j+1}},$$

and (derived in equation 13, page 100):

$$p(M_i < S \leq P_i) = A(1, m_i, \lambda) C(M_i, s_{m_i}, \lambda) - A(1, p_i, \lambda) C(P_i, s_{p_i}, \lambda).$$

So that:

$$f(s|\cdot) = \frac{A(1,j,\lambda)C(s,s_j,\lambda)e^{x_i\lambda_{j+1}}}{A(1,m_i,\lambda)C(M_i,s_{m_i},\lambda) - A(1,p_i,\lambda)C(P_i,s_{p_i},\lambda)}. \tag{16}$$

The next part of the integral to evaluate, is $p(R_i - s < T \leq L_i - s)$. Suppose that it can be identified that $R_i - s \in I_{r_{i,s}+1}$ and $L_i - s \in I_{l_{i,s}+1}$, then this probability is easily extracted from equation 13 (page 100):

$$\begin{aligned} p(R_i - s < T \leq L_i - s) &= \{A(1,r_{i,s},\beta)C(R_i - s, t_{r_{i,s}},\beta) \\ &\quad -A(1,l_{i,s},\beta)C(L_i - s, t_{r_{i,s}},\beta)\}. \end{aligned} \tag{17}$$

Considering the integration over one interval, say $\bar{I}_{j+1}$:

$$\int_{s\in\bar{I}_{j+1}} g(s)ds,$$

for a given $r_{i,s}$ and $l_{i,s}$, this will clearly reduce to the integral of an exponential (the multiple of equation 16 with equation 17). The integration is clearly tractable, even if slightly complicated. However $r_{i,s}$ and $l_{i,s}$ are not known, and must be calculated for each value of s. This complicates the integration considerably: the integral goes from being the integral of an exponential, to an integral of several complicated functions. To avoid this difficulty an approximation has been developed.

In this approximation, it is assumed that the initiating event, S, could have only occurred at a finite number of points. It will become clear as the likelihood is developed why it will then be possible to compute the likelihood,

where it was not possible in the exact case. A set G is created, which spans the time axis, and some new notation is introduced:

$$G(M,P) = G \cap (M,P].$$

The set G must be at least as fine as G_S, and the finer the set G, the more accurate the approximation will be. However at the same time the computational intensity of the model will increase, as $G(M,P)$ becomes increasingly continuous. Choosing the set G will be discussed within a practical application of the model in Chapter 6, section 6.3.

The likelihood is once again taken as the product over all observations, with each observation consisting of contributions from each possible s within the set $G(M_i,P_i)$:

$$L(T) = \prod_{i=1}^{n_T} \sum_{s \in G(M_i,P_i)} p(S=s|M_i < S \leq P_i)p(R_i - s < T \leq L_i - s).$$

Considering the summation over $G(M_i,P_i)$ as summations over the separate intervals from the time axis:

$$\begin{aligned} L(T) = \prod_{i=1}^{n_T} & \sum_{s \in G(M_i, s_{m_{i+1}})} g(s, R_i, L_i, \bar{I}_{m_{i+1}}) + \\ & \sum_{j=m_i+1}^{p_i-1} \sum_{s \in G(\bar{I}_{j+1})} g(s, R_i, L_i, \bar{I}_{j+1}) + \\ & \sum_{s \in G(s_{p_i}, P_i)} g(s, R_i, L_i, \bar{I}_{p_{i+1}}), \end{aligned} \tag{18}$$

where

$$g(s, R_i, L_i, \bar{I}_{j+1}) = p(S = s | M_i < S \leq P_i) p(R_i - s < T \leq L_i - s) \text{ for } s \in \bar{I}_{j+1}.$$

To evaluate this probability, firstly it is necessary to evaluate:

$$p(S = s | M_i < S \leq P_i) = \frac{f(s | M_i < S \leq P_i)}{n'},$$

where n' represents the number of points within the set $G(M, P)$. The conditional density function $f(s|M_i < S \leq P_i)$, was developed in equation 16 (page 107), so that:

$$p(S = s|\cdot) = \frac{A(1, j, \lambda) C(s, s_j, \lambda) e^{x_i \lambda_{j+1}}}{n' \{A(1, m_i, \lambda) C(M_i, s_{m_i}, \lambda) - A(1, p_i, \lambda) C(P_i, s_{p_i}, \lambda)\}}. \quad (19)$$

Next it is necessary to evaluate $p(R_i - s < T \leq L_i - s)$ for $s \in \bar{I}_{j+1}$. To evaluate this probability, it must be known within which intervals on the times axis, G_T, these two observations fall. Using similar notation to that which has already been used:

$$t_{r_{i,s}} < R_i - s \leq t_{r_{i,s}+1},$$

and

$$t_{l_{i,s}} < L_i - s \leq t_{l_{i,s}+1}.$$

Because s only takes a finite number of values, for each s it is possible to evaluate the above (it was at this point that the exact version of the

likelihood became too complex). For a given s this probability reduces to a regular interval censored data evaluation:

$$\begin{aligned} p(R_i - s < T \leq L_i - s) &= \{A(1, r_{i,s}, \beta)C(R_i - s, s_{r_{i,s}}, \beta) \\ &\quad -A(1, l_{i,s}, \beta)C(L_i - s, s_{r_{i,s}}, \beta)\}. \end{aligned} \tag{20}$$

It has been established that $g(s, R_i, L_i, I_{j+1})$ is the product of equation 19 with equation 20. Substituting $p(\cdot)$ into the likelihood developed at equation 18, produces the full form for the approximated likelihood for the incubation period.

It is noted that exact inferences for double interval censored data have been obtained (Kim *et al.*, 1993), under parametric assumptions. The difficulty of the model developed within this thesis stems from the model being based on a division of the time axis.

4.4 Frailty

There have been many applications of various forms of the frailty model in the statistical literature (Chapter 2, section 2.10). Most of these approaches have focused on the conditional proportional hazards. Some of the existing frailty models do however use the piecewise correlated baseline hazard function, although as in the case of the other data types, all model the baseline only by this dynamic process, and not the covariate effects. As previously

mentioned, using a posterior likelihood approach, Sinha (1998) discussed extending the conditional proportional hazards model to one with dynamic covariate effects (section 4). A natural extension of current frailty models, and the models developed in this thesis, is a conditional Normal Dynamic Survival Model. The piecewise correlated conditional proportional hazards model does have some similar features to the those models developed here, and so naturally the likelihoods developed within this section will have some similarities to those developed for example in Sinha (1998).

4.4.1 The Normal Dynamic Frailty Model (NDFM)

Definition 4.1 *The Normal Dynamic Frailty Model has a hazard at time t, for an observation with covariate vector z, in group g, defined by:*

$$h(t|z,u_g) = \exp(\beta_0(t) + \sum_{k=1}^{K} z_k\beta_k(t))u_g,$$

where

$$\beta_0(t) = \beta_{0j} \text{ for } t \in I_j,$$

$$\beta_k(t) = \beta_{kj} \text{ for } t \in I_j,$$

for $k = 1,...,K$. *Evolution of the parameters is defined by the evolution equations:*

$$\beta_{0j} = \beta_{0j-1} + w_{0j} \qquad \text{where } w_{0j} \sim N[0, W_0],$$

$$\beta_{kj} = \beta_{kj-1} + w_{kj} \qquad \text{where } w_{kj} \sim N[0, W_k].$$

Hyper-priors for the evolution variances are:

$$W_0^{-1} \sim G(\zeta_0, \eta_0) \text{ and } W_k^{-1} \sim G(\zeta_k, \eta_k).$$

The frailties u_g are distributed as gamma random variables with unit mean (this ensures identifiability) and variance μ^{-1}:

$$u_g \sim G(\mu, \mu).$$

The hyper-prior for the precision μ is:

$$\mu \sim G(\kappa, \nu).$$

For ease of computation, the whole term within the exponential part, is re-written in terms of a covariate vector multiplied by a covariate effect vector.

$$h(t|z, u_g) = \exp(z\beta(t))u_g,$$

where the notation $z\beta(t)$ was introduced in Chapter 3, section 3.1.

The Marginal Distribution

The hazard has been defined in terms of the conditional hazard (conditional on the frailties). As the frailty will generally be unknown, interest often focuses on the marginal distribution. The marginal distribution function is obtained by integrating out the frailty from the joint distribution $f(t, \mu_g)$. Thus for $t \in I_{m+1}$:

$$f(t) = \int_0^\infty f(t|u_g)p(u_g)du_g,$$

where

$$p(u_g) = \frac{\mu^\mu u_g^{\mu-1} e^{-\mu u_g}}{\Gamma(\mu)},$$

and (adapted from equation 9, page 74):

$$\begin{aligned} f(t|u_g) &= S(t|u_g)h(t|u_g) \\ &= \exp(-\sum_{j=1}^{m} e^{x\beta_j}(t_j - t_{j-1})u_g) \exp(-e^{x\beta_{m+1}}(t - t_m)u_g) e^{x\beta_{m+1}} u_g, \end{aligned}$$

which is re-written as:

$$f(t|u_g) = e^{-u_g B(t,m)} e^{x\beta_{m+1}} u_g,$$

where:

$$B(t,m) = \sum_{j=1}^{m} e^{x\beta_j}(t_j - t_{j-1}) + e^{x\beta_{m+1}}(t - t_m). \tag{21}$$

So that:

$$f(t) = \frac{\mu^\mu e^{x\beta_{m+1}}}{\Gamma(\mu)} \int_0^\infty u_g^\mu e^{-u_g(\mu + B(t,m))} du_g.$$

The integration is proportional to the integral of a gamma density, and so by finding the appropriate constant:

$$f(t) = \frac{e^{x\beta_{m+1}} \mu^{\mu+1}}{(\mu + B(t,m))^{\mu+1}}.$$

Within most applications, interest often centres around the survival or hazard function. It is therefore necessary to first calculate the distribution function $F(t)$:

$$F(t) = \int_0^t \frac{e^{x\beta_{m+1}} \mu^{\mu+1}}{(\mu + B(x,m))^{\mu+1}} dx.$$

Noting that $\frac{d}{dt}B(x,m) = e^{x\beta_{m+1}}$:

$$F(t) = 1 - \mu^{\mu}(\mu + B(t,m))^{-\mu}.$$

Using relations between the survival and hazard functions ($h(t) = \frac{f(t)}{S(t)}$):

$$h(t) = \frac{\mu e^{x\beta_{m+1}}}{\mu + B(t,m)}.$$

The gamma distribution was chosen to model the frailties primarily because of its conjugacy (enabling the marginal distributions to be computed straightforwardly). Other distributions could have been chosen, and a discussion of some other common distributions was given in section 2.10. It is noted however that choosing a distribution for the frailties which was not conjugate to the conditional density function, would result in further computational difficulties. These difficulties could perhaps be overcome by using Monte Carlo simulations, but this of course would increase the computation involved, in what is already a computational intensive model.

At this point there exist several possible options of how to fit the Normal Dynamic Survival Model, to what essentially is a right censored data set. Firstly we might consider carrying out a linear Bayes approximation, treating the frailty groups as dummy variables. However it would then not be possible (as no distribution would have been assumed for the frailty) to calculate the marginal distribution, which is in fact the primary goal. The linear Bayes ap-

proximation is therefore not an option under the frailty model. A frequentist method would be to use the EM algorithm or the MCEM algorithm (treating the frailties as missing data). The Bayesian approach is the preferred method of analysis in this thesis, and so continuing with the approach which is to be used in the other data types, MCMC will be used.

The Likelihood

It has already been mentioned in Chapter 3 that the temporal factorisation is more efficient when using MCMC simulations, for right censored data. By the same reasoning, the temporal factorisation is also the most efficient factorisation for the right censored frailty model. But this is only so for the temporal parameters, and not for the grouping parameters. For the grouping parameters, the most efficient factorisation is a factorisation over the groups (a fuller discussion on this is given in Chapter 5, section 5.1.1). So for the frailty model two likelihoods are constructed: the first based on a temporal factorisation; and the second based on a factorisation across the groups.

Again let r_j represent all individuals who fall into the risk set at the beginning of interval I_j; t_{ij} is a modified survival time; δ_{ij} is a modified right censoring indicator; and z_{ij} and u_{ij} are the covariate and grouping indicator for individual i in interval j (u_{ij} must be one of the grouping effects u_g

for $g = 1, ..., G$). All of the modified variables were defined at equations 3 and 4 (page 58). Using the temporal factorisation, the likelihood may be factorised into a product of contributions from each interval:

$$L = \prod_{j=1}^{N_T} L_j.$$

The likelihood contribution for interval I_j is simply a slight adaptation of the likelihood contribution L_j from a right censored data set (section 3.2.2), the modification arises as a result of the frailties:

$$L_j = \prod_{i=1}^{r_j} e^{x_{ij}\beta_j\delta_{ij}} u_{ij}^{\delta_{ij}} \exp(-e^{x_{ij}\beta_j}(t_{ij} - t_{j-1})u_{ij}). \tag{22}$$

When evaluating the full conditionals for the frailties, it is preferable (for reasons already mentioned) to consider the likelihood as a factorisation over each of the groups; and within each group, over all of the individuals who belong in that particular group:

$$\begin{aligned} L &= \prod_{g=1}^{G} L_g \\ &= \prod_{g=1}^{G} \prod_{i=1}^{n_g} S(t_{ig})h(t_{ig})^{\delta_{ig}}, \end{aligned}$$

where n_g denotes the number of individuals within group g; t_{ig} the survival time for individual i in group g, is such that $t_{ig} \in I_{m_{ig}+1}$; δ_{ig} is the right censoring indicator for observation i in group g; and L_g denotes the likelihood contribution for group g. Using conditional survival functions, each

individual likelihood contribution L_g is factorised over the time axis:

$$\begin{aligned} L_g = & \prod_{i=1}^{n_g} \exp(-u_g \sum_{j=1}^{m_{ig}} e^{x_{ig}\beta_j}(t_j - t_{j-1})) \\ & \times \exp(-e^{x_{ig}\beta_{m_{ig}+1}}(t_{ig} - t_{m_{ig}})u_g) e^{x_{ig}\beta_{m_{ig}+1}\delta_{ig}} u_g^{\delta_{ig}}, \end{aligned} \tag{23}$$

since all individuals in group g share the same hazard u_g. Using the notation introduced in the previous section, this may be re- written in the abbreviated form as:

$$L_g = \prod_{i=1}^{n_g} \exp(-u_g B(t_{ig}, m_{ig})) e^{x_{ig}\beta_{m_{ig}+1}\delta_{ig}} u_g^{\delta_{ig}}, \tag{24}$$

4.4.2 Interval Censoring and Frailty

The conditional non-proportional hazards frailty model (above) may be readily extended to incorporate interval censored data. The model definition does not change and neither does the resulting marginal distribution. However when the data are interval censored, the temporal factorisation is no longer an option (for the same reasons as under the interval censored model, section 4). Because the factorisation over the groups is more efficient when it comes to computing the full conditionals for the frailty parameters, this is the factorisation which will be used to compute all of the full conditionals under this model.

Interval censored observations once again take on the form $(R_{ig}, L_{ig}]$, for an observation i in group g. Assuming that $R_{ig} \in I_{r_{ig}+1}$ and $L_{ig} \in I_{l_{ig}+1}$, and

using a similar form of the likelihood used in the interval censoring model in section 4.2.1 (but this time with a factorisation over the groups):

$$L(T) = \prod_{g=1}^{G} \prod_{i=1}^{n_g} A'(1, r_{ig}, g, \beta)C'(R_{ig}, t_{r_{ig}}, g, \beta) - A'(1, l_{ig}, g, \beta)C'(L_i, t_{l_{ig}}, g, \beta). \tag{25}$$

The functions $A'(\cdot)$ and $C'(\cdot)$ represent the modified conditional survival function (modified by incorporating frailties):

$$A'(a, b, g, \beta) = \exp(-\sum_{j=a}^{b} e^{z\beta_j}(t_j - t_{j-1})u_g),$$

and

$$C'(t, t_j, g, \beta) = \exp(-e^{z\beta_{j+1}}(t - t_j)u_g).$$

Double Interval Censoring and Frailty

It is possible to extend the above frailty model to cater for double interval censored data. The likelihoods may be easily adapted, and in the following Chapter it will become evident that the MCMC method used, could in theory be extended to the double interval censored frailty model. The resulting MCMC simulations, for the double interval censoring model, with frailties, could be very computationally intensive, and convergence may be difficult to determine. For these reasons the double interval censored frailty model is not considered as an application in this thesis. However simply by substituting the modified survival functions $A'(\cdot)$ and $C'(\cdot)$ into the likelihoods developed

in section 4.3.2, the likelihood for the terminating event may be obtained (the likelihood for the initiating event would be the same as that derived above for the interval censored case).

4.4.3 Individual Frailties

It has been suggested that a frailty model may be used to avoid the assumptions of proportional hazards (Clayton and Cuzick, 1995): each individual is assigned to an individual frailty group, and an individual frailty included within a conditional proportional hazards frailty model. As the marginal hazards will not be proportional, it is clear that the proportional hazards assumption will have been avoided. But it must be considered whether firstly it is appropriate to model frailties with just one member within each family (Vaupel *et al.*, 1979); and secondly what are the marginal assumptions involved with the conditional proportional hazards frailty model (for the Gamma frailty model these are a convergence of hazards).

Such a model was fitted to the gastric cancer data. Clearly demonstrated within Chapter 3, proportional hazards is not an appropriate model for this data set. The estimated survival using the conditional proportional hazards frailty model is presented in figure 5. The true estimate of survival is known to cross at around 1000 months. This feature is not reflected within

the estimate using the conditional frailty model. Clearly although avoiding the proportional hazards assumption, the model includes other assumptions which are once again not representative of the data set. Also plotted within figure 5 is the survival estimate obtained using the Normal Dynamic Survival Model, again with individual frailties. The dynamic covariate effect structure of the model once again alows the interesting feature of crossing survival curves. As an additional comparision, figure 6 compares the survival estimated using the NDSM, with that obtained from the NDSM with frailties. The two survival curves are similar, with the frailty model estimating a slightly greater survival proportion throughout. Figure 8, shows how similar the estimated treatment effect is between the two models, although the difference between the estimated baseline hazards is greater (figure 7). The estimated frailty varibables are skewed to the left, thus pertaining a population which is on the whole less frail than the mean: resulting in an estimated survival slightly better than that estimated without accounting for heterogeneity.

A sample of estimated frailties are presented within the table 3. The frailties clearly decease with a later observed survival time. It is known (from the analysis within Chapter 3), that chemotherapy and radiation is the most hazardous treatment up until around 1000 days. The mean frailty under the

combined treatment, over the first 1000 months, is 1.21 compared with 1.12 under chemotherapy alone, again over the same time period; over the last 800 days the mean frailty under the combined treatment is 0.43 compared to 0.60 under chemotherapy. This simple analysis appears to confirm that the frailties have incorporated this interesting feature of the data. The estimated frailties obtained under the proportional hazards and NDSM both with individual frailties, are very similar.

Using the conditional proportional hazards model as a way of avoiding the proportional hazards assumption, is clearly not as effective at predicting survival, when compared with using the very flexible Normal Dynamic Survival Model.

Survival	Censor	covariate	u_g	Varu_g
17	1	1	1.495373	0.861244
42	1	1	1.462847	0.784244
44	1	1	1.471534	0.847566
1622	0	1	0.406339	0.100519
1626	0	1	0.554389	0.126074
1736	0	1	0.372066	0.086371
1	1	0	1.527972	0.915784
63	1	0	1.463782	0.817383
105	1	0	1.423087	0.795385
1511	1	0	0.614840	0.142355
1690	0	0	0.407004	0.098125
1694	1	0	0.591110	0.132237

Table 3: Gastric Cancer with Individual Frailties: A selection of frailties estimated for various individuals.

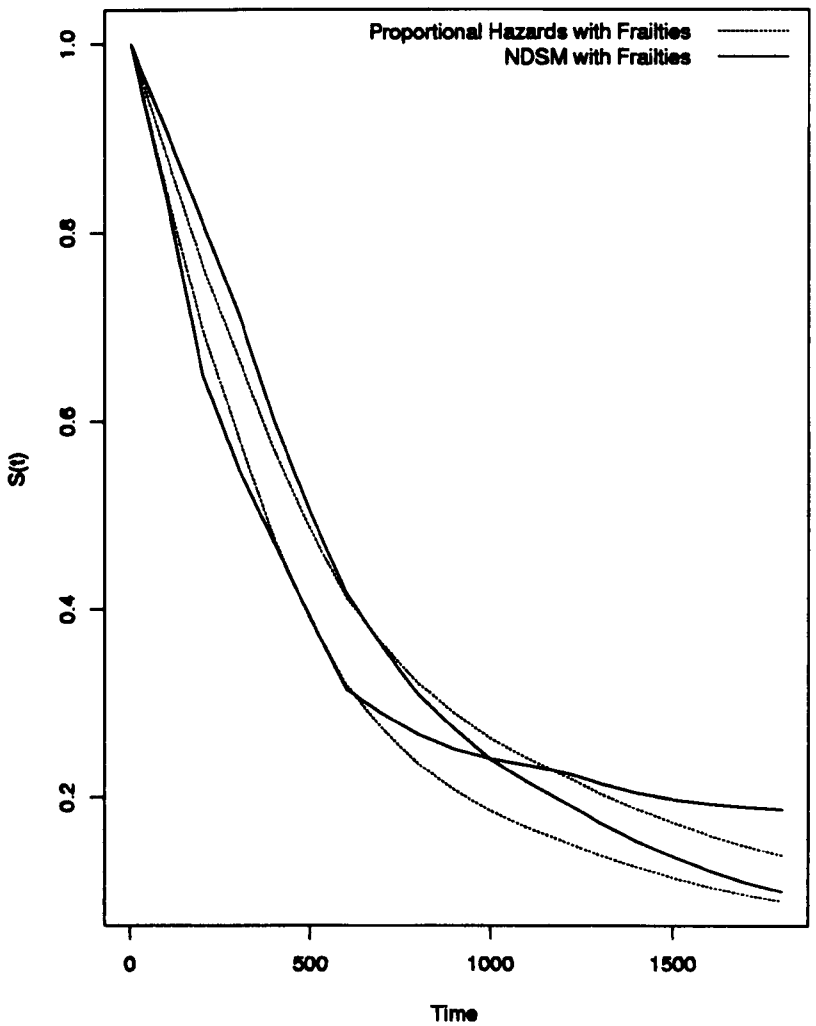

Figure 5: Gastric Cancer with Individual Frailties: Estimated survival using individual frailties. The conditional NDSM estimates survival curves to cross with chemotherapy the initial superior treatment. The conditional proportional hazards model estimates chemotherapy to be the superior treatment throughout.

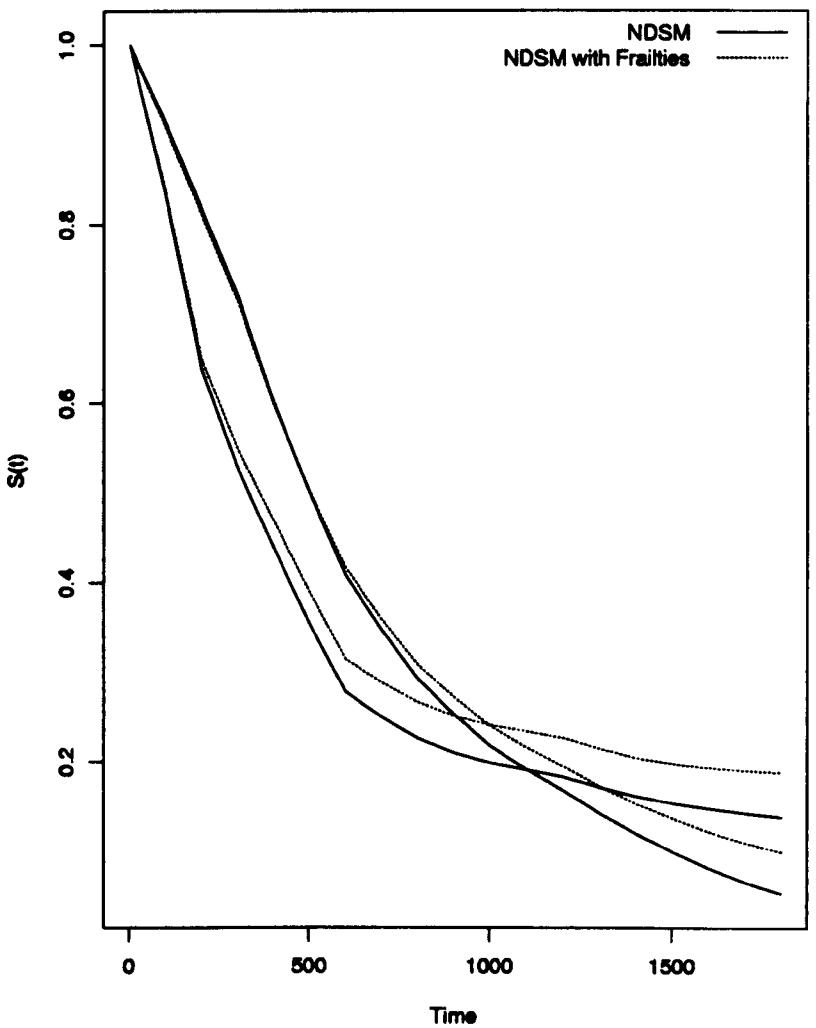

Figure 6: Gastric Cancer with Individual Frailties: A comparision of survival using the NDSM and the conditional NDSM with individual frailties. Survival under the conditional NDSM is estimated to be slightly better than that obtained by the NDSM. Those treated with chemotherapy have an estimated better prognosis throughout.

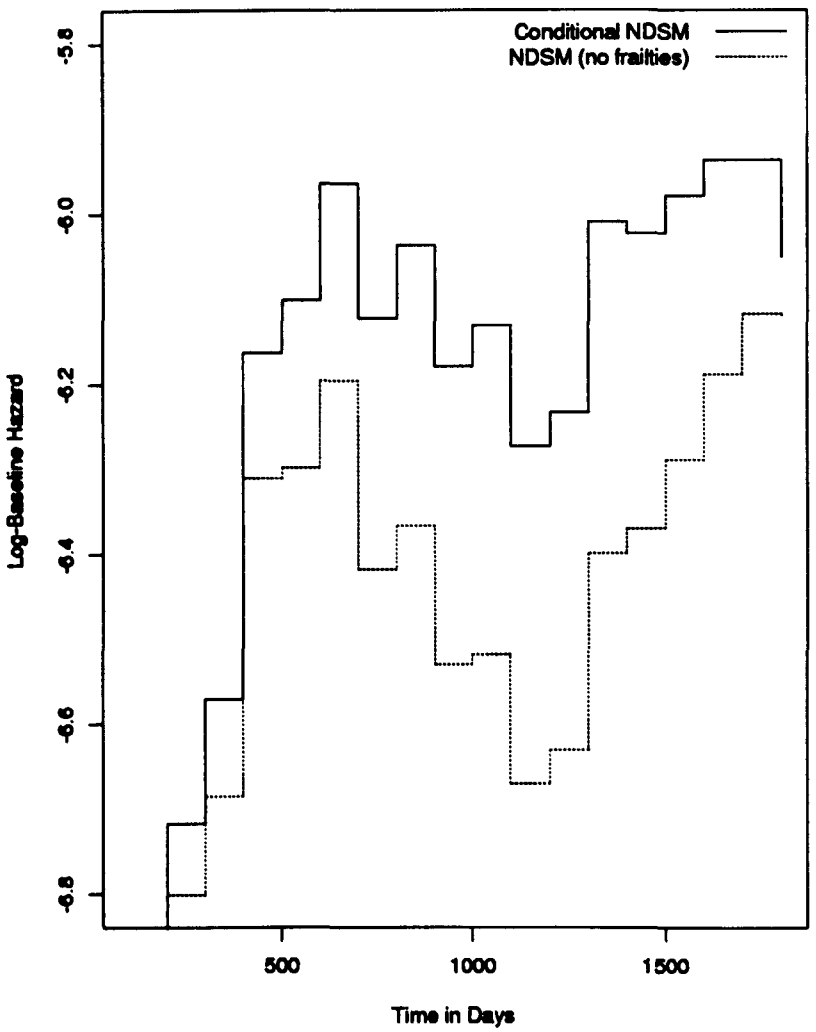

Figure 7: Gastric Cancer with Individual Frailties: The estimated log-baseline hazard is estimated to decrease and then level off after around 500 days. This pattern is fairly consistent between both the NDSM and the NDSM with individual frailties.

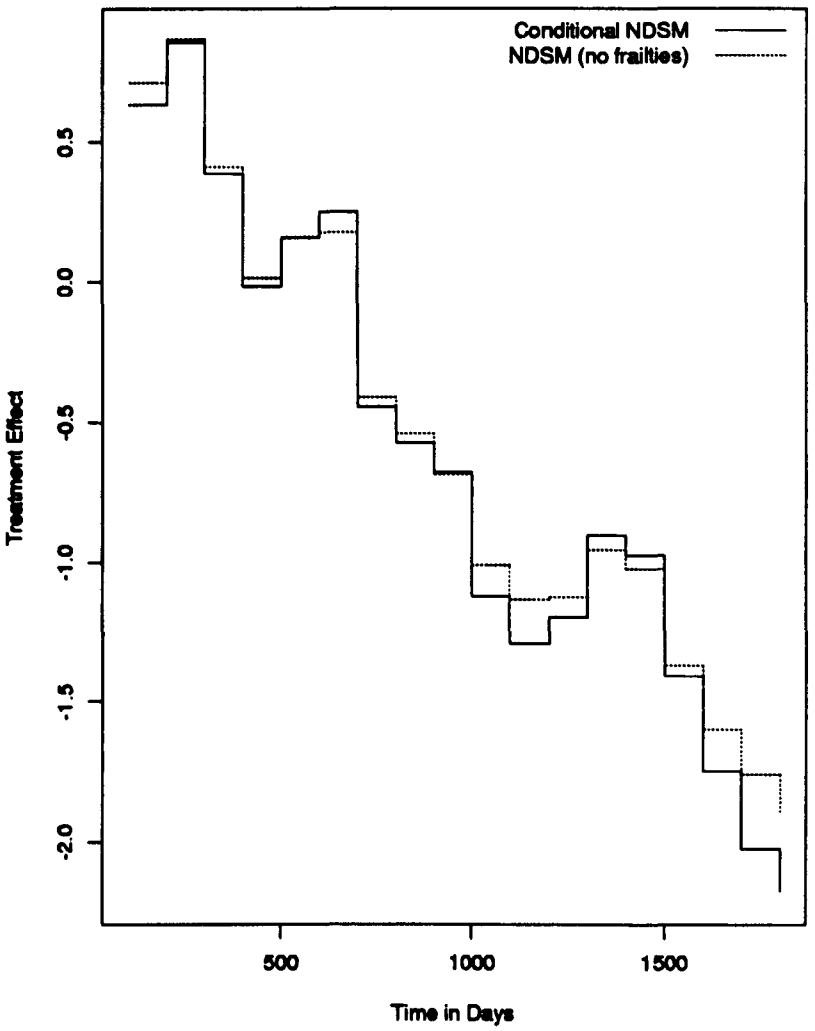

Figure 8: Gastric Cancer with Individual Frailties: The estimated treatment effect is very similar between both the NDSM and the NDSM with individual frailties: a decreasing effect with chemotherapy the initial superior treatment.

4.5 Summary

Likelihoods have now been developed under the Normal Dynamic Survival Model so as the model can accomodate interval censoring, double interval censoring and non-independent observations. The likelihood developed under the interval censored and frailty models were straightforward to derive. For the double interval censored model, integration complications resulted in an approximation being developed. The frailty model was then applied to the gastric cancer data set, with each individual assigned an individual frailty. Although the model appeared to capture well those individuals who were frail, it was still essential to use a dynamic covariate effect to represent accurately how the effect of treatment changes over time.

5 Model Fitting

It has been proposed throughout this thesis, that the method to estimate the model parameters will be by Markov chain Monte Carlo simulations: this will be the focus of this current Chapter. In Chapter 3, MCMC was used in an application of the right censored Normal Dynamic Survival Model. Here the same basic principals established in Chapter 3, will be used in the analysis of the Normal Dynamic Survival Model when applied to interval censored, double interval censored, and frailty data sets. For interval censoring the form of the full conditionals changes only slightly; for double interval censoring there exists a choice of augmenting the data to either interval or right censored data, or performing a full MCMC analysis on the approximate likelihood (section 4.3.2); for frailty models various factorisations are used with the aim of creating the most computationally efficient full conditionals. At the end of this Chapter, the methods will be applied to a data set simulated from a known survival distribution, and the impact of various degrees of censoring investigated.

5.1 Gibbs Sampling

All of the models considered in this thesis are Dynamic Survival Models, and consequently they all have in common the same basic structure. For

the Gibbs sampler, this results in the same basic set of parameters which must be estimated. This basic set of parameters may further be subdivided into two groups, the main parameters and the evolution variances. If $T \sim NDSM(\beta, N, W)$, then the main set of parameters are (all for $k = 1, ..., K$):

$$\beta_{01}, w_{02}, ..., w_{0N},$$

$$\beta_{k1}, w_{k2}, ..., w_{kN},$$

and the evolution variances are:

$$W_0, W_1, ..., W_K.$$

5.1.1 The Full Conditionals

The full conditionals for the main and hyper-parameters, assuming a likelihood L are all derived below. These full conditionals differ somewhat to those derived under the right censored model (Chapter 3). This is because under the right censored model, the likelihood was constructed using a temporal factorisation; for interval and double interval censoring, the individual factorisation has been used instead: this means that the full conditionals do not reduce to the same format as in the right censored model. Where the full conditionals reduce to a standard distribution form then this is derived. Where the full conditionals do not reduce to a standard distribution, then

it will be described within the next section how Metropolis-Hastings will be used to sample from the relevant distribution.

- For the paramater β_{01}:

$$[\beta_{01}|\cdot] \propto Lp(\beta_{01}).$$

- For the evolution of the log of the baseline hazard:

$$[w_{0j}|\cdot] \propto Lp(w_{0j}) \text{ for } j = 2, ..., N.$$

- The full conditional for the inverse of the evolution variance of the log-baseline, is exactly as under the right censored model (equation 11, page 83). Without deriving this distribution again, the full conditional is of a standard gamma form:

$$[W_0^{-1}|\cdot] \sim G(\zeta_0', \eta_0'),$$

where

$$\zeta_0' = \zeta_0 + \frac{N-1}{2} \text{ and } \eta_0' = \eta_0 + \sum_{j=2}^{N} \frac{w_{0j}^2}{2}.$$

- For the parameter β_{k1}:

$$[\beta_{k1}|\cdot] \propto Lp(\beta_{k1}).$$

- For the evolution of the covariate effect parameters:

$$[w_{kj}|\cdot] \propto Lp(w_{kj}) \text{ for } j = 2, ..., N.$$

- Once again the full conditional for the inverse of the evolution variance of the covariate effects is (derived at equation 12, page 84):

$$[W_k^{-1}|\cdot] \sim G(\zeta_k', \eta_k'),$$

where

$$\zeta_k' = \zeta_k + \frac{N-1}{2} \text{ and } \eta_k' = \eta_k + \sum_{j=2}^{N} \frac{w_{kj}^2}{2}.$$

Where different distributions are used within the dynamic model (i.e. other than normal and gamma), naturally the form of the full conditionals will change, although the same basic structure outlined will always hold.

The Full Conditionals for the Interval Censored Model

The likelihood L, for a set of interval censored data is given in section 4.2 (equation 14, page 102), and simply by substituting the likelihood into the above, all of the relevant full conditionals may be derived.

The Full Conditionals for the Double Interval Censored Model

To recap the initiating times, S, are modelled via $S \sim NDSM(\lambda, N_S, V)$, and the terminating event times, T, are modelled via $T \sim NDSM(\beta, N_T, W)$. Using a Gibbs sampling approach essentially two analyses must be performed:

- Using Gibbs sampling obtain final estimates for the parameters for the initiating event S.

- Using the estimates for the parameters in the initiating time model, establish the log-likelihood for the incubation time (using the likelihood developed for the terminating event in section 4.3.2, equation 18). By substituting this log-likelihood into the full conditionals outlined above, estimates for the model parameters may then be made using Gibbs sampling.

Alternative methods for estimating the parameters in a double interval censored model will be outlined in section 5.2 of this Chapter.

Full Conditionals for the Frailty Model

The parameters which must be estimated within this model, may be divided into three groups. These consist of the main paramaters, hyper-parameters (as under the interval censored and double interval censored models), and in addition the frailty parameters: the grouping effects $u_1, ..., u_G,$; and the hyper-paramater for the frailty distribution, μ. For the basic set of parameters it is most efficient in terms of computing the full conditionals to use a factorisation across the time axis (section 4.4). For parameters in interval I_j, the only part of the likelihood which is not a constant of proportionality, is the likelihood contribution for that particular interval. For the first group of parameters, the full conditionals are therefore as in the right censored model (Chapter 3, section 3.2.6), with the one difference being L_j (the contribution

to the likelihood from interval j), which must be adapted to account for the frailties. The contribution to the likelihood in interval I_j, denoted here by L_j, was defined in equation 22 (page 116):

$$L_j = \prod_{i=1}^{r_j} \exp(-e^{x_{ij}\beta_{ij}}(t_{ij} - t_{j-1})u_{ij})e^{x_{ij}\beta_j\delta_{ij}}u_{ij}^{\delta_{ij}}.$$

The full conditionals for the evolution variances do not change, and therefore may be found on page 84 (equation 12). The full conditionals for the frailties are based on a factorisation over the groups (essentially for the same reasons as why the interval parameters were based on a factorisation over the time axis). Using this factorisation the full conditionals for the frailty parameters are as follows:

$$[u_g|\cdot] \propto L_g p(u_g)$$

The likelihood contribution L_g was derived in equation 24 (page 117), and $p(u_g)$ was defined within the model definition 4.1 (page 111). So that:

$$[u_g|\cdot] \propto \exp(-u_g(\mu + \sum_{i=1}^{n_g} B(t_{ig}, m_{ig}))u_g^{\sum_{i=1}^{n_g}\delta_{ig}+\mu-1},$$

where $B(t_{ig}, m_{ig})$ was defined in section 4.4, equation 21. This density is proportional to a gamma density, so that:

$$[u_g|\cdot] \sim G(\mu_1, \mu_2), \tag{26}$$

where

$$\mu_1 = \mu + \sum_{i=1}^{n_g}\delta_{ig} \text{ and } \mu_2 = \mu + \sum_{i=1}^{n_g} B(t_{ig}, m_{ig}).$$

The full conditional for the frailty hyper-paramater μ is:

$$[\mu|\cdot] \propto \prod_{g=1}^{G} p(u_g|\mu)p(\mu)$$

Using model definitions (definition 4.1, page 111):

$$[\mu|\cdot] \propto \frac{\mu^{G\mu+\kappa-1}\exp(-\mu(\nu+\sum_{g=1}^{G}\mu_g))\prod_{g=1}^{G}\mu_g^{\mu-1}}{\Gamma(\mu)^G},$$

unfortunately this density is not of a standard form, and so a Metropolis-Hastings step will be used to take samples from this full conditional (the proposals will be discussed in the following section).

Using a proportional hazards model based on a piecewise correlated baseline hazard, Aslanidou and Dey (1996) developed full conditionals similar to those developed here for the frailty paramaters.

The Full Conditionals for an Interval Censored Frailty Model

For the interval censored frailty model, it is not possible to use the temporal factorisation (as was possible for the right censored frailty model). The alternative individual factorisation, derived in section 4.4.2, must be used instead. The full conditionals for the main and hyper-parameters therefore follow the forms given on page 83, with the likelihood L as defined in section 4.4.2, equation 25. For the frailty parameters the full conditionals are similar to those in the right censored frailty model (above), although with a slight

modification (to account for the interval censoring), and are again based on a factorisation over the groups. As in the right censored frailty case, the full conditional for the frailty parameters u_g reduce to the likelihood contribution for that particular group multiplied by the prior for the frailty paramater u_g:

$$[u_g|\cdot] \propto L_g p(u_g).$$

By substituting the form of L_g for the interval censored frailty model (section 4.4.2, equation 24), as in the right censored frailty model (above), the full conditional is proportional to a gamma density so that:

$$[u_g|\cdot] \sim G(\mu_1, \mu_2),$$

where

$$\mu_1 = \mu + \sum_{i=1}^{n_g} \delta_{ig},$$

and

$$\mu_2 = \mu + \sum_{i=1}^{n_g} \{B(R_{ig}, r_{ig}) - B(L_{ig}, l_{ig})\},$$

and where the function $B(\cdot)$ was defined at equation 21 (page 113). The full conditional for the hyper-parameter μ does not change and is as defined in equation 26.

5.1.2 Metropolis-Hastings

The same basic format for the Metropolis-Hastings proposals will be used as were used in the right censored model (Chapter 3, section 3.2.6). For

the hyper-parameters of the frailty distribution (which are positive random variables), a gamma proposal was used, with similar features as described under the normal proposal. That is, if the current value of the parameter is u_g^m (where the m refers to the iteration number), then the proposal is:

$$G(\alpha, \gamma),$$

where $\alpha = cu_g^m$ and $\gamma = c\mu_g^m u_g^m$, and where the parameter c will control the variance of the proposal (whilst constraining the mean of the proposal to the current value of the parameter). The variance paramater was chosen to provide roughly forty to sixty percent acceptance rates. Combining this method with vague hyper-priors, along with checks on convergence, proposals based on this form fared well.

5.2 Imputation Methods

The analysis proposed so far for double interval censored data sets, is based on a computationally intensive approximation to the likelihood. Experience showed that the computation time to estimate model paramaters in such a data set, was substantially more than that involved within the analysis of interval censored and right censored data sets. This is the motivation behind the alternative method of model fitting described here. The method suggested is based on iteratively augmenting the double interval censored

data set into an interval censored one, and analysing this interval censored data set using the MCMC techniques proposed within this thesis. Experience proved that not only does the proposed method reduce the computation time involved, but also provides estimates very close to those obtained using the approximate likelihood. The method proposed has some similarities to, the data augmentation method of Tanner and Wong (1987). The main difference is that here at each iteration there exists only one augmented data set; the method proposed by Tanner and Wong (1987) samples several. Pan (2000) used a similar form of augmentation when analysising interval censored data using Cox's proportional hazards model.

We start by considering augmenting an interval censored data set. Not because it is expected that the computation time will be reduced by much: it will not be as the computation involved within the analysis of interval censored and right censored data sets are similar, and using the augmentation method will involve the additional computation involved of the imputation procedure itself. But the approach will identify the impact of augmentation on the resulting estimates, before the method is applied to double interval censored data.

5.2.1 Interval Censoring

The survival times T are again assumed to follow the Normal Dynamic Survival distribution: $T \sim NDSM(\beta, N, W)$, and the proposed augmentation method takes the following form:

1. Obtain initial estimates for all parameters (these may be obtained as either subjective estimates or they may be based on a midpoint analysis).

2. For each observation, augment the interval censored data $(R, L]$, by sampling a survival time, t, from the conditional survival function $S(t|R < T \leq L)$. The interval censored observation is replaced by an exact or right censored survival time.

3. Based on the augmented data set, use right censoring methods to estimate the parameters in the survival model.

4. Steps 2 and 3 are iteratively repeated.

The Conditional distribution

For each interval censored observation an estimated failure time must be sampled from the interval $(R, L]$. The conditional survival distribution from

which samples must be taken is (for t $in(R, L]$:

$$\begin{aligned} p(T > t | R < T \leq L) &= \frac{p(T>t, R<T\leq L)}{p(R<T\leq L)} \\ &= \frac{p(t<T\leq L)}{p(R<T\leq L)}. \end{aligned}$$

For $t \in I_{j+1}$, $R \in I_{r+1}$, and $L \in I_{l+1}$, then using equation 4.1 (page 99):

$$p(t < T \leq L) = A(1, j, \beta)C(t, t_j, \beta) - A(1, l, \beta)C(L, t_l, \beta),$$

and

$$p(R < T \leq L) = A(1, r, \beta)C(R, t_r, \beta) - A(1, l, \beta)C(L, t_l, \beta).$$

Sampling for Multiple Imputation

- Generate a uniform random variable u with the aim of solving for t:

$$S(t | R < T \leq L) = u \text{ where } T \sim NDSM(\beta, N, W).$$

- To establish within which interval t lies compute:

$$\begin{aligned} A_j &= p(T > t_j | R < T \leq L) \\ &= \frac{A(1, j, \beta) - A(1, l, \beta)C(L, t_l, \beta)}{\{A(1, r, \beta)C(R, t_r, \beta) - A(1, l, \beta)C(L, t_l, \beta)\}}. \end{aligned}$$

The A_j's are a series of non-increasing cumulative probabilities:

$$A_{r+1} > A_2 > .... > A_{l+1},$$

so find j such that

$$A_j < u < A_{j+1},$$

concluding that $t \in I_{j+1}$.

- Solving for t:

$$\begin{aligned} u &= p(T > t | R < T \leq L) \text{ for } t \in I_{j+1} \\ &= \frac{A(1,j,\beta)C(t,t_j,\beta) - A(1,l,\beta)C(L,t_l,\beta)}{A(1,r,\beta)C(R,t_r,\beta) - A(1,l,\beta)C(L,T_l,\beta)}, \end{aligned}$$

gives the generated failure time as:

$$t = t_j - \frac{1}{e^{x\beta_{j+1}}} \log \frac{Up(R < T \leq L) + A(1,l,\beta)C(L,t_l,\beta)}{A(1,j,\beta)}.$$

5.2.2 Double Interval Censoring

For double interval censored data sets, there are essentially two ways in which the data set may be augmented: either to an interval censored data set, or to a right censored one. Firstly the data set must be augmented on the left (i.e. by substituting a time of the initiating event), with the resulting data being either interval censored or right censored (this will depend on whether the initial data was doubly censored or double interval censored). When the augmented data set is interval censored then either the data set may be analysed as such (using the techniques developed for interval censoring), or alternatively the data may be augmented once more to create a right censored data set.

The method is very similar to modifying an interval censored data set by augmentation, and so is described only briefly here:

1. For each observation which is interval censored on the left, an initiating event time is sampled from the conditional survival function:

$$S(s|M < S \leq P),$$

where $S \sim NDSM(\lambda, N_S, V)$, and where the parameters for S will have already been estimated.

2. The doubly censored data is augmented to $(R - s, L - s]$ using the current sampled value s.

3. The parameters in the model $T \sim NDSM(\beta, N_T, W)$ are estimated using interval censored techniques.

4. Steps 2 to 3 are iteratively repeated.

A modification exists where at step 2 the data are augmented to a right censored data set (this will be based on the methods described above for augmenting an interval censored data set).

Sinha (1998) applied a very similar approach, modelling a set of double in-

terval censored data, although the data were augmented to grouped data, and advantages taken of tractable methods using a beta Levy process.

5.3 Simulation Study

Within this next section, for each type of censoring, data sets will be simulated with increasing degrees of censoring, so that impact of the degree of censoring may be studied. For each data set a survival time is simulated from a specified survival distribution. The method used to generate the failure times is very similar to the procedure used to impute death times in the previous section, simulating from the distribution $S(t)$ rather than $S(t|R < T \leq L)$.

When the data set which is to be generated involves right censoring, then for each observation a right censoring indicator must be generated. In this thesis this will be done very simply by specifying a cut-off date (T_{MAX}), and generating a uniform variable on the interval $[0, T_{MAX}]$. This will then represent the censoring time, and observations will become right censored where the censoring time is greater than the simulated survival time. Other more complicated (and also more realistic) censoring schemes could easily be incorporated, where perhaps the proportion of observations which are right censored increases overtime.

For interval censored data a censoring mechanism could be based on patients being requested to attend a series of examination times, at which time failure may be detected. For each observation the censoring interval could be identified as the two examination times within which the failure time lies. A simple version which would generate such a process could be based on a series of Bernoulli random variables: a finite series of possible examination times is specified, and at each time a Bernoulli random variable is generated, which will represent whether or not the individual attended the examination time.

For double interval censored data it is necessary to simulate two survival times: the first for the time of the initiating event; and the second for the survival time of the terminating event. From this the calendar time of the terminating event may be obtained as the sum of the two survival times. A series of examination times may then be generated as in the interval censoring case, with two intervals identified: containing firstly the time of the initiating event; and secondly the time of the terminating event.

Each data set consisted of 90 observations, equally divided by treatment group. The estimated survival under each data set is compared to the ob-

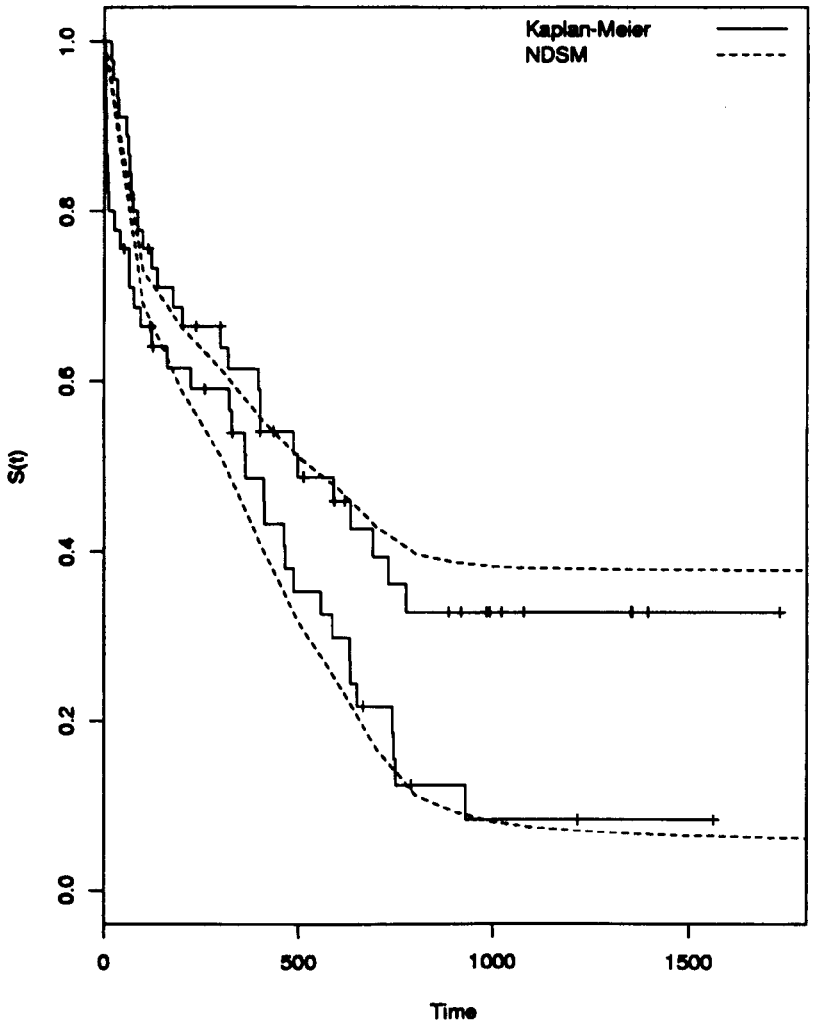

Figure 9: Right Censored Simulated Data (Group1):The data consist of approximately 30 percent right censored observations, equally divided by covariate group. Survival is over estimated under the NDSM at early times, due to a poor choose of time axis.

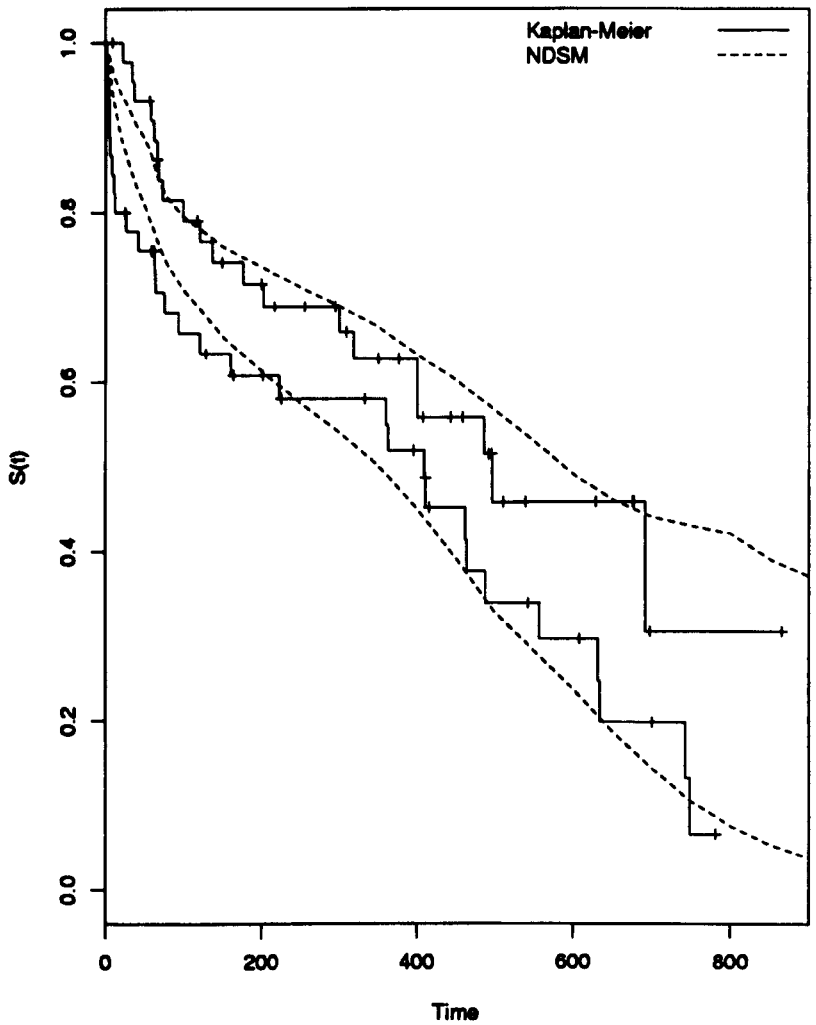

Figure 10: Right Censored Simulated Data (Group2): The data consist of approximately 44 percent right censored observations, equally divided by covariate group.

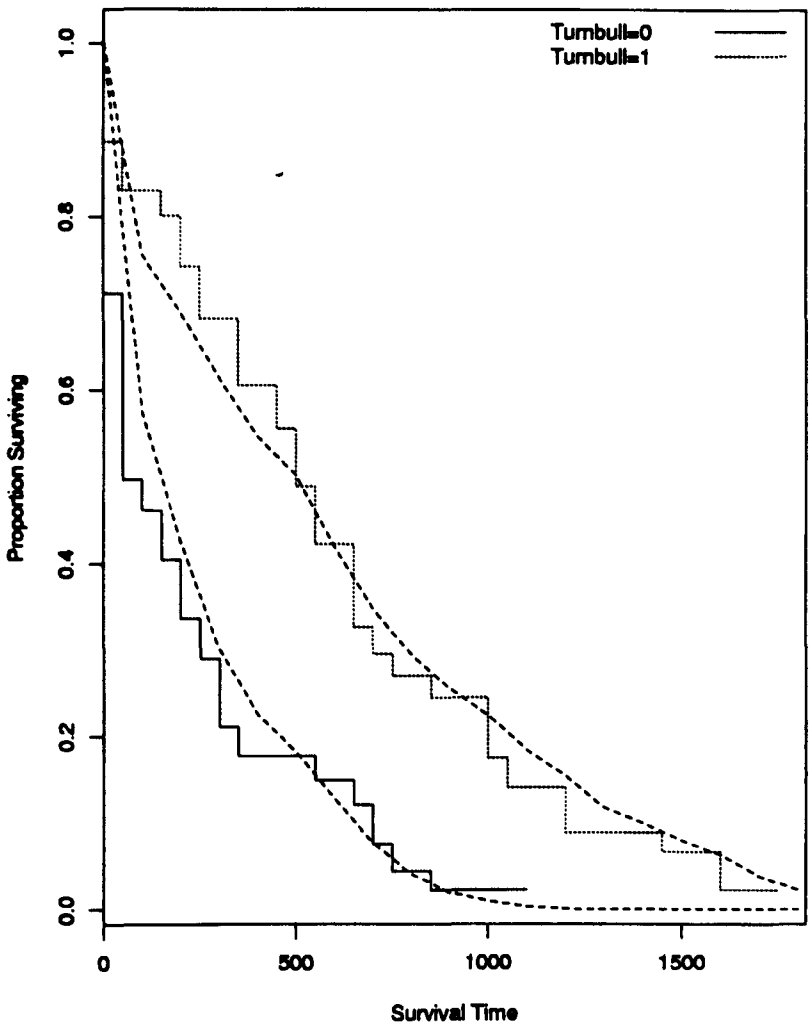

Figure 11: Interval Censored Simulated Data (Group1):The data consist of interval censored observations, equally divided by covariate group. The average length of the censoring interval is 87 units.

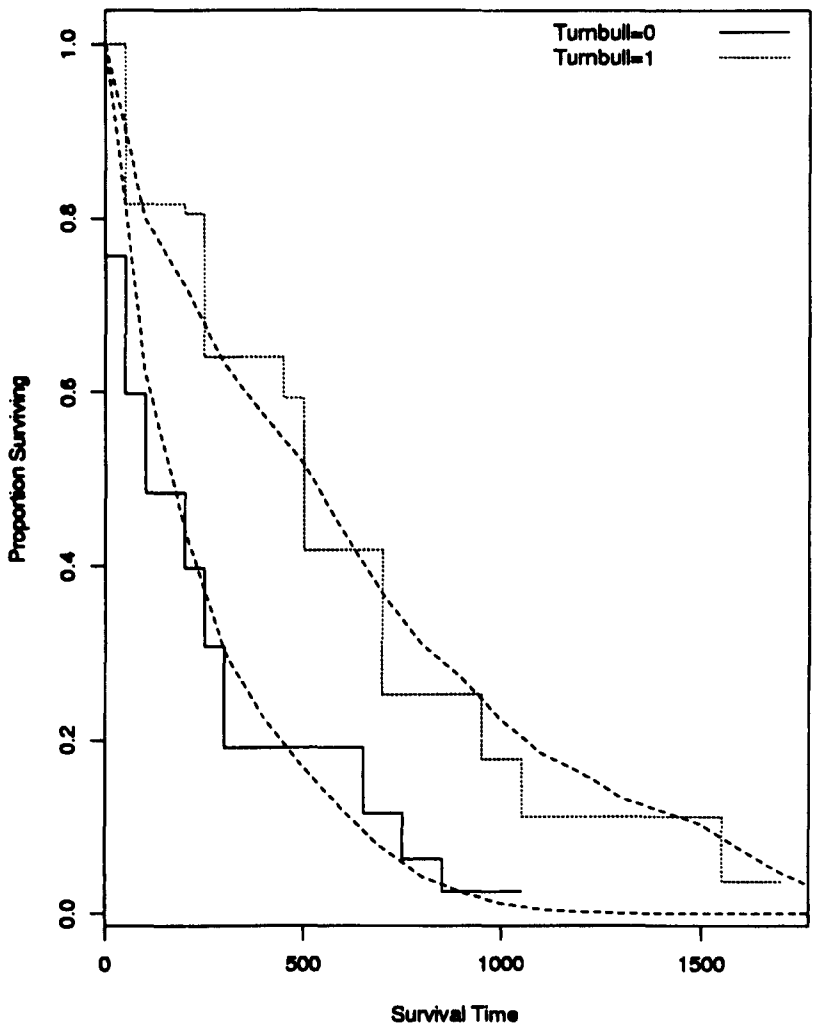

Figure 12: Interval Censored Simulated Data (Group2): The data consist of interval censored observations,equally divided by covariate group. The average length of the censoring interval is 251 units.

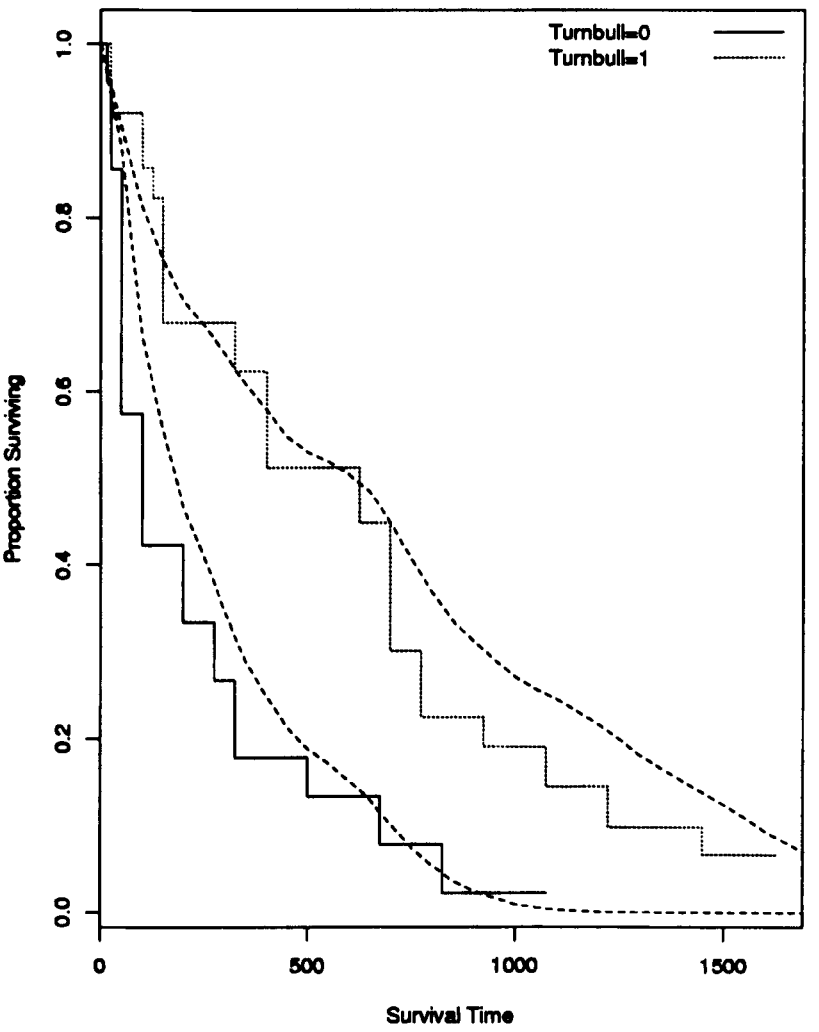

Figure 13: Double Interval Censored Simulated Data: The data consist of double interval censored observations, equally divided by covariate group. The average length of the censoring interval is 105 units on the left and 87 units on the right. Turnbull's estimate is based on midpoints for the initiating event.

served survival (as estimated by the relevant non-parametric technique).

For the right censored data set, the first group (Group 1), consisted of 30 percent right censored observations, and Group 2, 44 percent right censored. The initial division used for the estimation of the survival under Group 1 was an even division: $100, 200, ..., 1800$. The corresponding estimated survival is presented in figure 9: there appears to be a slight over-estimation of survival for very early times. This is almost certainly because the time axis division, is not fine enough to capture the amount of change within these very early days. Consequently estimation under Group 2 (figure 10), is based on a division of the time axis with a much finer division at early times: there appears to be an improved fit, highlighting the need for caution and care when choosing the time axis division. The model copes well with the increased proportion of censoring within the second group.

Figures 11 and 12, similarly demonstrate how the models can accommodate quite different degrees of interval censoring, and yet produce similar survival estimates. In the first example of interval censored data the average length of the censoring interval is 87 units, and in the second example it is increased to 251 units. The estimated survival changes very little when interval censoring is also introduced on the left (figure 13), although there does exist

some discrepency when comparing the survival estimated using the NDSM with Turnbull's estimate. The most likely cause for this discrepancy is due to the fact that the estimate produced by the Turnbull method was based on a data set with mid-point substitutes for the times of infection with HIV. The average length of the censoring interval on the left is 105 units and on the right 87 units.

5.4 Summary

Within this Chapter methods for model fitting to be used within the application of the NDSM have been discribed (a continuation of that in Chapter 3). The full conditionals for parameters have been outlined, and where necessary Metropolis-Hastings steps have been incorporated. For interval censoring and double interval censoring an alternative method based on a data augmentation algorithm has been discussed. It is expected that this method will reduce considerably the computation time involved within the estimation of the parameters for a double interval censored data set. The methods were successfully applied within a range of simulated data sets, and the impact of the degree of censoring considered.

6 Data applications

The focus of this Chapter will be on applying the Normal Dynamic Survival Model to interval censored, double interval censored, and frailty data sets. All of the methods for estimating model parameters, as described in Chapter 5, will be considered, and results compared. These applications will build on the principles already established through previous applications (Chapter 3 and 5, for the right censored and simulated data sets respectively).

All of the data frames, non-parametric survival curves (Kaplan-Meier and Turnbull), survival curves assuming the proportional hazards assumption (Cox, 1972), along with tests for proportionality, may be found within appendix A. For some of the data sets, the preliminary analysis is based on approximated data (e.g. using midpoints or by ignoring dependencies).

Vague priors have been used throughout for all of the covariate and baseline parameters (the same numerical values as in Chapter 3 for the gastric cancer data analysis were used). The number of iterations carried out was always greater than that recommended by the Raftery and Lewis (1992) diagnostic. Samples of the simulation plots are given in the appendix.

It is reiterated, that even if covariate effects are found not to change over

time, then the Normal Dynamic Survival Model may still be used. In a right censored application the approach would then provide an alternative to Cox's proportional hazard model (and indeed also an alternative to Gamerman's approach). With interval censoring, the method would provide a comparison with Finkelstein's (1985) extension of the proportional hazards model to interval censoring. With double interval censoring similar comparisons could be made with the models developed by Kim *et al.* (1993). Within the data analysis likelihoods are often compared between models, it must be remembered that comparing likelihoods in this way is a guide only, as the likelihoods are themselves based on estimates.

6.1 Breast Cancer Data

Surgical removal of an invasive breast cancer tumor, is an alternative option to a mastectomy; generally being preferred as it is supposed to retain a better cosmetic appearance. In the following study, breast cancer patients were followed after they had undergone surgery to remove the tumor. Patients were requested to visit the clinician every 4 to 6 months, when cosmetic appearance would be assessed. Although the patients were requested to return regularly, actual intervals between visits were sometimes much longer, especially for patients who were geographically remote. At each visit a cosmetic appearance score for the breast was derived. Remission was defined to oc-

cur, for the purpose of this data analysis, when the overall cosmetic score deteriorated by a significant amount, as determined by the study organisers. Remission is the survival event of interest which can be identified to lie within an interval, generating a set of interval censored survival data. Interest lies in comparing the survival of patients treated with radiation to those treated with chemotherapy and radiation. The data set is contained in table 7(page 198) with time measured in months.

Finkelstein and Wolfe (1985) were among the first to analyse this data set. From the fitted semi-parametric survival regression model (adapted for interval censoring), there is a clear indication that the treatment which consistently increases the chances of survival, is radiation alone. Ghosh and Sinha (1995) later considered a non-parametric analysis for one of the covariates from this same data set, using a piecewise correlated baseline hazard model (they did not consider a dynamic covariate effect), although failing to display any numerical output from their fitted model. Sinha (1997) used a semi-parametric approach (no covariates and a baseline hazard modelled by a Lévy process), estimating that the survival curves cross at an early time. Sinha concluded by plotting a credible band of difference in two $\log(-\log(S(t)))$ against time, that there was not enough evidence to reject proportional hazards. Although when considering mean estimated values

only, then the assumption would be rejected. There is a noticeable difference in the estimated survival under the model proposed by Sinha (1997) and that proposed by Finkelstein and Wolfe (1985). The first difference probably occurs as a result of the parametric assumptions made by Finkelstein and Wolfe (1985): the impact of this is to predict radiation as the superior treatment throughout; the model proposed by Sinha (1997) is not restricted by such assumptions, estimating that the survival curves cross at an early time. The second larger difference occurs between the prediction under the radiation group: Sinha (1997) estimates less than 5 percent surviving after 60 months; whereas Finkelstein and Wolfe (1985) estimate the proportion to be 40 percent. Goggins *et al.* (1998) compared a proportional hazards model fitted to the interval censored data (based on an enumeration of all possible rankings), to a midpoint imputation approach, concluding that the exact approach provided a better fit. Using the EM algorithm, the treatment effect was estimated to be 1.45 (standard error 0.371). Pan (2000) using a proportional hazards model estimated the treatment effect to be 0.90 (standard error 0.29).

The estimated survival using the non-parametric approach of Turnbull (1976) is given in figure 31 (page 199). The survival curves are estimated to cross at an early time, and the final estimated proportion surviving at time 48,

are 41 and 0 percent. The estimated survival under Cox's model, using the midpoints of the data, is given in figure 32 (page 200). Estimated survival under radiation at 60 months is 35 percent, again very different to that estimated by Sinha (1997). The Grambsch and Therneau (Grambsch and Therneau, 1994) test for proportional hazards is statistically significant at the five percent level (chi-squared value of 4.5 on 1 degree of freedom). The accompanying plot indicates that the treatment effect follows a quadratic like shape (figure 33, page 201).

6.1.1 Analysis using the NDSM

The value 1 is assigned to the combined treatment and 0 to radiation. A division of the time axis was chosen, with interval lengths increasing as the amount of information decreased. The log-likelihoods observed after 1,000 and 10,000 iterations changed only slightly (-137.27 and -135.42), and as convergence checks were also satisfactory, it was decided not to consider 100,000 iterations. Two different methods of model fitting were used: the exact likelihood approach and the imputation method (Chapter 5). As can be seen from the estimated survival (figure 14), estimates from the two approaches were very similar. The largest difference observed is between the treatment effect for the first time interval, although variance estimates for this parameter are quite large, and the impact on the predicted survival is slight (see table 19

for standard errors of estimates). A summary of the numerical output comparing the two methods is presented in table 19 (page 225). The estimates obtained using the method of imputation have a slightly larger variance than those estimates obtained using the exact approach.

Estimates for the treatment effect are given in figure 15. The treatment effect increases sharply over the first 20 months after surgery, changing from a negative to positive effect at around 10 months. After 20 months the treatment effect begins to decrease, although always remaining above zero. The estimates for the evolution variances are 0.21 and 0.51 (variances 0.05 and 0.21). This estimate is consistent with that produced by the Grambsh and Therneau (1994) method (figure 33, page 201). The conclusion is that chemotherapy appears to increase the probability of acute skin reactions when administered in conjunction with radiotherapy, although this is not so certain in the earlier days, where the effect could possibly be reversed. A complete numerical output is given in the appendix (table 19, page 225).

The corresponding survival curves are given in figure 14. The estimate compares well with that estimated by Sinha (1997), up until around 45 months: Sinha estimates that survival under radiation drops to zero between months 45 and 60. There seems however to be little evidence from the data to support

this, with very little information observed after month 45. Indeed Turnbull's estimate (figure 31) also does not support that estimated by Sinha (1997).

The estimated survival, constraining the NDSM to a proportional hazards model, is given in figure 17, with the estimated treatment effect being 0.93 (variance 0.083), which is similar to that estimated by Pan (2000) (estimated treatment effect 0.90 and standard error 0.29). This estimate is lower than that estimated by Goggins *et al.* (1998), although within two standard errors of their estimate. It is also noted that the data set does not meet the requirements of the proportional hazards model (covariate effects have been estimated to change over time), and so by fitting an inappropriate model it is possible that poor parameter estimated may be obtained. The estimated log-likelihood is -140.79, slightly worse than that observed under the non-proportional hazards model.

With this data set it seems that a proportional hazards analysis, although it may not be completely appropriate, does not draw such extreme, incorrect conclusions, as were observed in the gastric cancer data analysis. On the other hand the method has shown very clearly that the effect of the treatment is different from a constant function.

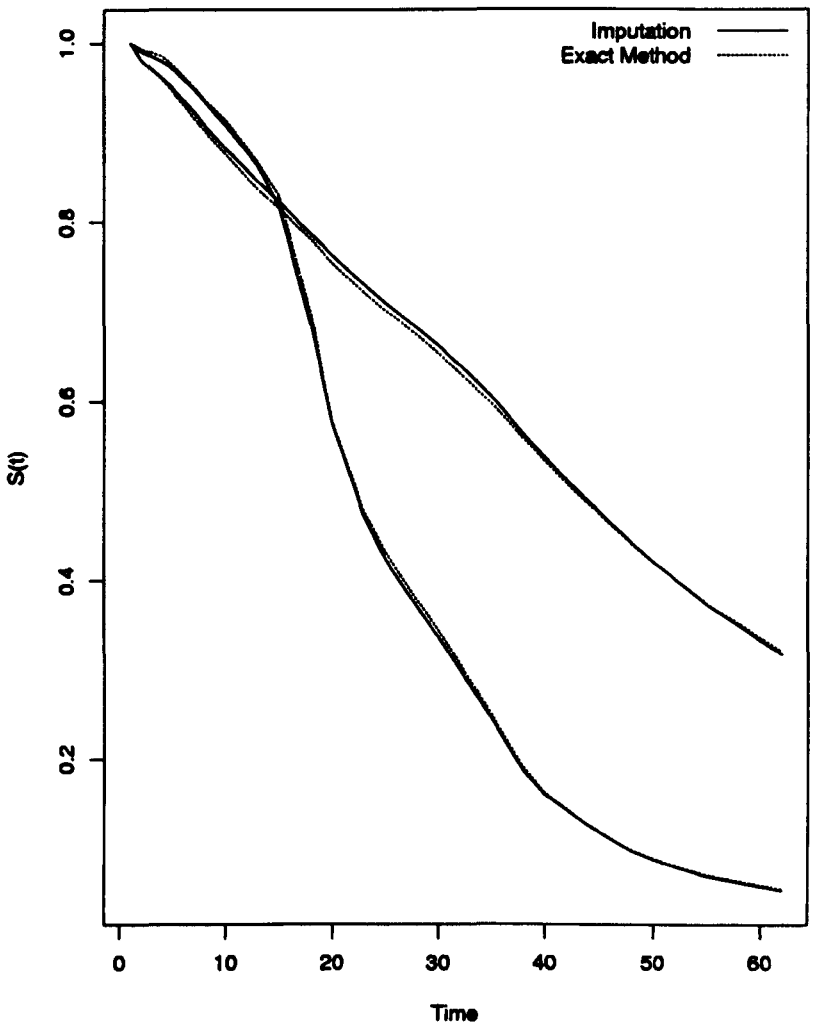

Figure 14: Breast Cancer: Survival estimates show survival under radiation is better than under the combined treatment, except during the first 20 months after diagnosis.

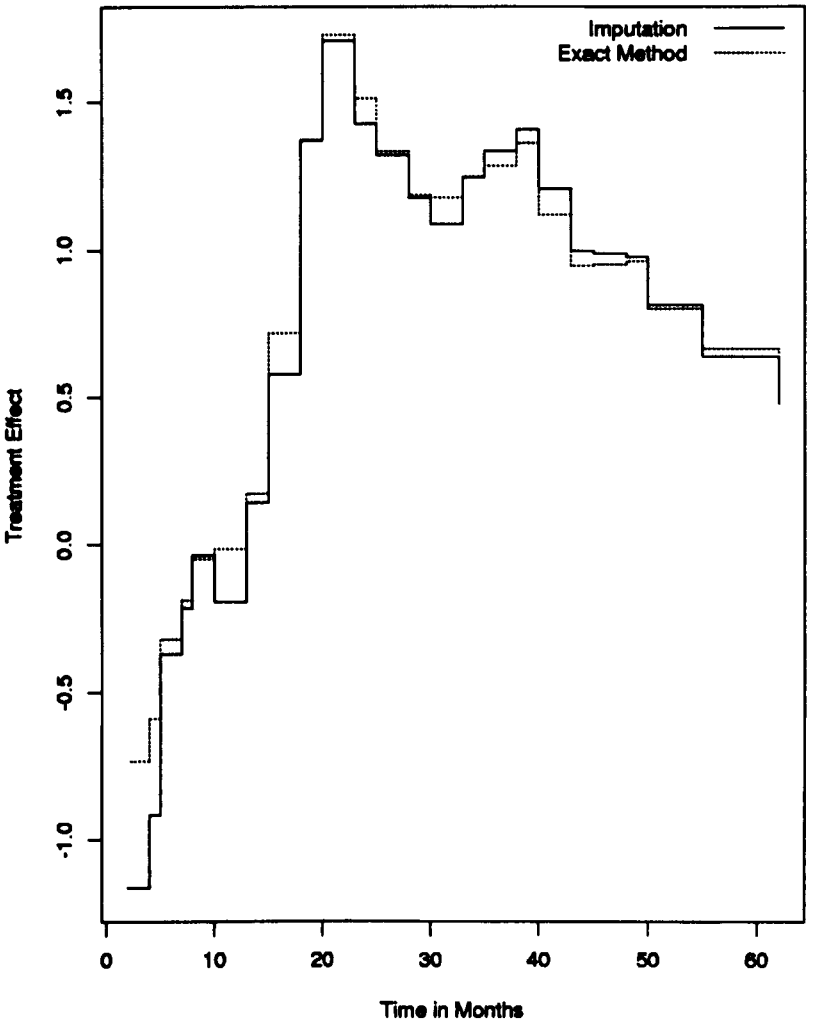

Figure 15: Breast Cancer: The treatment effect indicates that radiation has an initial negative effect, but changes to a positive effect after around 15 months.

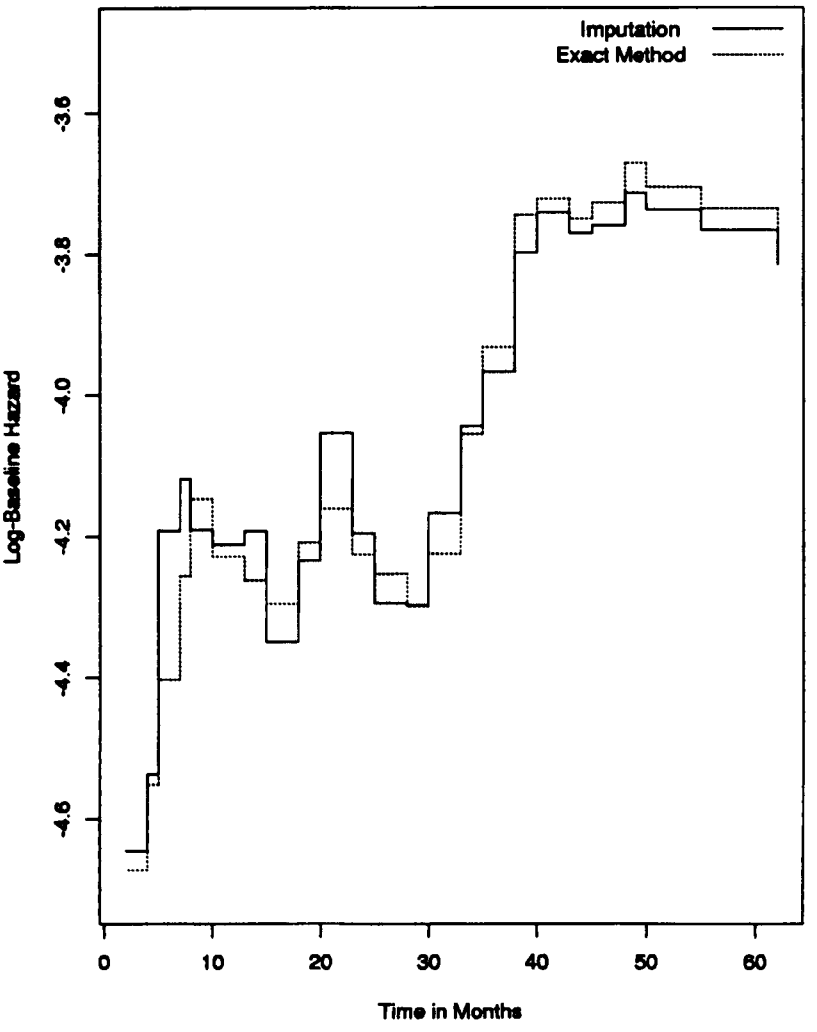

Figure 16: Breast Cancer: The log-baseline hazard decreases over time.

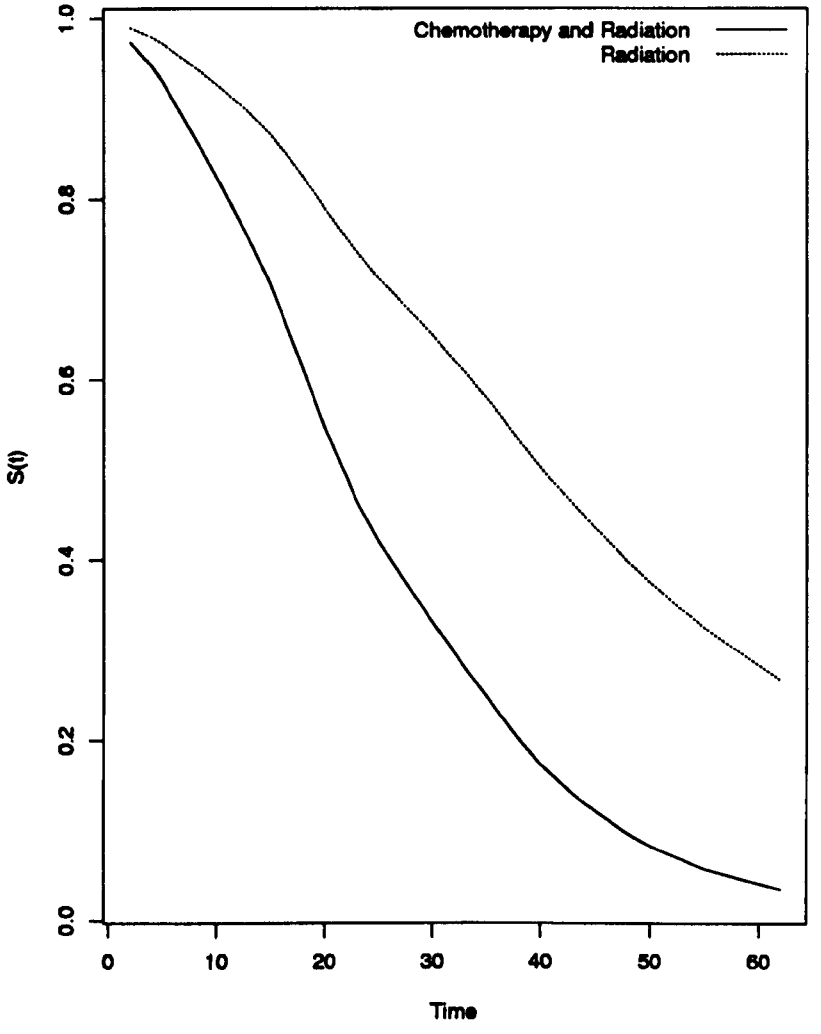

Figure 17: Breast Cancer: Survival estimated using the NDSM constrained to return proportional hazards estimate survival under radiation is better than under the combined treatment throughout.

6.2 Kidney Infection Data

This data set is based on the time to infection of 38 kidney patients (table 10, with survival times recorded in days). For each individual two infection times are recorded, either exact or right censored. There is a strong possibility that the two infections times for any individual will be positively correlated, the data are therefore analysed using frailties. In the terminology described in the previous Chapters, the data set consists of 38 groups with 2 observations in each group, and just one recorded covariate for each observation, which represents gender (age is a covariate that has sometimes been used within this data set although is not used here). The data set was introduced in this thesis as an example in Chapter 2 (example 2.9).

Initially ignoring the fact that the data are grouped, and using 0 to represent males and 1 for females, Kaplan-Meier curves were plotted for both males and females (figure 35, page 206). Cox's proportional hazards model was also fitted, and again sex was found to be a significant covariate (figure 36, page 207). Grambsch and Therneau (1994) tests of proportionality yielded a chi-squared value of 11.2 on 1 degree of freedom, giving a clear indication of lack of proportionality. The estimated plot (figure 37, page 208) of the covariate effect over time, showing a very evident decreasing effect.

There have been several analyses of this data set: the data have been used in numerous papers as an example of a random effects data set. McGilchrist and Aisbett (1991) modelled the data using a conditional proportional hazards model, with log-normal frailties. Aslanidou and Dey (1996) fitted a gamma frailty model to the data set, comparing the conditional proportional hazards model (using a piecewise correlated process to model the baseline), with the conditional Weibull proportional hazards. Sinha (1998) used a posterior likelihood approach, modelling the data by a conditional proportional hazards model, again using a piecewise correlated process modelling the baseline hazard. Qiou (1997) compared the positive stable frailty to the gamma frailty model on the same data set, although as mentioned in Chapters 2 and 4, the positive stable frailty model is not always considered as an appropriate model. All of the above mentioned papers found that after accounting for the dependencies within individuals, the effect of the covariate representing sex increased.

To investigate the time dependencies of the covariate effect sex, a Normal Dynamic Survival Model was initially fitted to the data (with no frailties, and assuming that all of the observations were independent). The resulting survival curves, covariate effect and predicted baseline hazard may be found within figures 18 to 20. Consistent with the estimate produced by

the Grambsch and Therneau (1994) estimate, the effect of the covariate decreases over time, indicating that proportional hazards is not an appropriate assumption to make (at least when not accounting for the dependencies). The estimated log-likelihood under the proportional hazards model was -332.73 compared to -329.50 when fitting a non-proportional hazards model.

Allowing for the fact that the data are grouped, two further survival models were fitted to the data: the first a conditional proportional hazards model with gamma frailties (using the NDSM constrained to return proportional hazards); and the second a Normal Dynamic Frailty Model. The estimated survival under both models are very similar (figure 18). The estimated covariate effect over time, under the NDFM, is presented within figure 20. The effect of sex, once again is observed to decrease over time, but to a lessor extent than under the NDSM (with no frailties). Furthermore it is observed from this same figure that after accounting for the dependencies, the effect of the covariate has increased. Similarly the effect of sex under the conditional proportional hazards model increased to -1.66, compared to an estimated effect of -0.84 under the proportional hazards model with no frailties. The log-likelihood increased from -315.36 under the conditional proportional hazards model to -313.9 under the conditional Normal Dynamic Survival Model, indicating that the latter provides a slightly better

fit. The estimated frailties for each individual were extremely similar under both models, and are presented in table 12. These estimates also compare well to those estimated in the papers cited above.

Numerical values for the estimated parameters over time are presented within the appendix (table 11, page 209). Simulated values at each iteration of the Gibbs sampler are also presented for a selection of the parameters. The estimated variance of the frailties was 2.12 (variance 0.6) under the conditional proportional hazards model, and 2.21 (variance 0.7) under the conditional NDSM: both indicating that there exists a large amount of heterogeneity within the model. The estimated evolution variances under the Normal Dynamic Frailty Model were 0.059 and 0.03 (variances 0.01 and 0.004). All results within this section were based on 10,000 iterations, with a burn in of 1,000 iterations.

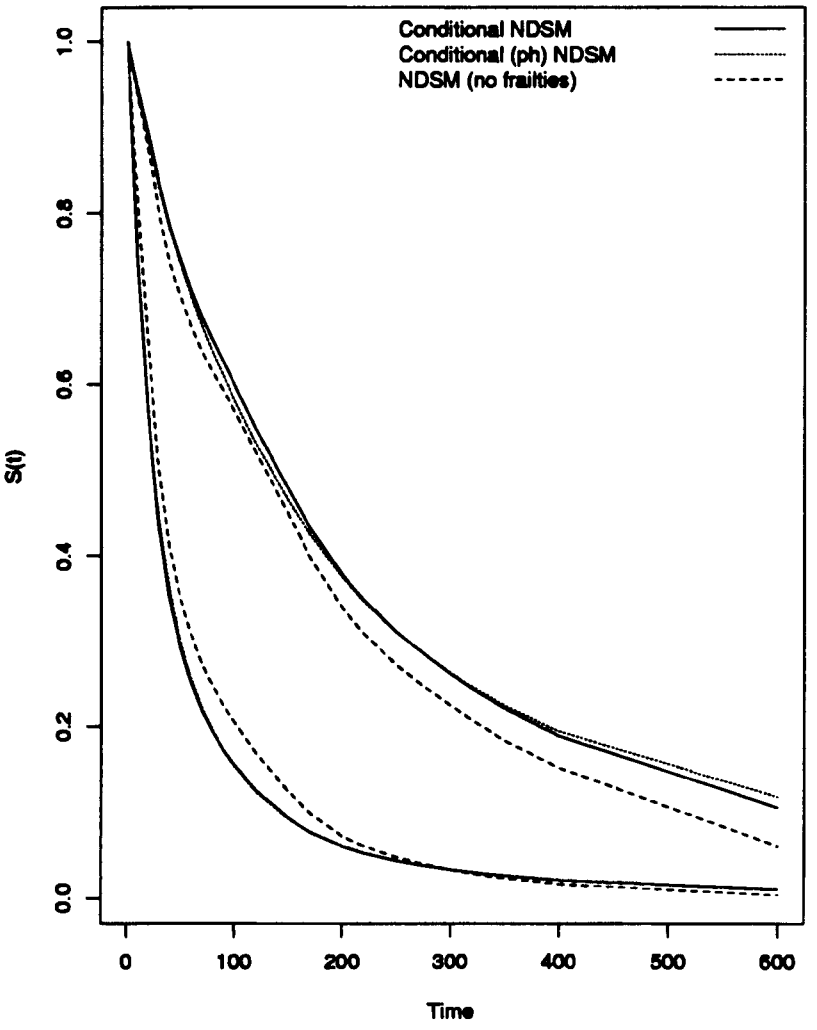

Figure 18: Kidney Data: Estimated survival under the conditional NDSM, the NDSM (without accounting for any heterogeneity), and the conditional NDSM constrained to return proportional hazards. Under all models females have a much better prognosis than males (upper curve for all 3 models).

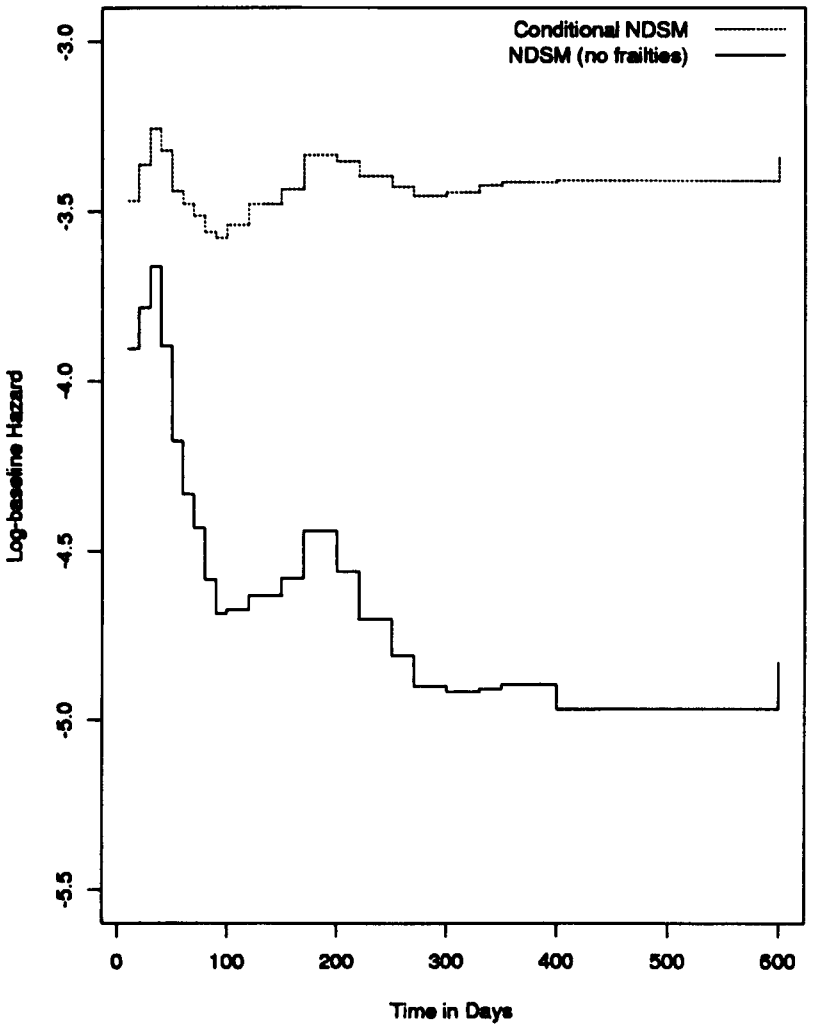

Figure 19: Kidney Data: Estimated log-baseline hazards under the conditional NDSM, and the NDSM without accounting for any heterogeneity. The baseline hazard decreases under the conditional NDSM due to the inclusion of the frailty parameters. Hazards under both models follow a similar pattern.

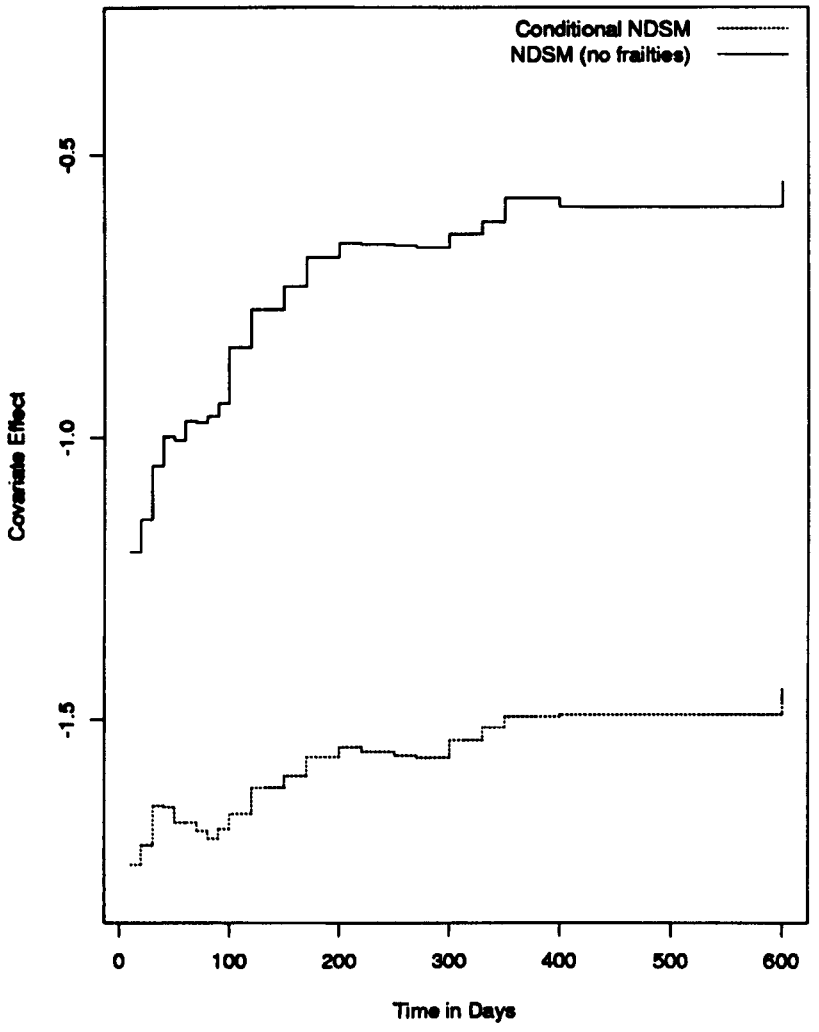

Figure 20: Kidney Data: Estimated effect of the sex parameter under the conditional NDSM, and the NDSM without accounting for any heterogeneity. The covariate effect increases under the conditional NDSM, although under both models the effect follows a similar pattern over time.

6.3 AIDS Data

Hemophiliacs became at risk of receiving contaminated (by HIV) blood samples during the late 1970s to the mid 1980s, when the introduction of heat treating and screening blood, virtually eliminated the risk of future contaminations. Blood samples from hemophilia patients, from several clinics in the United States and elsewhere, were stored during the 1970s and 1980s, for reasons completely independent of the HIV virus (Brookmeyer and Goedert,1989). Also available are blood samples from a group of homosexual men (another high risk group), participating in a cohort study for a Hepatitis B vaccine.

All of these data sets permit the time of HIV infection to be determined up to an interval. Once the patients were found to have been infected with HIV, they were then followed over time, and data are available on the times of their progression to AIDS (either recorded as interval censored or exact).

The data set considered here, is the one previously analysed by Kim *et al.* (1993), who used a proportional hazards model for the incubation distribution. There are two groups of patients, divided by the type of treatment that they received: heavy and light (which reflects the severity of their hemophilia). The data may be found in tables 14 and 15 (pages 213 and

214). For the times of HIV infection time zero represents 1978 and the data are recorded in 6 month units; so that an infection time of 20 uints refers to 1988.

Possible covariates which may effect either or both of the infection and incubation distributions are age, type and severity of hemophilia (which will effect the amount of replacement blood clotting factor given). Treatment centre or geographical location is not a factor in this data set, as all of the patients are from the same treatment centre. Treatment center may be an important factor in other data sets, as it is possible that some treatment centers may have received contaminated factors before others (although in the USA there are only a handful of manufacturers of the replacement blood clotting factor, which distribute across the whole of the country). Kim *et al.* (1993), found that treatment type was a significant factor, but did not find age to be. In the analysis on a similar data set, Brookmeyer and Goedert (1989), found that both age and treatment type were influential factors on survival. They commented that initially treatment center was influential, although after accounting for the other two covariates, the effect was found to be slight.

Preliminary analysis of survival curves for both the initiating and termi-

nating event times are given in the appendix. Treatment type is clearly a significant covariate for both the initiating and terminating event distributions. Age is also a possible significant covariate. It was included within this analysis, firstly as it is interesting to consider its effect, and whether does have a significant impact, and secondly as it is the only data set considered within the thesis that has more than one covariate. It is noted that age was found to be an insignificant covariate in the analysis of Kim *et al.* (1993); this could possibly be because their model was restricted by the proportional hazards assumption. For the initiating event the Grambsh and Therneau (1994) estimate of the treatment effect over time, suggests that the parameter decreases in an almost linear form. For the incubation distribution however the estimate is remarkably different, almost quadratic in shape (although for this parameter the chi-squared test is not statistically significant). The Grambsch and Therneau (1994) test indicates that the effect of age does not vary over time.

6.3.1 Analysis of Times of Infection

The estimated survival curves for the time of the initiating event is presented in figure 22. As indicated in the preliminary analysis, the lightly treated group have a much increased chance of survival. There appears to be a slight difference between the age groups, under both treatment types, with with

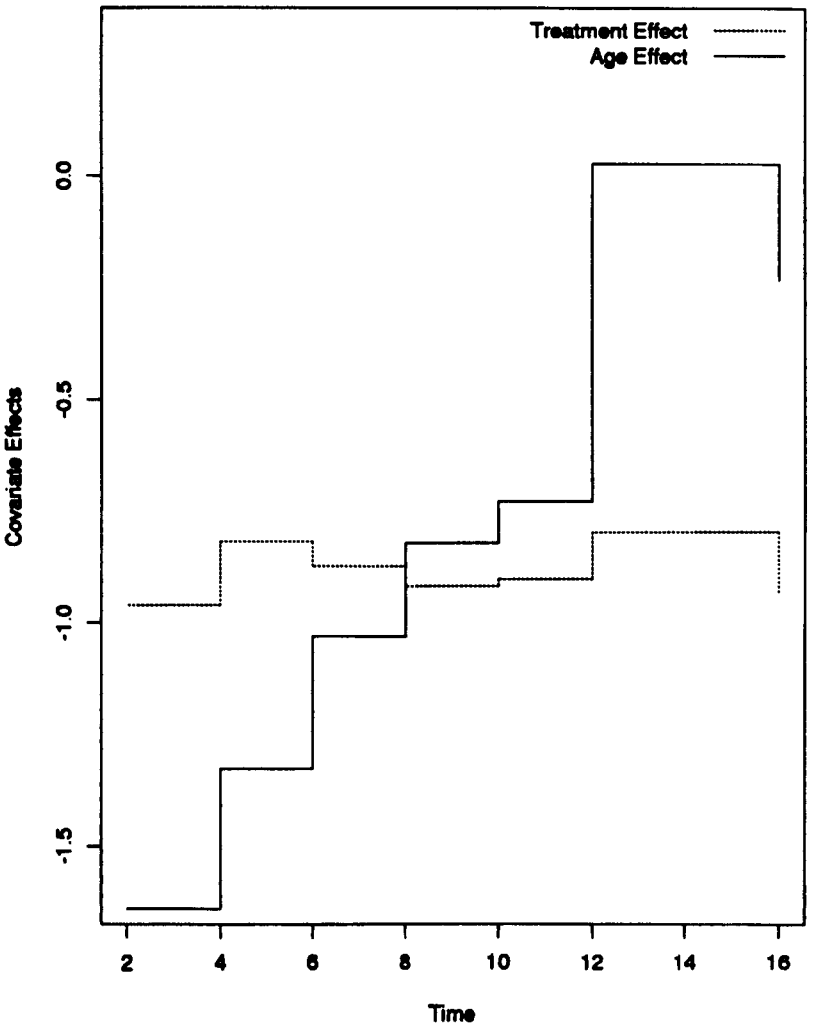

Figure 21: AIDS Data Initiating Event: The negative treatment effect (for time to infection with HIV) indicates that heavily treated patients have a poorer prognosis than those who are lightly treated. The negative value of the age effect indicates that those who are in the younger age group have a poorer prognosis.

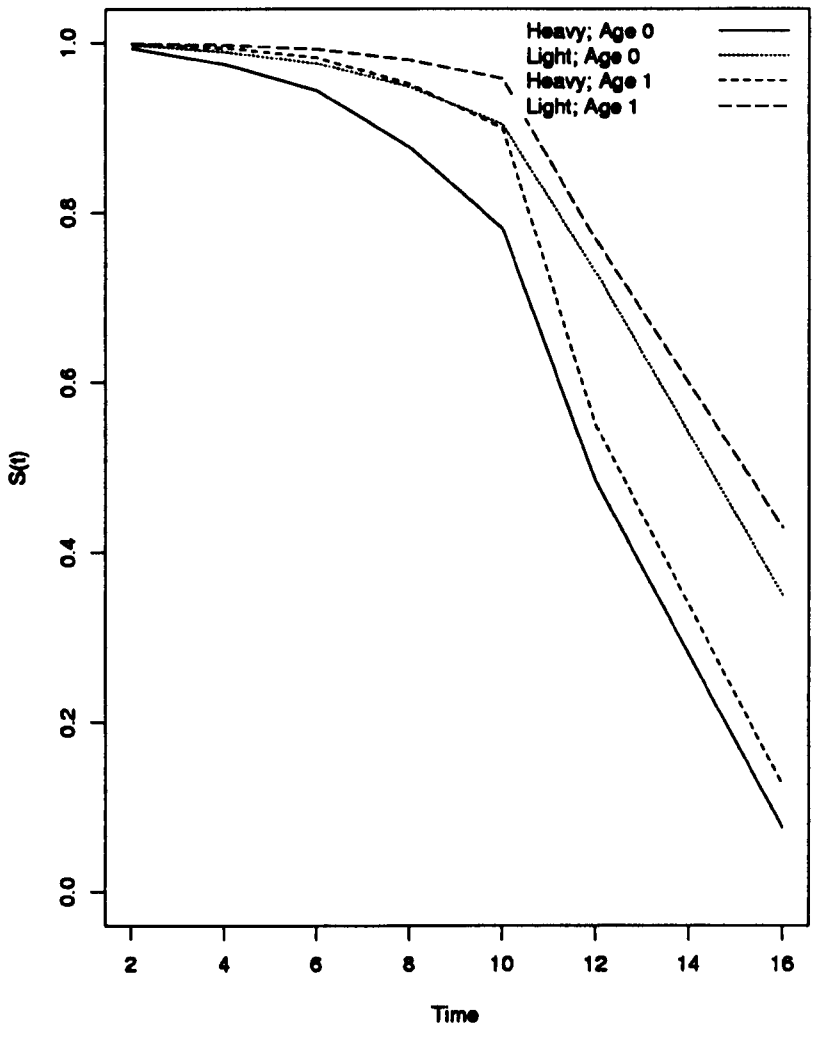

Figure 22: AIDS Data Initiating Event: Estimated survival for all combinations of treatment type and age group, for time to infection with HIV. Age 0 refers to the younger group; and heavy refers to the heavily treated group.

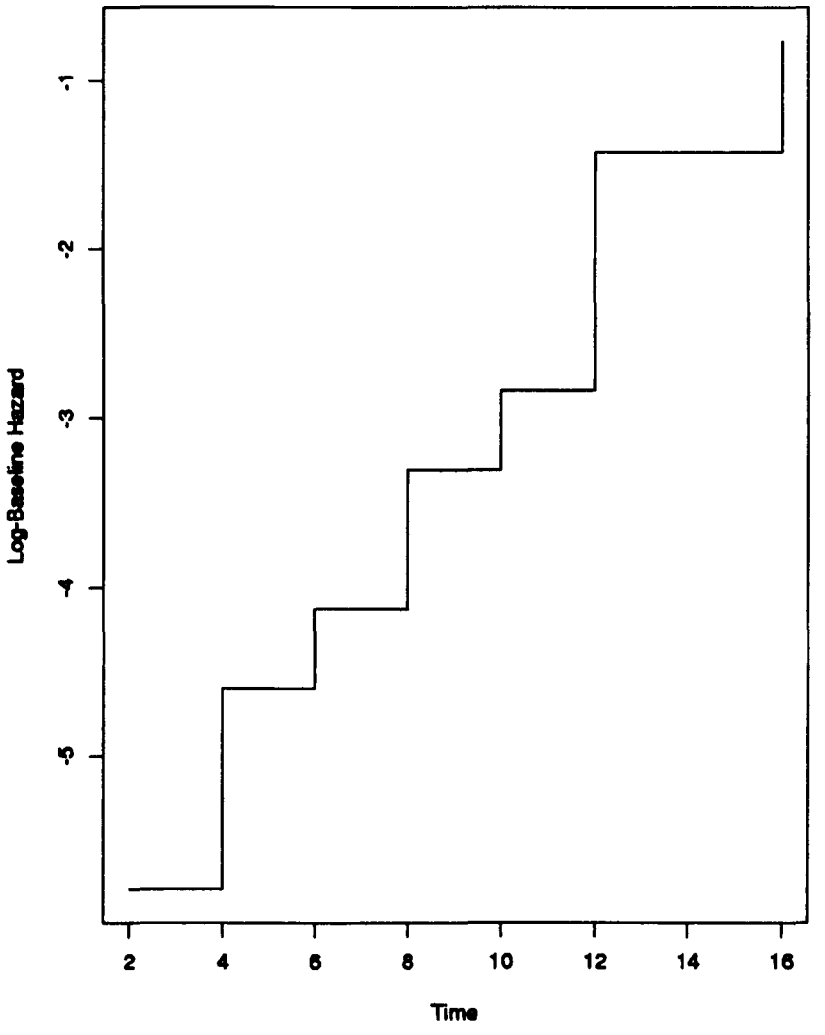

Figure 23: AIDS Data Initiating Event: The estimated log-baseline hazard for time to infection with HIV suggests that the hazard increases continuously over time.

those in age group 1 (older than 20 years) having the slightly better survival rate. The most noticeable differences between the age groups occurs under the heavily treated group, at early times. The estimated log-likelihood without including age was -384.69, and when including age increased slightly to -383.4.

The estimated plots of the both the treatment and age effect over time is presented in figure 21. The treatment effect appears to be fairly steady at around -1, showing no particular increase or decrease. The age effect on the other hand, decreases sharply with time, until at around time 10, when its effect has reduced to almost zero. The estimated covariate effects over time, do not show as much similarity to the Grambsch and Therneau (1994) plots as was observed within the other data applications. This is very probably because none of the estimates are remarkably non-constant, and small changes are more difficult to detect. A plot of the log-baseline hazard is given in figure 23. The hazard of developing HIV increasing continuously for all combinations of covariate groups, which is consistent with results cited in Brookmeyer and Goedert (1989), and also confirms that the Weibull distribution is possibly an adequate model for this data set.

All results cited were based on 10,000 iterations (1,000 burn in) and us-

ing the exact interval censored approach. The imputation method (Chapter 5) was considered, and results found to be very similar.

6.3.2 Analysis of Incubation Time

Again both of the covariates age and treatment were included within the analysis. The estimated survival is presented in figure 24. As with the times of infection, there exists a large difference between the treatment types, and a much smaller difference between the age groups. The most noticeable differences between the age groups is at later times, although it is at these times that the amount of information reduces, and the standard errors of the estimates increase.

The estimated treatment and age effects are given in figure 25. The effect of age is concentrated around zero, except at later times. The treatment effect follows a quadratic like shape, consistent with the Grambsch and Therneau (figure 44) estimate. Disregarding the estimate after time 12, when the variability of the estimate is high, the conclusion would be that treatment has a decreasing effect. It would be interesting to consider this hypothesis, with a more recent data set, where estimates made after time 12 would be more reliable. The estimated hazard (figure 26) suggests that the hazard of developing AIDS reduces 7-10 years after infection with HIV, although this

is almost certainly a result of very few deaths during the later years.

Again a selection of the Gibbs output is given in the appendix (figure 46, page 224), the results within this section are based on 100,000 iterations using the exact approach. A sample of the numerical output from both the exact and an approximate likelihood method are given in tables 17 and 18. There is only a small difference between estimates, with those produced using the approximate method having more often than not a slightly larger variance than compared to the same parameter estimated using the exact approach (although this is not as consistent as under the interval censored comparison, section 6.1). There was quite a large difference between the log-likelihoods under both methods (-914 and -1096), although this was mainly a contribution from later time parameters where uncertainty was large, after saying this it is noted that the difference in survival between the two methods was small.

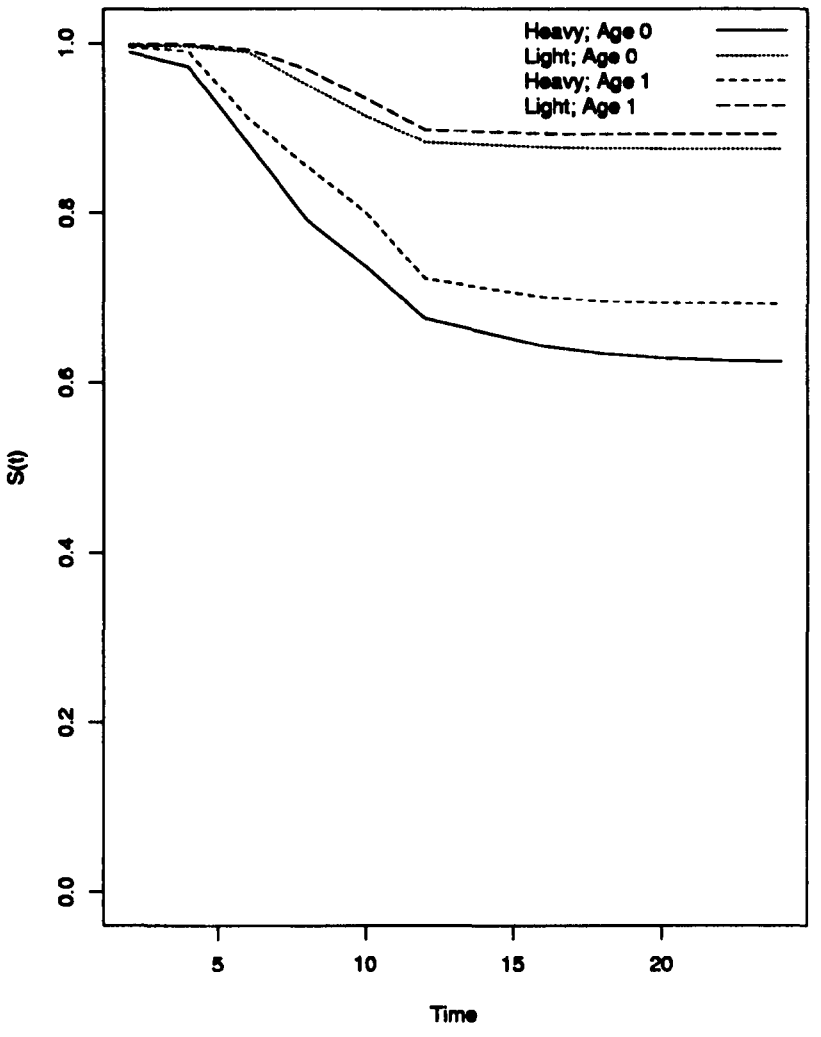

Figure 24: AIDS Data Terminating Event: Estimated survival for all combinations of treatment type and age group, for time to diagnosis of AIDS. Age 0 refers to the younger group; and heavy refers to the heavily treated group.

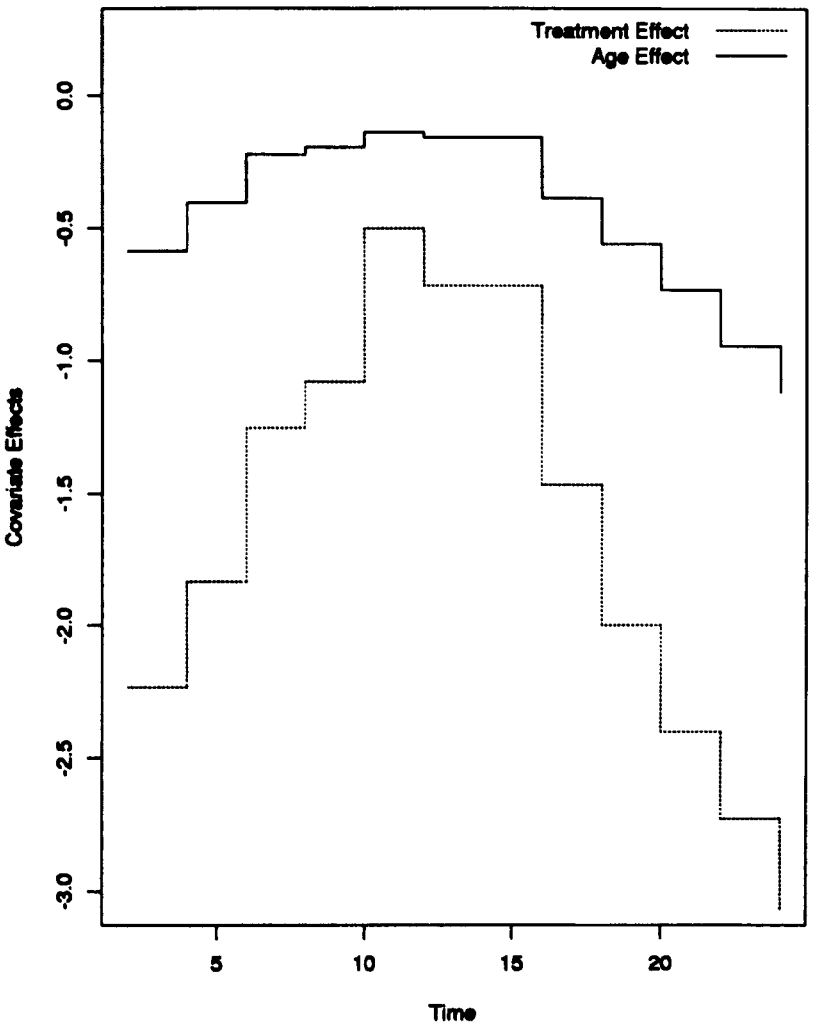

Figure 25: AIDS Data Terminating Event: The negative treatment effect (for time to diagnosis of AIDS) indicates that heavily treated patients have a poorer prognosis than those who are lightly treated. The negative value of the age effect indicates that those who are in the younger age group have a poorer prognosis.

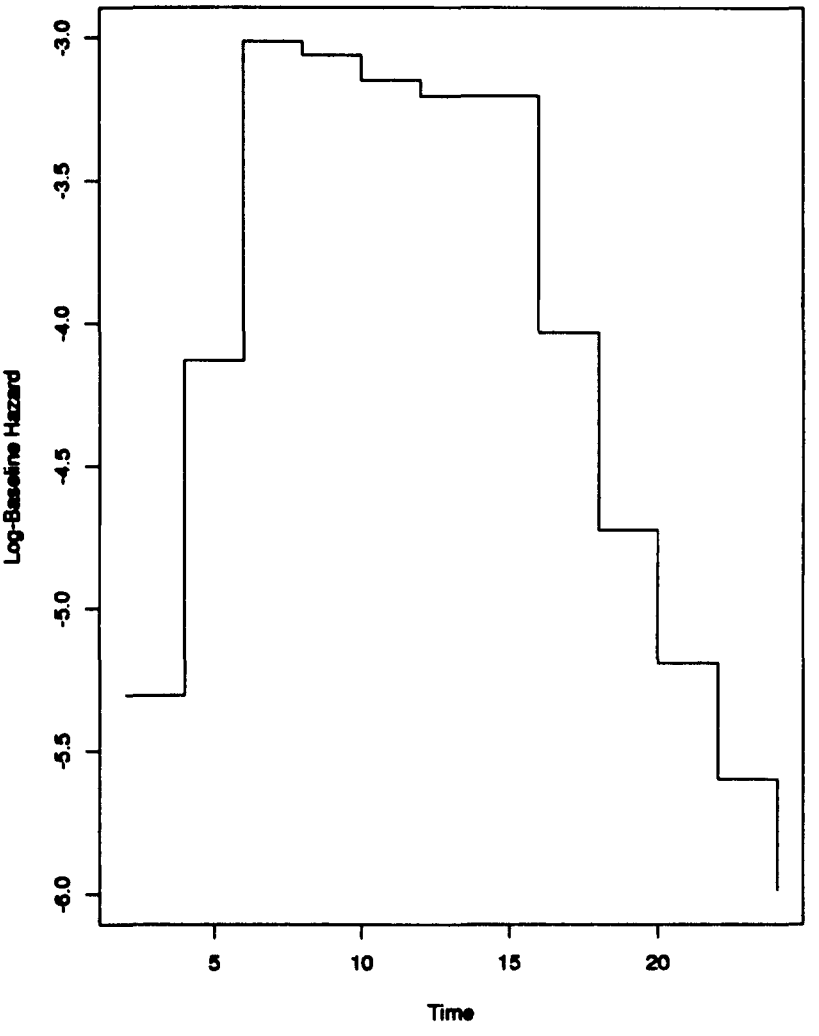

Figure 26: AIDS Data Terminating Event: The estimated log-baseline hazard for time to diagnosis of AIDS suggests that the hazard increases continuously until time 10. After time 10 the hazard decreases, this is almost certainly a result of lack of follow up.

7 Conclusions and Discussion

Piecewise correlated functions are by no means new to survival analysis, although their use seems to have been limited to modelling the baseline hazard function in a proportional hazards setting. The immense flexibility of the proportional hazards model should not be underestimated, but not all data frames meet this assumption. Gamerman (1987) sought to rectify the situation, by introducing a non-proportional hazards model, based on a piecewise-constant (correlated) process modelling both the baseline hazard and covariate effects. Constraints of the method of model fitting used, mean that the model is not easily adapted to handling additional types of censoring (interval censoring and double interval censoring), and many of the multivariate models for such data types are again restrained by the proportional hazards assumption.

It is easy to see the appeal of a model which retains the flexibility of the proportional hazards model, but which will also return non-proportional hazards when the assumption is not appropriate. The fundamental principals underlying the model developed, the NDSM, are fairly straightforward. The model fitting techniques less so, and unfortunately require in-depth programs before they can be put into practice. In some of the data applications the model produced remarkably different results to those obtained by using a

proportional hazards model (in terms of predicted survival curves). In other applications, only slightly different results were observed; although even in these cases the covariate effect was found not to be constant. The dynamic covariate effects may be of interest in themselves, and the models have a wider use than merely estimating the survival. The techniques developed here could for instance be extended into a formal hypothesis test for proportional hazards. Alternatively the effect of the covariate over time may be of interest to a medical investigator.

The appeal of the model is possibly over shadowed by the complicated nature of the estimation involved. In the past models were undoubtedly restricted by computational power, and the model developed here is certainly complicated and time consuming to program. But today computational power is not such an issue. Computation times ranged from several minutes (a right censored data application with 90 observations) to several days (exact double interval censored approach with 257 observations). What is important however, whatever the method of computation, is that accurate estimates should be obtained. This can be done be combining common sense, experience, and the use of convergence diagnostics, all increasing the accuracy and stability of the estimates.

The non-statistical medical world is already applying the new Bayesian techniques of MCMC to real data applications. And in the case of Bayesian smoothed incidence maps, it is now a valuable tool (avoiding small number problems involved with the older methods). The models developed here, have an almost equal potential: the full conditionals are log-concave and so BUGS (Spiegelhalter *et al.*, 1996) could be used as in Bayesian mapping (with the exception of the double interval censored model). Perhaps with the continuing developments of MCMC in Bayesian analysis, the models developed in this thesis will in time become easier to implement.

Extensions of the NDSM

Because of the approach used, some parametric assumptions had to be made (otherwise Gibbs sampling could not have been used). In this thesis normal and gamma distributions were used. These distributions were chosen because of their interpretability, conjugate properties (reducing to standard full conditionals and tractable marginals), and wide use in similar applications: West and Harrison (1997) used normality in Bayesian time series modelling, and Clayton and Cuzick (1985) amongst others, have modelled frailties by the gamma distribution. This is not to say however that other parametric distributions could not have been used, and their impact should not be underestimated,

For the frailty model clear extensions exist in the type of distribution modelling the frailties. The gamma distribution was used here mainly because it is conjugate to the likelihood, thus providing a tractable marginal distribution. If the integration involved in computing the marginal distribution was not tractable, it would be possible to use Monte Carlo simulations to calculate the posterior moments (perhaps using MCMC). But including additional Monte Carlo simulations, would once again increase the computational times involved within the model fitting. The gamma distribution has often been applied successfully in frailty modelling, and so it can be used here fairly confidently without extensive investigation. The major aim of this thesis was to apply a non-proportional hazards model and not to investigate frailty modelling, but a very interesting continuation of the work developed here would certainly be to investigate the impact of the frailty distribution on the marginal hazards. Hougaard (1986) and others have considered using the positive stable frailty distribution, as it has the added advantage of modelling a hazard which is proportional both conditionally and marginally. It would be interesting to consider the implications of using the positive stable frailty distribution within a NDSM. Specifically would the positive stable frailty distribution model a conditional Normal Dynamic Survival Model both conditionally and marginally?

One further point worthy of discussion under the topic of frailty models are dynamic frailty effects. For a reasonable sized group it may not be a valid assumption that the frailty effect is constant across the time axis: just as covariate effects change over time, so too may the frailty effects (they can after all be interpreted as unobservable covariates). Unfortunately with the applications considered in this thesis, there do not exist sufficient observations within each group to consider dynamic frailties.

Frailty models are often interpreted within the statistical literature to model either unobservable covariates or to account for non-independent observations. The fundamental difference which exists between usual covariates and "unobservable covariates" modelled as frailties, is that usually there exist only a small number of observations to estimate each frailty, and secondly inferences focus on the marginal distribution. But interest could also lie within the marginal distribution, in many survival applications which would not always be considered to be frailty models. Suppose that a data set consists of a group of cancer patients, some of whom have participated within screening program, and some that have not. Along with an indicator of screening, the data set would almost certainly consist of various other covariates, such as age and deprivation score to name just a couple. If interest

focuses on whether those screened have a higher survival rate than those not screened: it will then be necessary to compute the marginal distribution, i.e. integrating out the effect of all other covariates. For a large data set, the covariates such as age etc. could be included as frailties: given that there would then be a larger amount of data to estimate each of the frailty effects, it would then become possible to consider including dynamic frailties. Indeed Paik *et al.* (1991) have considered applications of dynamic frailties. Such an application as the one mentioned above introduces one further issue: multivariate frailties (i.e. age and deprivation), yet again another area of interesting research.

Leaving aside parametric assumptions, one of the other fundamental issues concerns efficiency and accuracy of the model as it stands. The accuracy of estimate obtained should be an integral part of any model: confidence may certainly be placed in the estimates produced for all of the data applications considered, but care was always essential: if the chain had failed to accept a sufficient amount of values (or had accepted too many), then it would have not converged. A more efficient type of MCMC algorithm may be of benefit.

A further improvement, with estimation, not necessarily concerning accuracy, but rather efficiency, would be to develop a more efficient approximation to

the double interval censored likelihood. Taking summations rather than integrating, essentially means that a form of numerical integration has been used. There exits much research and investigation on how best to choose the summation points, although the points used within this thesis were not chosen in any of these ways. Including an efficient form of numerical integration could be of great benefit: as probably the major disadvantage of the model as it stands is the computation involved with the double interval censored model.

Perhaps one of the other most natural extensions of the Normal Dynamic Survival Model would be to include the division of the time axis as an unknown parameter to be estimated (as did Arjas and Gasbarra, 1994). Different divisions of the time axis were considered within the data applications in the thesis, and in the majority of cases, as long as divisions were sensible, then the impact appeared to be small. Although in one application, the right censored simulated data, it was noted how the survival was over predicted under a division with too few points at early times. Bayes factors could be used, although would increase computational times; perhaps the method advocated by Arjas and Gasbarra (1994), could have avoided this potential problem.

Further Applications

There exist several interesting areas of further application, which particularly concern the AIDS data set.

One assumption made within the AIDS data analysis, is that the distribution of the incubation time is independent of the chronological time that an individual developed HIV. This becomes increasingly unrealistic with the introduction of AZT, a treatment which has been shown to increase the incubation period of the virus. Furthermore virus strains are also known to vary over time, and different strains could have differing incubation periods. One possible way to avoid this assumption is to include a covariate, with time varying effect, which would represent the chronological time of infection. Following naturally from the AIDS applications within Chapter 6, the NDSM could be extended to accommodate truncated data. Many of the more recent data sets which may be used to estimate the incubation period of the HIV virus are truncated (section 2.9). Using the NDSM, the likelihood may be derived simply by extending the likelihoods which were developed within Chapter 5. Suppose that each observation is truncated on the left (similar extensions exist for right truncation) by a_i where $a_i \in I_{\tilde{a}_i}$. Then the likelihood, once again as a factorisation over individual contributions, will

be:

$$\begin{aligned} L(T) &= \prod_{i=1}^{n} p(T > t_i | T > a_i) \\ &= \prod_{i=1}^{n} \frac{p(T>t_i)}{p(T>a_i)} \\ &= \prod_{i=1}^{n} \frac{A(1,m_i,\beta)C(t_i,t_{m_i},\beta)}{A(1,\bar{a}_i,\beta)C(a_i,t_{\bar{a}_i},\beta)}. \end{aligned}$$

It appears that such a likelihood could be readily incorporated into the model fitting techniques developed in Chapter 6. Additional care would be needed to ensure sufficient data within all of the intervals (to avoid the danger of estimates being based on little or no information).

7.1 Summary

In brief, the dynamic Bayesian survival model (Gamerman, 1987) has been re-analysed, changed slightly and re-applied to the gastric cancer data. The motivation behind this work, was to create a model which could include interval censoring, double interval censoring and non-independent observations. The model developed, the Normal Dynamic Survival Model, has successfully been applied within a range of data applications. The model, when applied to right censored data, although not producing a remarkable improvement on that developed by Gamerman, does have some advantages: it introduces the evolution variance as a hyper-parameter; avoids the linear Bayes approximation, by using the more up to date technique of MCMC; and avoids having to carry out both an online and retrospective analysis. The model requires

specific programming, but then so too does Gamerman's version, and in fact the computation involved within the right censored model is moderate (10 minutes for 10,000 iterations). More importantly, the NDSM was extended within Chapter 4, to accommodate interval censoring, double interval censoring, and non-independent observations, which Gamerman's model can not do. A non-proportional hazards model thereby exists to model data applications which were previously restricted to either non-parametric methods or proportional hazards. The range of data which can now be modelled using a non-proportional hazards technique has therefore been extended, and inferences can thus be drawn on the effect of the covariates over time.

A Data Frames and Preliminary Analysis

A.1 Gastric Cancer Data

The data frame represents the survival times of 90 gastic cancer patients. Right censored observations are denoted by $*$. All survival times are recorded in days.

Treatment	Survival Time
Chemotherapy and radiation	17,42,44,48,60,72,74,95,103,108,
	122,144,167,170,183,185,193,
	195,197,208,234,254,307,315,401,
	445,464,484,528,542,
	567,577,580,795,855,1174*,1214,1232*,
	1366,1455*,1585*,1622*,
	1626*,1736*
Chemotherapy	1,63,105,125,182,216,250,262,301,
	342,354,356,358,380,383,
	383,388,394,408,460,489,499,
	523,524,535,562,675,676,
	748,778,786,797,955,968,977,1245,
	1271,1420,1460*,1516*,1551,
	1690*,1694

Table 4: Gastric Cancer Data

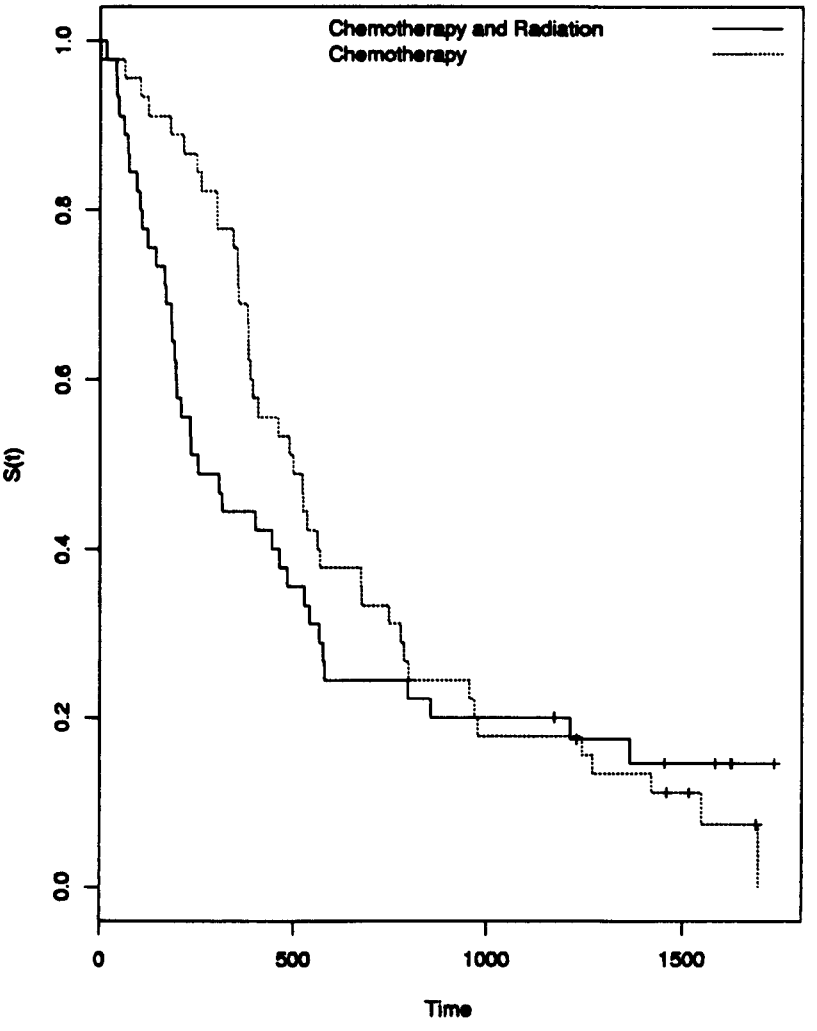

Figure 27: Gastric Cancer: Survival estimated using the non-parametric technique known as the Kaplan-Meier curve.

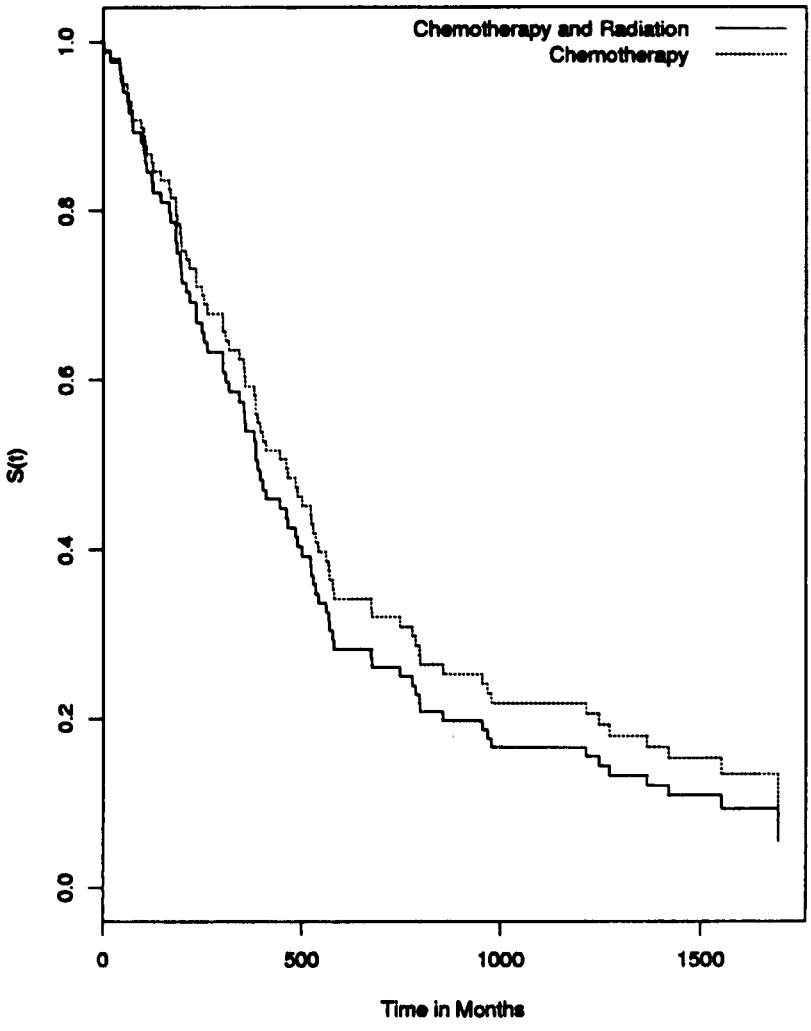

Figure 28: Gastric Cancer: Survival estimated using the proportional hazards model developed by Cox (Cox, 1972).

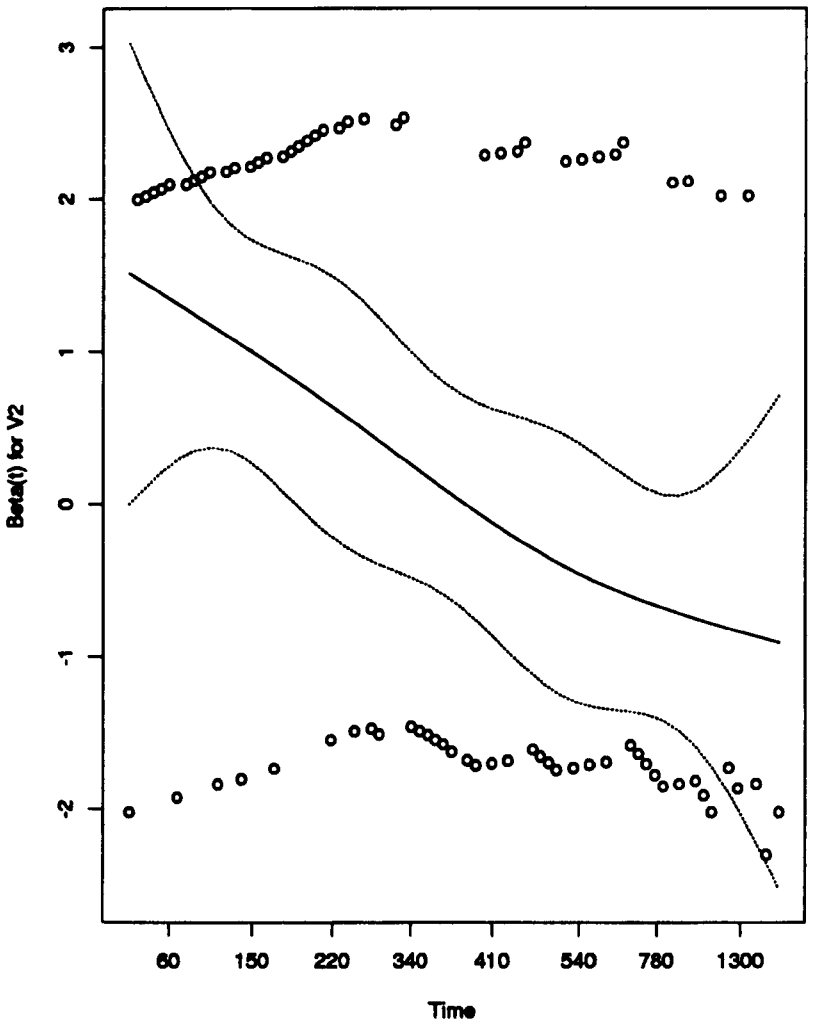

Figure 29: Gastric Cancer: The Grambsch and Therneau (1994) test for proportionality, yields a chi-squared value of 11.1 on 1 degree of freedom. The plot indicates that the treatment effect decreases from a positive to negative effect: chemotheraphy and radiation have initially the poorer prognosis, although change to the better prognosis after around 1000 days.

Time	β_0	var(β_0)	β_1	var(β_1)
100	-6.998	0.140	0.710	0.210
200	-6.802	0.080	0.865	0.154
300	-6.686	0.071	0.412	0.198
400	-6.310	0.102	0.015	0.249
500	-6.298	0.072	0.161	0.243
600	-6.196	0.078	0.175	0.192
700	-6.418	0.105	-0.413	0.424
800	-6.366	0.094	-0.541	0.285
900	-6.530	0.114	-0.687	0.296
1000	-6.519	0.106	-1.012	0.397
1100	-6.670	0.135	-1.135	0.345
1200	-6.631	0.118	-1.126	0.349
1300	-6.399	0.139	-0.959	0.374
1400	-6.370	0.122	-1.024	0.343
1500	-6.290	0.122	-1.370	0.530
1600	-6.186	0.129	-1.599	0.471
1700	-6.118	0.140	-1.76	0.489
1800	-6.119	0.166	-1.893	0.489

Table 5: Gastric Cancer: The estimated parameters in the NDSM based on 10,000 iterations, and division 1 of the time axis (Chapter 3).

Model	Time	β_0	var(β_0)	β_1	var(β_1)
Division 1	100.000000	-6.997919	0.140056	0.710458	0.209706
	500.000000	-6.298003	0.071899	0.161492	0.242679
	1000.000000	-6.518715	0.106141	-1.011860	0.397129
	1500.000000	-6.289463	0.121734	-1.369720	0.529549
	1800.000000	-6.119363	0.165888	-1.893697	0.489560
Division 2	100.000000	-7.198092	0.081542	0.907102	0.231332
	500.000000	-6.167043	0.077111	0.049965	0.209826
	1000.000000	-6.472663	0.090322	-1.034699	0.286311
	1500.000000	-6.238798	0.098772	-1.364340	0.348384
	1800.000000	-6.019131	0.126056	-1.784977	0.353103
Division 3	103.000000	-7.347658	0.069857	1.254830	0.184632
	499.000000	-6.101015	0.069747	-0.250224	0.193807
	977.000000	-6.438448	0.071180	-1.080969	0.195890
	1551.000000	-6.173302	0.075005	-1.541757	0.219938
	1800.000000	-6.036098	0.085855	-1.656929	0.224201
1000 Iterations	100.000000	-6.743143	0.082463	0.475765	0.165553
	500.000000	-6.272270	0.027244	-0.083461	0.135886
	1000.000000	-6.393534	0.036533	-1.257496	0.321202
	1500.000000	-6.289953	0.040743	-1.670100	0.329928
	1800.000000	-6.230968	0.084879	-2.268828	0.372559
10000 Iterations	100.000000	-6.910058	0.117889	0.619454	0.211675
	500.000000	-6.183795	0.062162	-0.118117	0.251824
	1000.000000	-6.430449	0.084848	-1.448498	0.620833
	1500.000000	-6.368153	0.103135	-1.842821	0.667507
	1800.000000	-6.426151	0.118865	-2.648087	0.720709
NDSM (PH)	100.000000	-6.648217	0.066480	0.171631	0.052414
	500.000000	-6.343220	0.045168	0.171631	NA
	1000.000000	-6.734523	0.066716	0.171631	NA
	1500.000000	-6.601112	0.078410	0.171631	NA
	1800.000000	-6.422539	0.098857	0.171631	NA

Table 6: Gastric Cancer: Summary Output

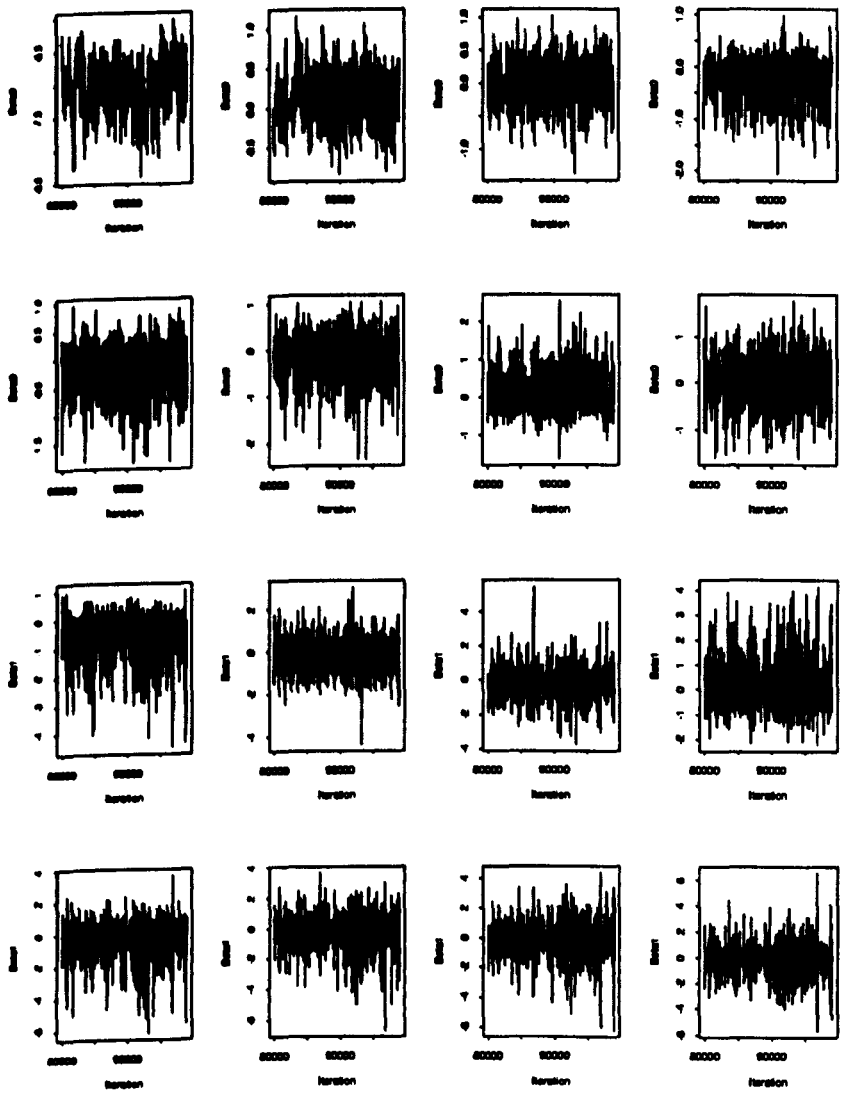

Figure 30: A selection of the simulated values from the Gibbs sampler. The top half of the plot refers to the baseline parameter, and the latter half to the treatment effect.

A.2 Breast Cancer Data

The data set contains 94 remission times for breast cancer patients, with survival times recorded in months. Right censored observations are denoted by $-$, and exact observations by $*$. Treatment 0 will represent radiation and treatment 1 will represent the combined treatment of chemotherapy and radiation. An observation of the form 6,10 refers to an interval censored observation as refered to by the notation $(6, 10]$ within the thesis.

Radiotherapy			Combined		
45*	25,37	37,-	8,12	0,5	30,34
6,10	46,-	0,5	0,22	5,8	13,-
0,7	26,40	18,-	24,31	12,20	10,17
46,-	46,-	24,-	17,27	11,-	8,21
46,-	27,34	36,-	17,23	33,40	4,9
7,16	36,44	5,11	24,30	31,-	11,-
17,-	46,-	19,35	16,24	13,39	14,19
7,14	36,48	17,25	13,-	19,32	4,8
37,44	37,-	24,-	11,13	34,-	34,-
0,8	40,-	32,-	16,20	13,-	30,36
4,11	17,25	33,-	18,25	16,24	18,24
15,-	46,-	19,26	17,26	35,-	16,60
11,15	11,18	37,-	32,-	15,22	35,39
22,-	38,-	34,-	23,-	11,17	21,-
46,-	5,12	36,-	44,48	22,32	11,20
46,-			14,17	10,35	48,-

Table 7: Breast Cancer Data

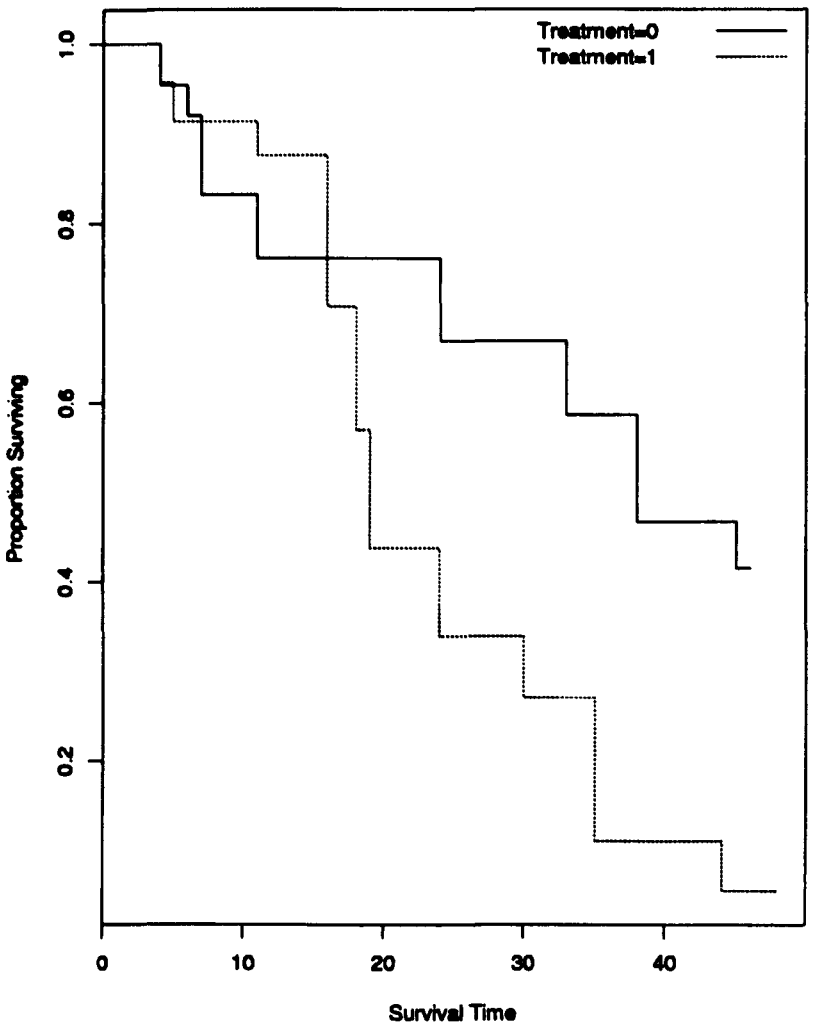

Figure 31: Breast Cancer: Survival is estimated using the non-parametric technique for interval censored data as developed by Turnbull (1976). Radiation is denoted by 0, the combined treatment by 1.

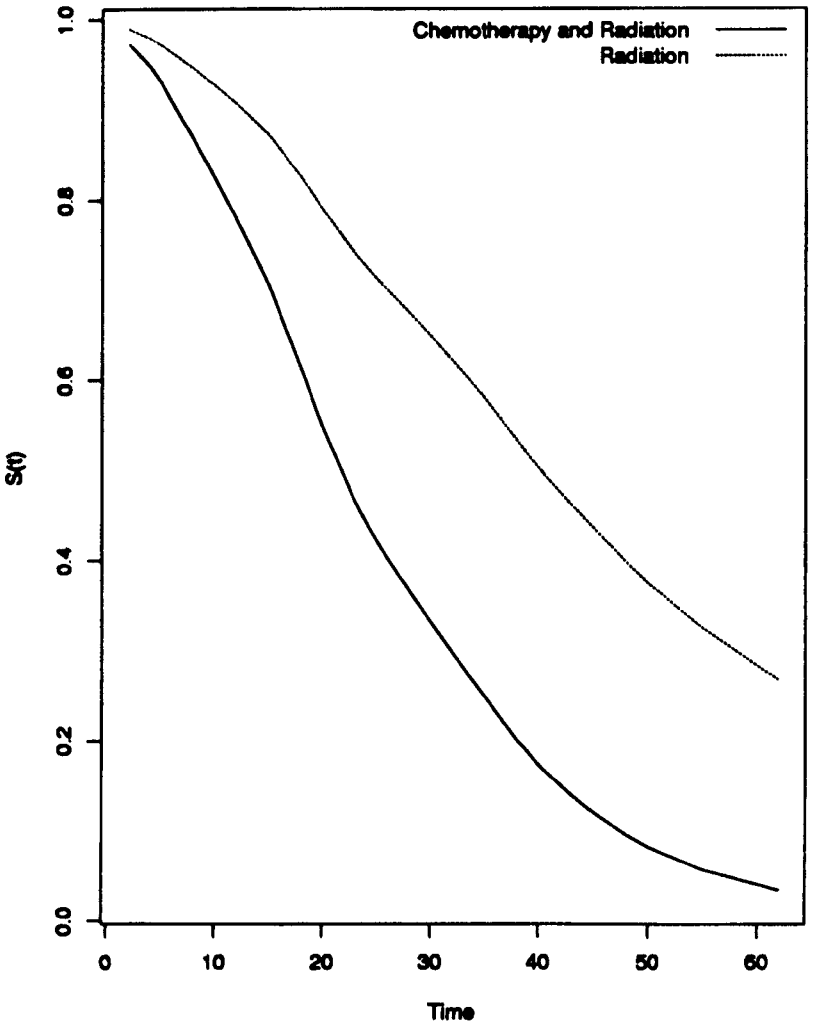

Figure 32: Breast Cancer: Survival is estimated using Cox's proportional hazards model, and the analysis is based on the midpoints of the interval censored observations.

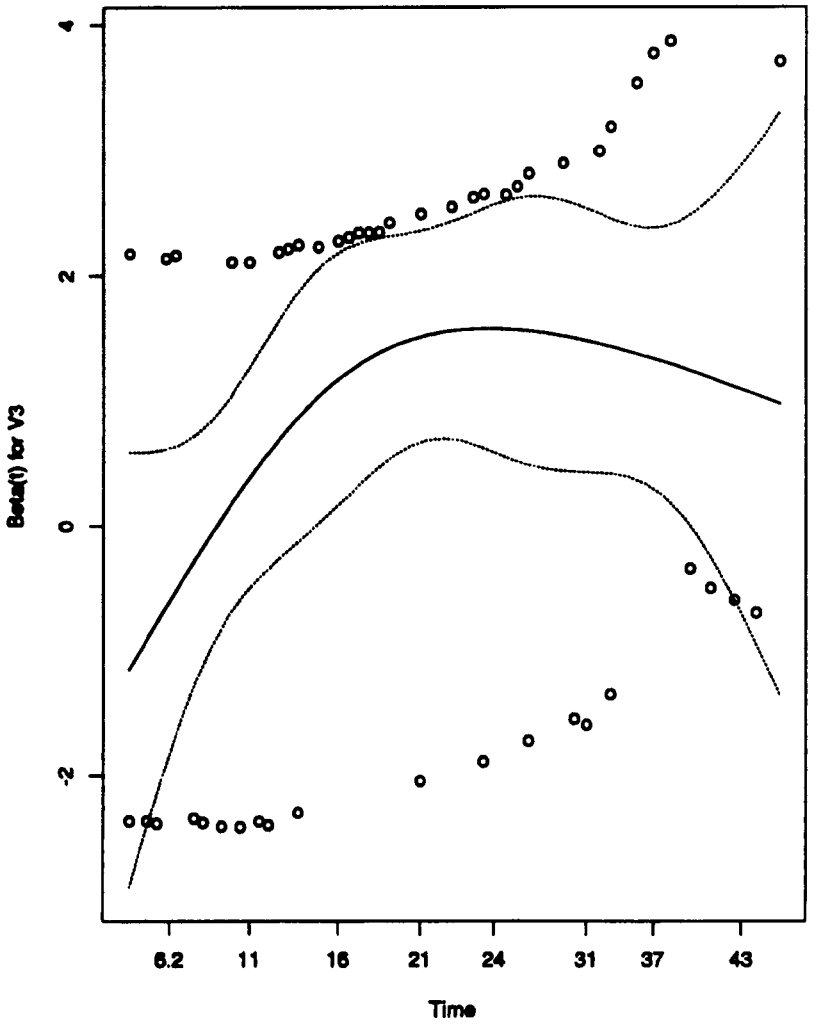

Figure 33: Breast Cancer: The Grambsch and Therneau (1994) test for proportionality, yields a chi-squared value of 4.5 on 1 degree of freedom. The analysis was based on the mid-points. For the short time that the treatment effect is negative, the combined treatment has the better pronosis, although after around 10 months the treatment effect becomes positive and radiation has the better prognosis.

Time	β_0	var(β_0)	β_1	var(β_1)
2.000000	-4.672	0.329	-0.737	1.054
4.000000	-4.551	0.162	-0.591	0.367
5.000000	-4.402	0.228	-0.322	0.451
7.000000	-4.256	0.188	-0.190	0.421
8.000000	-4.147	0.183	-0.0490	0.486
10.000000	-4.230	0.164	-0.014	0.351
13.000000	-4.263	0.142	0.173	0.458
15.000000	-4.296	0.151	0.718	0.320
18.000000	-4.209	0.181	1.372	0.378
20.000000	-4.162	0.167	1.727	0.375
23.000000	-4.226	0.175	1.512	0.425
25.000000	-4.259	0.168	1.334	0.312
28.000000	-4.300	0.155	1.182	0.358
30.000000	-4.225	0.141	1.174	0.422
33.000000	-4.055	0.160	1.242	0.353
35.000000	-3.932	0.176	1.280	0.348
38.000000	-3.745	0.155	1.359	0.360
40.000000	-3.722	0.204	1.115	0.495
43.000000	-3.751	0.191	0.942	0.419
45.000000	-3.728	0.195	0.948	0.419
48.000000	-3.672	0.203	0.957	0.496
50.000000	-3.707	0.216	0.797	0.505
55.000000	-3.736	0.224	0.661	0.502
62.000000	-3.764	0.234	0.494	0.505

Table 8: Breast Cancer: Estimated parameters for the NDSM based on 10,000 iterations.

Model	Time	β_0	var(β_0)	β_1	var(β_1)
Exact	2.000000	-4.672285	0.329380	-0.737070	1.054595
	10.000000	-4.228912	0.164141	-0.013802	0.351617
	30.000000	-4.225524	0.141269	1.174146	0.421863
	50.000000	-3.706744	0.216431	0.796942	0.505562
Imputation	2.000000	-4.645090	0.491784	-1.165036	1.100107
	10.000000	-4.212164	0.268752	-0.194692	0.598277
	30.000000	-4.168136	0.181635	1.084502	0.394358
	50.000000	-3.737493	0.350477	0.810028	0.631038
NDSM (PH)	2.000000	-5.255736	0.247787	0.935386	0.083692
	10.000000	-4.597068	0.213016	0.935386	NA
	30.000000	-3.921287	0.210437	0.935386	NA
	50.000000	-3.520420	0.261745	0.935386	NA

Table 9: Breast Cancer: A sample of the estimated parameters, from the NDSM using the exact approach, the NDSM using the imputation method, and the NDSM constrained to return proportional hazards.

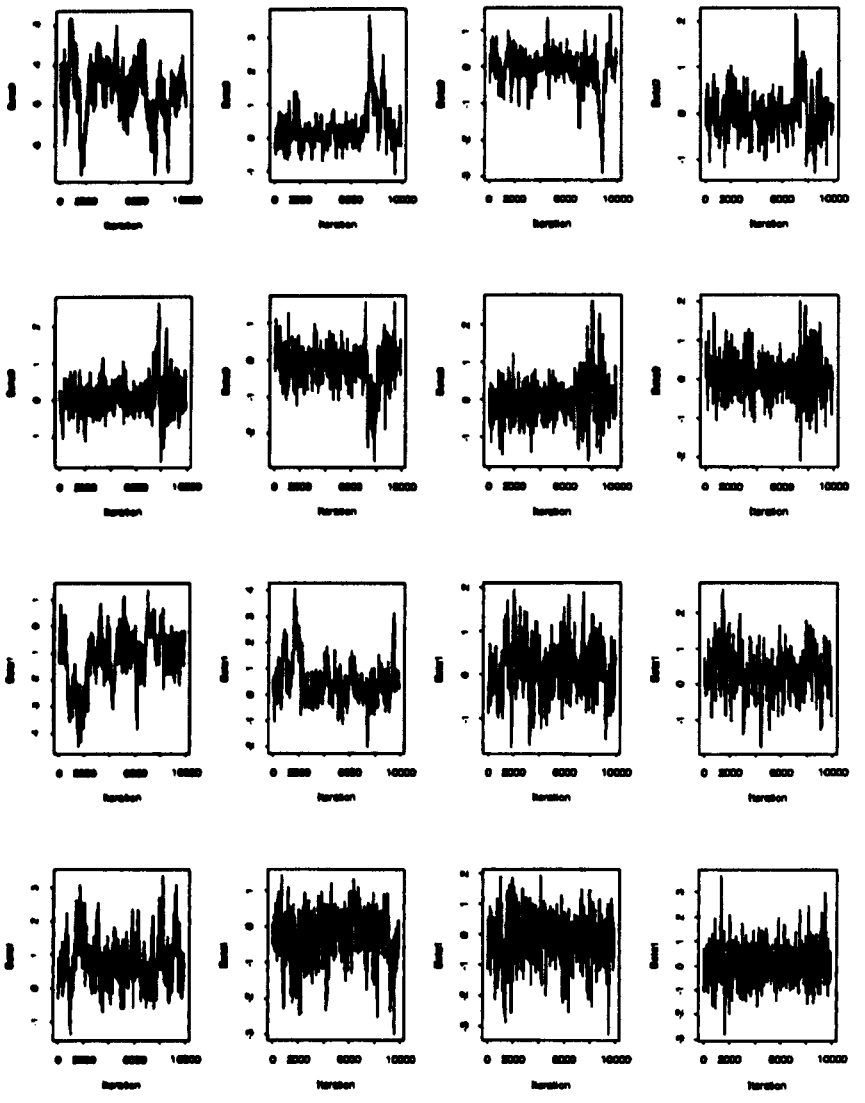

Figure 34: Breast Cancer: A selection of the simulated values from the Gibbs sampler. The top half of the plot refers to the baseline parameter, and the latter half to the treatment effect.

A.3 Kidney Data

The kidney data set contains the times of two infections for 38 individuals, with survivla times recorded in months. Females are denoted by 1 and males by 0. For each observation there exists a right censoring indicator.

Individual	Times	Censoring	Sex	Individual	Times	Censoring	Sex
1	8,16	1,1	0	20	15,108	1,0	1
2	33,13	1,0	1	21	152,562	1,1	0
3	22,28	1,1	0	22	402,24	1,0	1
4	447,318	1,1	1	23	13,66	1,1	1
5	30,12	1,1	0	24	39,46	1,0	1
6	24,245	1,1	1	25	12,40	1,1	0
7	7,9	1,1	0	26	113,201	0,1	1
8	511,30	1,1	1	27	132,156	1,1	1
9	53,196	1,1	1	28	34,30	1,1	1
10	15,154	1,1	0	29	2,25	1,1	0
11	7,333	1,1	1	30	130,26	1,1	1
12	141,8	1,0	1	31	27,58	1,1	1
13	96,38	1,1	1	32	5,43	0,1	1
14	149,70	0,0	1	33	152,30	1,1	1
15	536,25	1,0	1	34	190,5	1,0	1
16	17,4	1,0	1	35	119,8	1,1	1
17	185,117	1,1	1	36	54,16	0,0	1
18	292,114	1,1	1	37	6,78	0,1	1
19	22,159	0,0	1	38	63,8	1,0	0

Table 10: Kidney Data

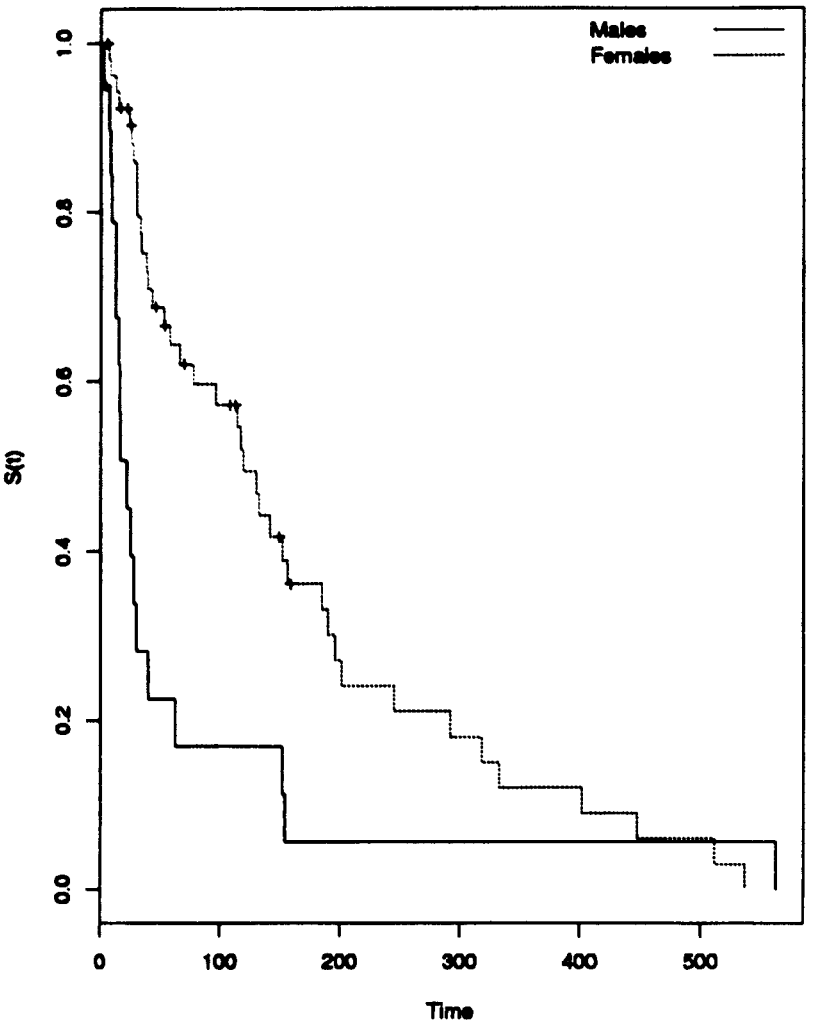

Figure 35: Kidney Data: Survival estimated using the non-parametric Kaplan-Meier method, ignoring the dependencies which may exist within the data.

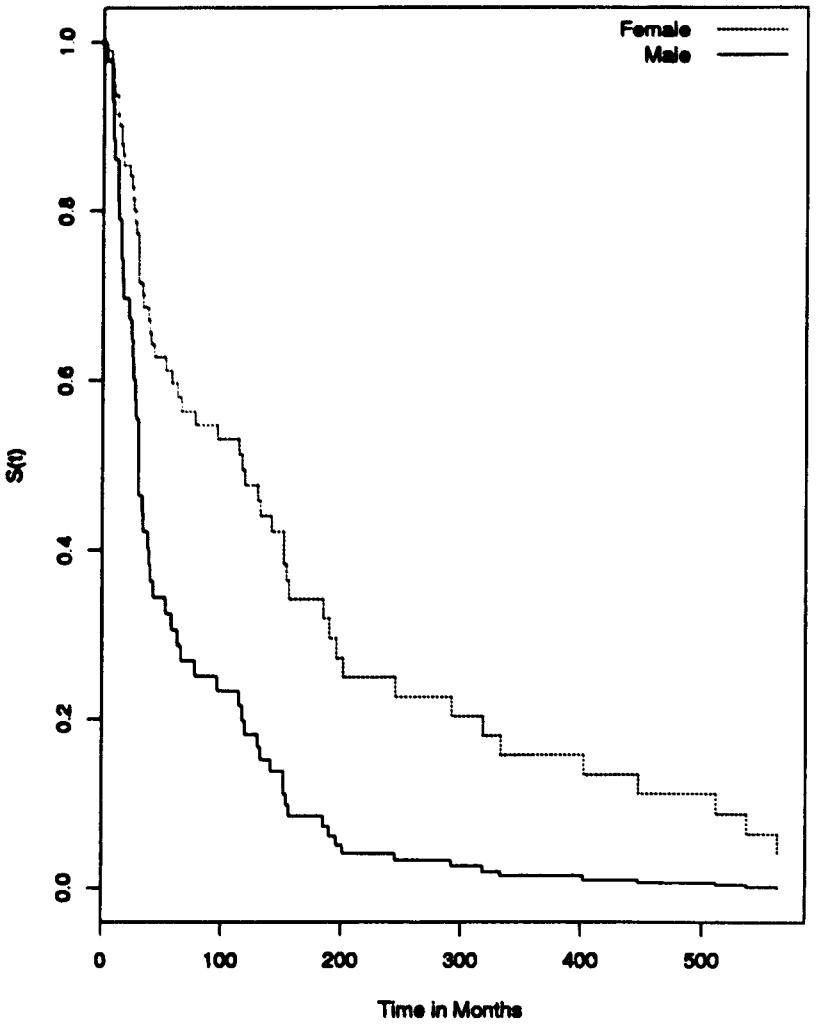

Figure 36: Kidney Data: Survival estimated using Cox's proportional hazards model, ignoring the dependencies which may exist within the data.

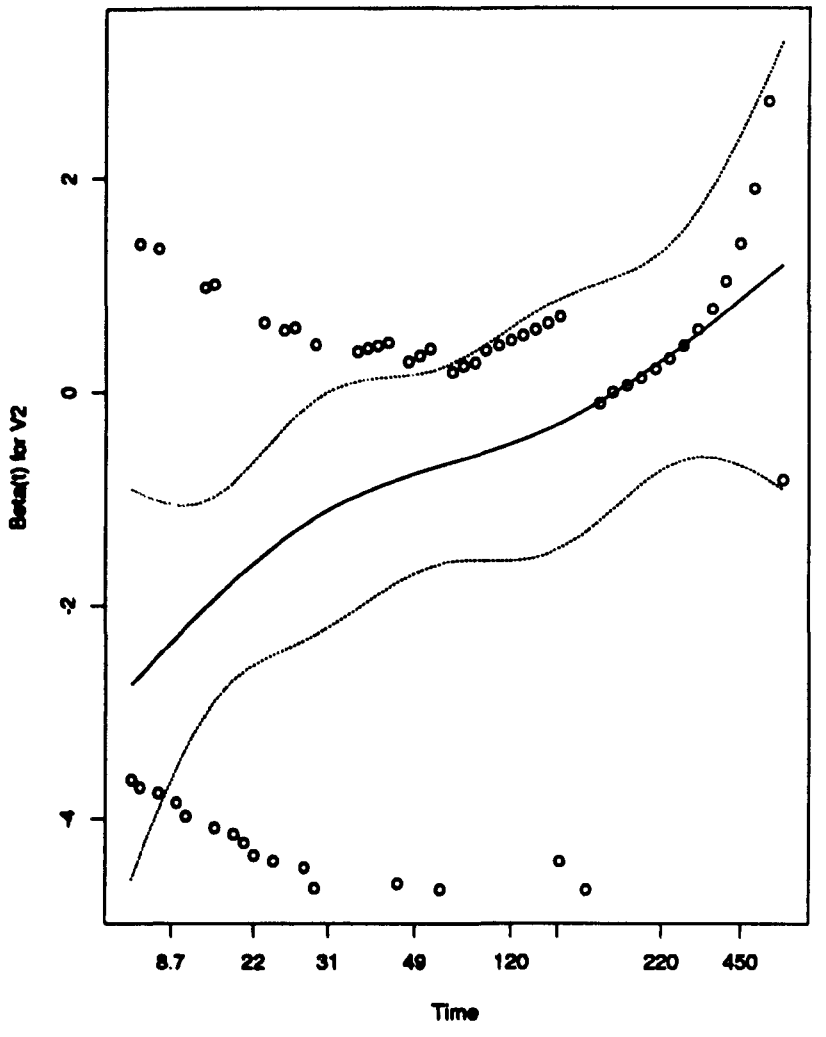

Figure 37: Kidney Data: The Grambsch and Therneau (1994) test for proportionality, yields a chi-squared value of 11.2 on 1 degree of freedom. The analysis ignored any dependencies which may exist within the data. The difference between the sexes decreased over time although always remaining positive: females always have the better prognosis.

	NDSM (Frailties)				NDSM			
Time	β_0	var(β_0)	β_1	var(β_1)	β_0	var(β_0)	β_1	var(β_1)
10.000	-3.470	0.157	-1.757	0.248672	-3.904	0.110	-1.203	0.180
20.000	-3.362	0.061	-1.722	0.024577	-3.784	0.074	-1.145	0.037
30.000	-3.254	0.046	-1.653	0.031744	-3.660	0.076	-1.050	0.049
40.000	-3.319	0.049	-1.655	0.023809	-3.895	0.107	-0.997	0.038
50.000	-3.440	0.080	-1.682	0.028589	-4.175	0.112	-1.005	0.044
60.000	-3.478	0.052	-1.681	0.02988	-4.332	0.105	-0.971	0.038
70.000	-3.512	0.044	-1.696	0.027074	-4.433	0.090	-0.974	0.052
80.000	-3.559	0.048	-1.710	0.034756	-4.584	0.130	-0.964	0.043
90.000	-3.577	0.053	-1.694	0.02616	-4.686	0.101	-0.940	0.049
100.000	-3.538	0.039	-1.667	0.023844	-4.676	0.103	-0.840	0.082
120.000	-3.479	0.047	-1.621	0.027509	-4.632	0.126	-0.773	0.047
150.000	-3.433	0.035	-1.600	0.02503	-4.582	0.103	-0.733	0.043
170.000	-3.331	0.064	-1.566	0.029421	-4.441	0.124	-0.682	0.061
200.000	-3.351	0.043	-1.550	0.026219	-4.562	0.124	-0.658	0.044
220.000	-3.396	0.048	-1.557	0.031787	-4.703	0.121	-0.660	0.051
250.000	-3.429	0.044	-1.565	0.028122	-4.812	0.097	-0.661	0.044
270.000	-3.455	0.053	-1.568	0.033395	-4.901	0.103	-0.665	0.048
300.000	-3.445	0.052	-1.537	0.031981	-4.917	0.106	-0.641	0.043
330.000	-3.424	0.042	-1.515	0.027949	-4.911	0.106	-0.620	0.051
350.000	-3.415	0.045	-1.496	0.030603	-4.898	0.116	-0.577	0.058
400.000	-3.408	0.040	-1.492	0.027319	-4.970	0.136	-0.593	0.060
600.000	-3.341	0.059	-1.447	0.040533	-4.832	0.156	-0.548	0.075

Table 11: Kidney Data: Estimated parameters for the NDSM with frailties, and the NDSM without frailties, both sets of estimates are based on 10,000 iterations.

	Conditional Proportional NDSM		Conditional NDSM	
Individual	u_g	var (u_g)	u_g	var (u_g)
1	1.431	0.566	1.432	0.562
2	1.323	0.617	1.310	0.603
3	1.081	0.344	1.059	0.319
4	0.558	0.093	0.556	0.094
5	1.169	0.374	1.151	0.366
6	1.044	0.296	1.049	0.294
7	1.601	0.729	1.587	0.723
8	0.695	0.136	0.680	0.138
9	1.096	0.316	1.100	0.314
10	0.536	0.106	0.547	0.101
11	0.932	0.240	0.929	0.229
12	1.010	0.356	1.036	0.365
13	1.406	0.541	1.392	0.527
14	0.588	0.192	0.605	0.199
15	0.516	0.106	0.504	0.099
16	1.116	0.461	1.096	0.430
17	1.005	0.269	1.012	0.273
18	0.863	0.207	0.851	0.193
19	0.632	0.254	0.635	0.242
20	1.080	0.414	1.086	0.395
21	0.162	0.014	0.173	0.016
22	0.616	0.142	0.607	0.134
23	1.609	0.720	1.594	0.697

Table 12: Kidney Data: Estimated frailties obtained using the conditional NDSM and the conditional NDSM constrained to return proportional hazards. The output is continued overpage.

	Conditional Proportional NDSM		Conditional NDSM	
Individual	u_g	var u_g	u_g	var u_g
24	1.177	0.485	1.168	0.472
25	1.054	0.323	1.044	0.313
26	0.740	0.197	0.743	0.195
27	1.039	0.289	1.042	0.293
28	1.684	0.815	1.664	0.762
29	1.362	0.522	1.348	0.487
30	1.333	0.487	1.334	0.474
31	1.577	0.705	1.581	0.680
32	1.313	0.614	1.299	0.593
33	1.260	0.444	1.270	0.425
34	0.895	0.274	0.911	0.289
35	1.428	0.553	1.436	0.547
36	0.806	0.416	0.813	0.333
37	1.188	0.499	1.184	0.497
38	0.676	0.179	0.688	0.181

Table 13: Kidney Data: Estimated Frailties (continued)

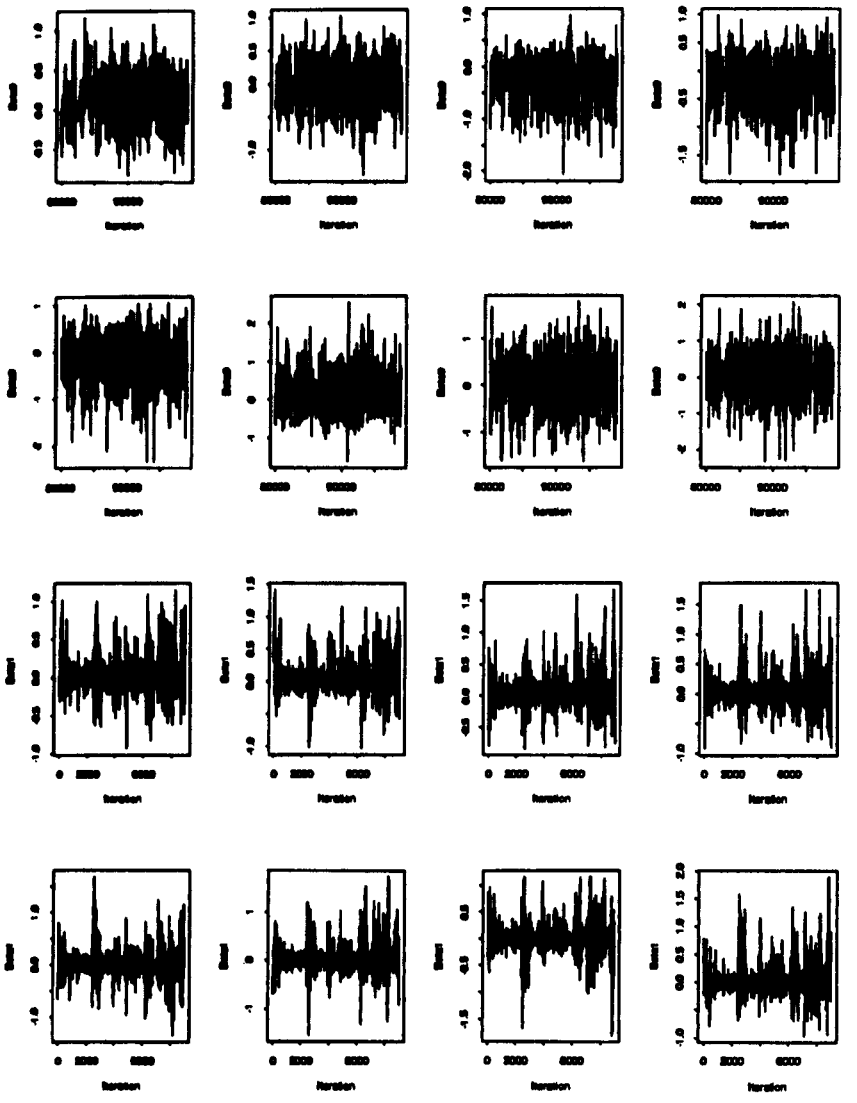

Figure 38: Kidney Data: A selection of the simulated values from the Gibbs sampler. The top half of the plot refers to the baseline parameter, and the latter half to the covariate representing sex.

A.4 AIDS Data

Times	Times	Age	Times	Times	Age	Times	Times	Age
15,∞		0(6)	16,∞		0(15)	17,∞		0(9)
15,∞		1(13)	16,∞		1(16)	17,∞		1
18,∞		1	1,5	23,∞	0(2)	1,11	23,∞	0(2)
1,12	23,∞	1(2)	1,13	23,∞	1	1,14	23,∞	0
1,15	23,∞	0(7)	1,16	23,∞	0	1,17	23,∞	0
1,18	23,∞	0	1,15	23,∞	1(2)	1,16	23,∞	1(3)
3,10	23,∞	0	4,11	23,∞	1	5,11	23,∞	1
6,13	23,∞	1	6,8	23,∞	0	7,12	23,∞	0
7,10	23,∞	0	8,16	23,∞	1	8,12	23,∞	0
8,15	23,∞	0	8,13	23,∞	0	9,12	23,∞	1
9,14	23,∞	1(2)	9,13	23,∞	0	9,14	23,∞	0
10,15	23,∞	1	10,16	23,∞	0	10,12	23,∞	1
10,12	23,∞	0	11,14	23,∞	1(2)	11,15	23,∞	1
11,12	23,∞	1(2)	11,14	23,∞	0(2)	11,16	23,∞	1
12,14	23,∞	1(3)	12,15	23,∞	1	12,13	23,∞	0(4)
12,14	23,∞	0	13,15	23,∞	0(2)	13,14	23,∞	0(4)
14,15	23,∞	0(4)	14,16	23,∞	1	15,16	23,∞	1(3)
13,15	23,∞	1	15,16	23,∞	0(2)	1,12	23,∞	0
1,3	7,8	0	1,12	14,15	0	6,12	16,16	0
7,13	16,17	1	10,11	19,20	0	11,12	20,20	0
13,14	20,20	1	3,16	21,21	1	6,13	21,21	1
8,16	21,21	1	10,12	21,21	0	13,15	21,22	0
13,13	22,22	0	12,13	23,23	1			

Table 14: AIDS Data Lightly Treated Group: The indicator 0 will denote those individuls who are younger than 20, and 1 will denote those who are older than 20. Multiplicities are denoted in brackets. The first interval censored time refers to the time of infection with HIV and the second the time of developing AIDS. All times are recorded in units of 6 months.

Times	Times	Age	Times	Times	Age	Times	Times	Age
15,∞		0(2)	16,∞		1(2)	16,∞		0
17,∞		1(2)	17,∞		0	1,16	23,∞	0
1,7	23,∞	0	1,11	23,∞	0	1,12	23,∞	0(2)
1,13	23,∞	1	1,13	23,∞	0	1,14	23,∞	0(3)
1,15	23,∞	0(2)	5,7	23,∞	1	5,7	23,∞	0
3,14	23,∞	0	3,15	23,∞	1	7,9	23,∞	0
7,10	23,∞	0	7,15	23,∞	1	8,10	23,∞	0
8,15	23,∞	1	9,10	23,∞	0(2)	9,12	23,∞	0(3)
10,11	23,∞	0(4)	10,11	23,∞	1(3)	10,12	23,∞	0(1)
10,15	23,∞	0	11,12	23,∞	0	11,13	23,∞	1
11,13	23,∞	0(3)	12,13	23,∞	0(6)	12,13	23,∞	1
12,14	23,∞	0	12,14	23,∞	1	13,15	23,∞	0(2)
13,15	23,∞	1	13,16	23,∞	0	14,15	23,∞	1(2)
14,15	23,∞	0(6)	14,16	23,∞	0(2)	15,16	23,∞	0(4)
15,16	23,∞	1	1,7	12,13	0	1,7	16,16	1
1,10	11,12	0	3,7	17,17	0	5,7	11,12	0
1,15	20,21	0	7,9	19,19	0	8,10	15,15	0
5,8	13,13	0	9,11	17,18	1	9,12	17,18	0
9,13	14,15	0	9,12	22,22	0	10,11	14,15	0
9,13	17,18	0	10,12	15,16	1	10,14	15,16	1
10,11	15,16	0	10,12	23,23	0	12,13	20,20	0
10,12	16,17	0	13,15	17,18	0(2)	13,14	19,20	0
13,14	17,18	1	14,15	23,23	1			
13,14	20,21	0						

Table 15: AIDS Data Heavily Treated Group: The indicator 0 will denote those individuls who are younger than 20, and 1 will denote those who are older than 20. Multiplicities are denoted in brackets. The first interval censored time refers to the time of infection with HIV and the second the time of developing AIDS. All times are recorded in units of 6 months.

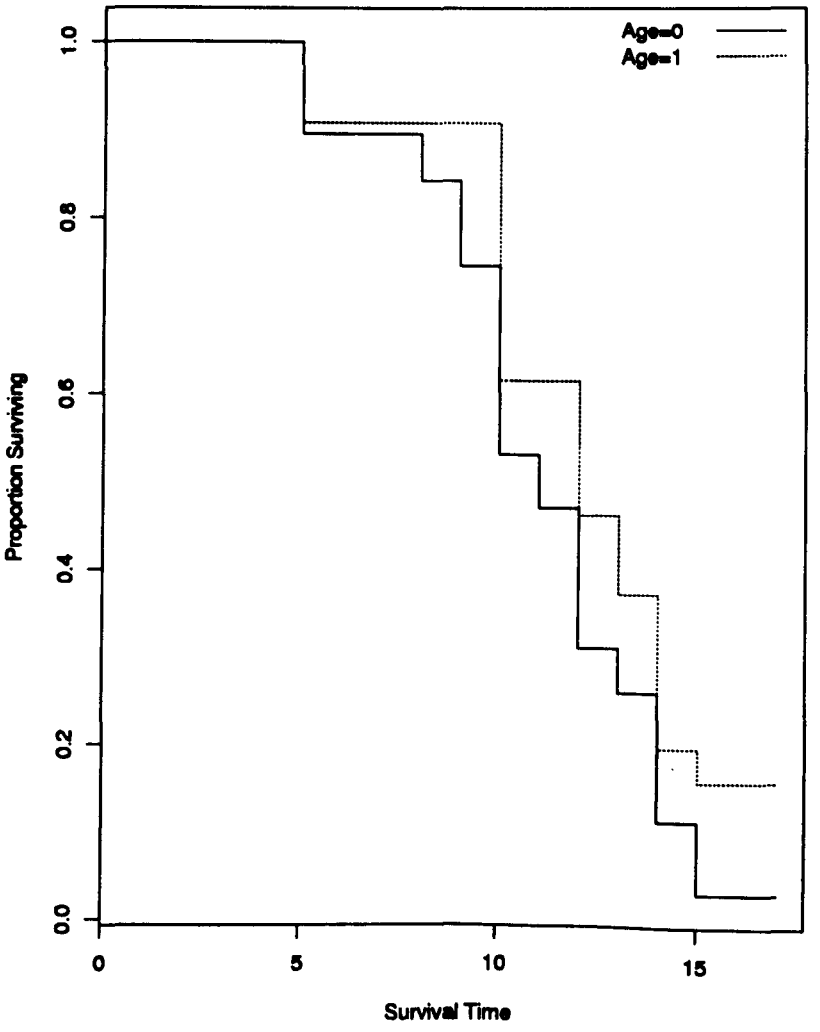

Figure 39: AIDS Data Lightly Treated Group: Survival estimated for time of infection with HIV, using Turnbull's non-parametric estimate. Age 0 refers to the young , and 1 to the old: the younger age group has a slightly worse prognosis. Time 0 refers to the baseline year of 1976.

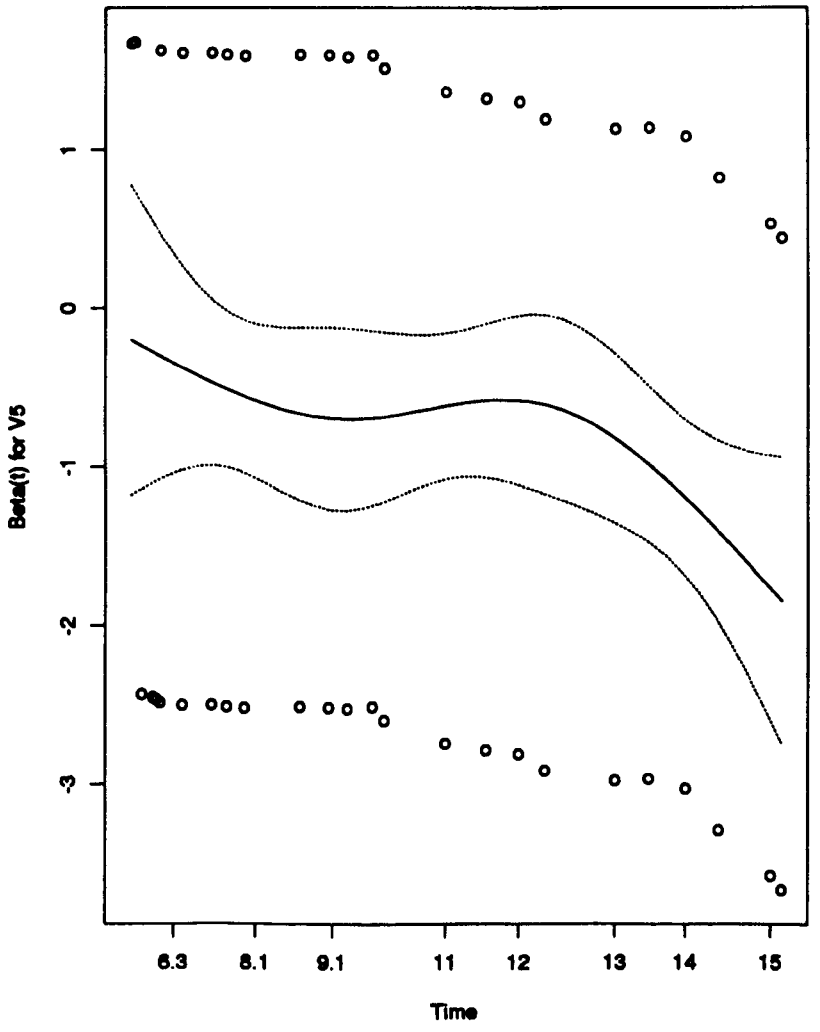

Figure 40: AIDS Data Lightly Treated Group: Test of proportionality for the treatment covariate, using the Grambsch and Therneau (1994) method (Chi-squared 5.84 on 1 degree of freedom), on the midpoints. The effect of the treatment group is always below zero: those heavliy treated have a poorer prognosis.

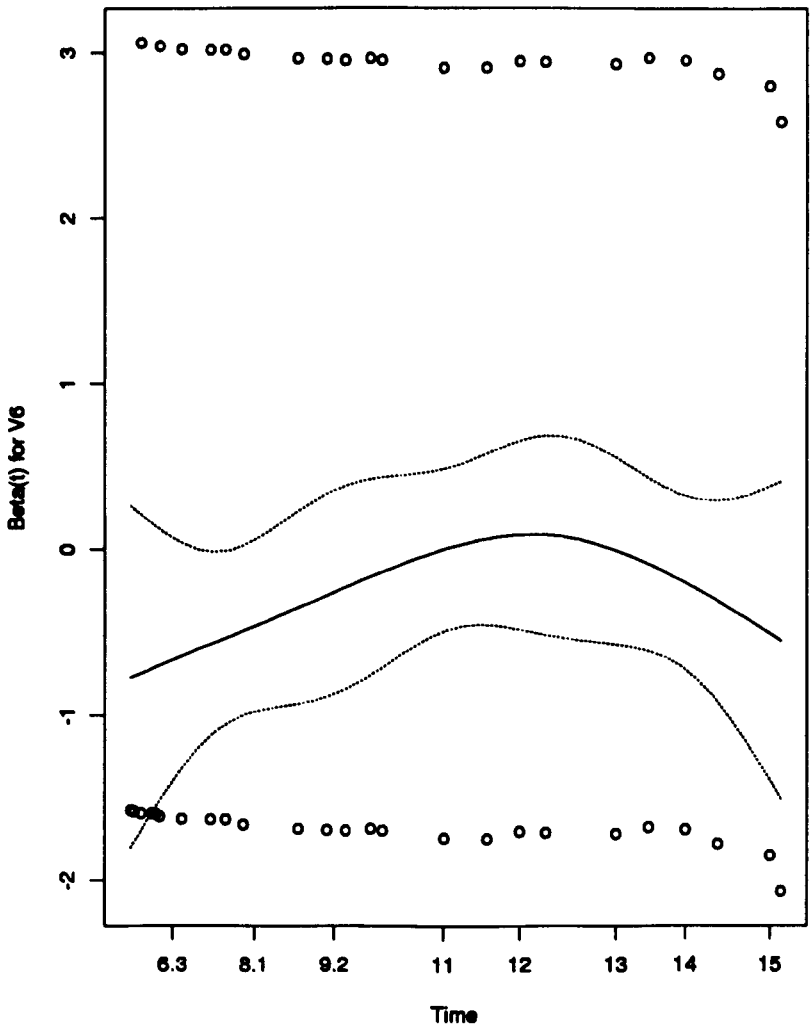

Figure 41: AIDS Data Lightly Treated Group: Test of proportionality for the age covariate, using the Grambsch and Therneau (1994) method (chi-squared 0.57 on 1 degree of freedom), on the midpoints. The age coefficient is always below zero: the younger age group have a poorer prognosis.

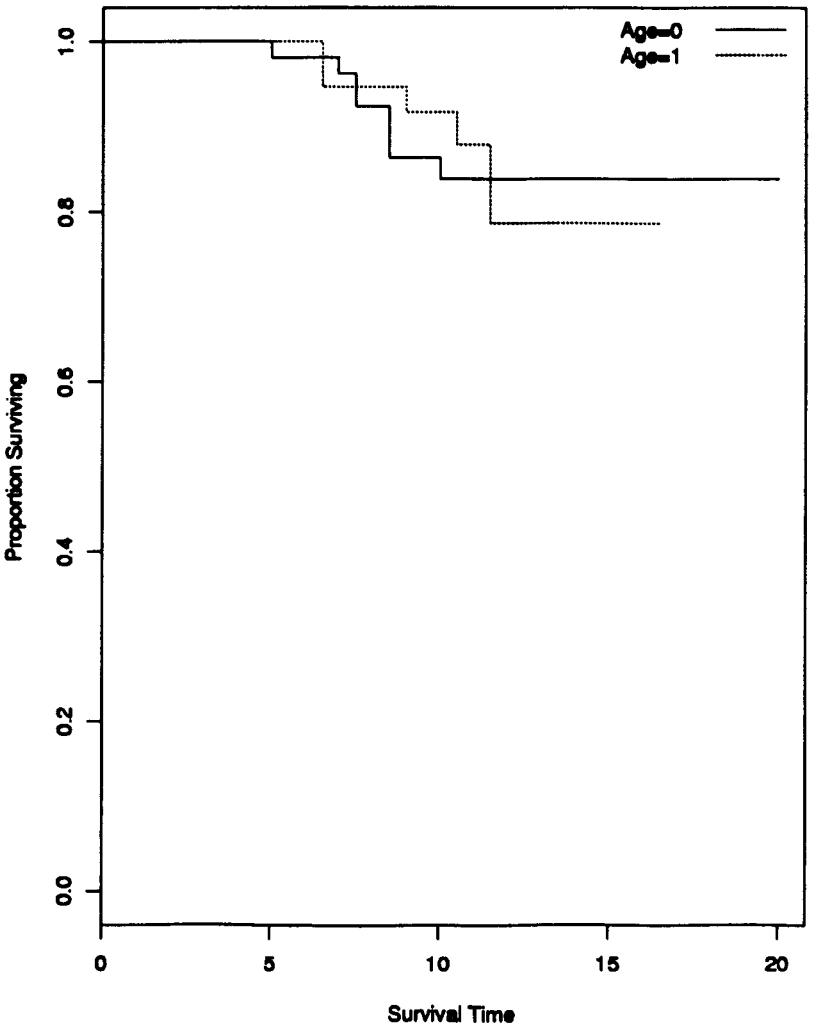

Figure 42: AIDS Data Lightly Treated Group: Survival estimated using the non-parametric technique of Turnbull (1976), using the mid-point estimate of the time of infection with HIV. Age 0 refers to the young , and 1 to the old.

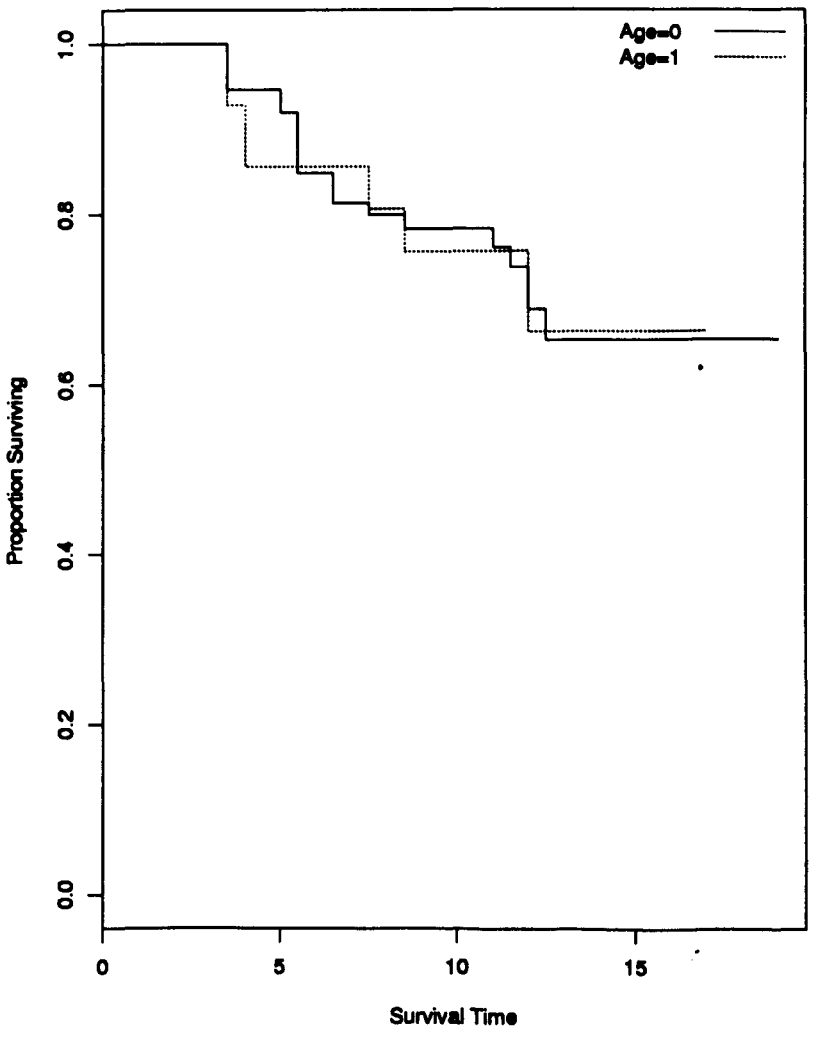

Figure 43: AIDS Data Heavily Treated Group: Survival estimated using the non-parametric technique of Turnbull (1976), using the mid-point estimate of the time of infection with HIV. Age 0 refers to the young , and 1 to the old.

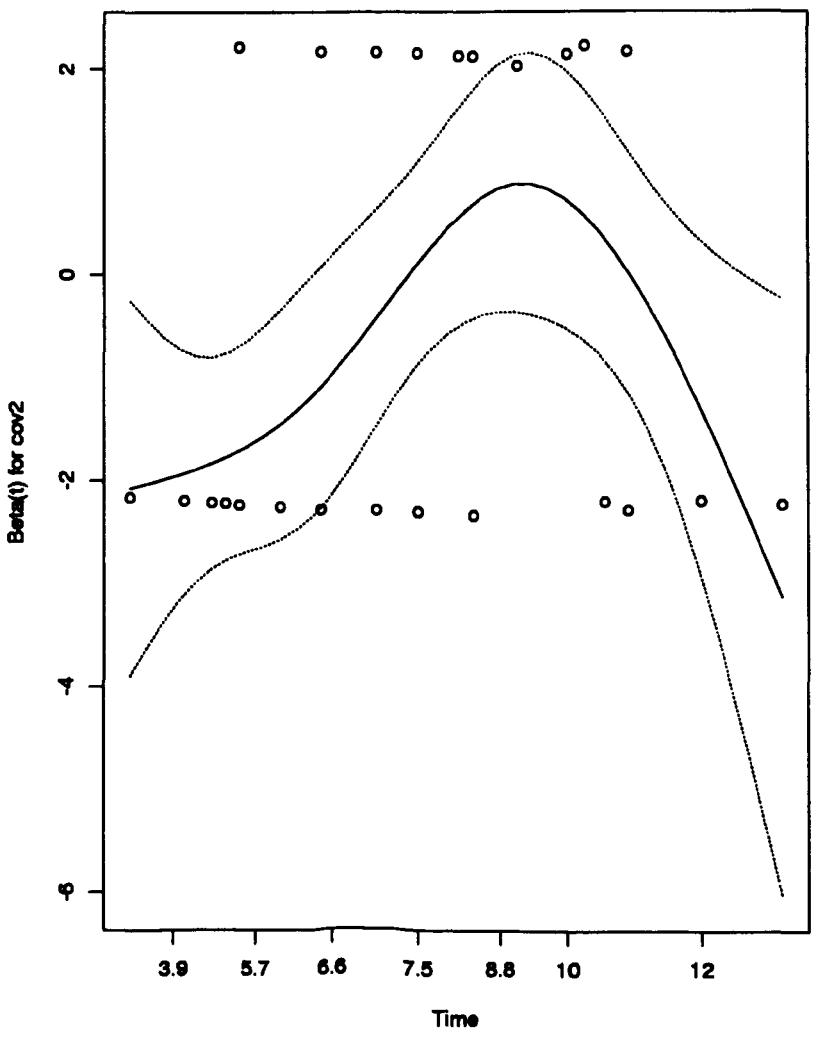

Figure 44: AIDS Data Heavily Treated Group: Test of proportionality for the treatment covariate, using the Grambsch and Therneau (1994) method yields a chi-squared value of 1.82 on 1 degree of freedom (the analysis is based on the midpoints for both the initiating and terminating events). The effect of treatment for the most part lies below zero: those who are heavily treated have a poorer prognosis.

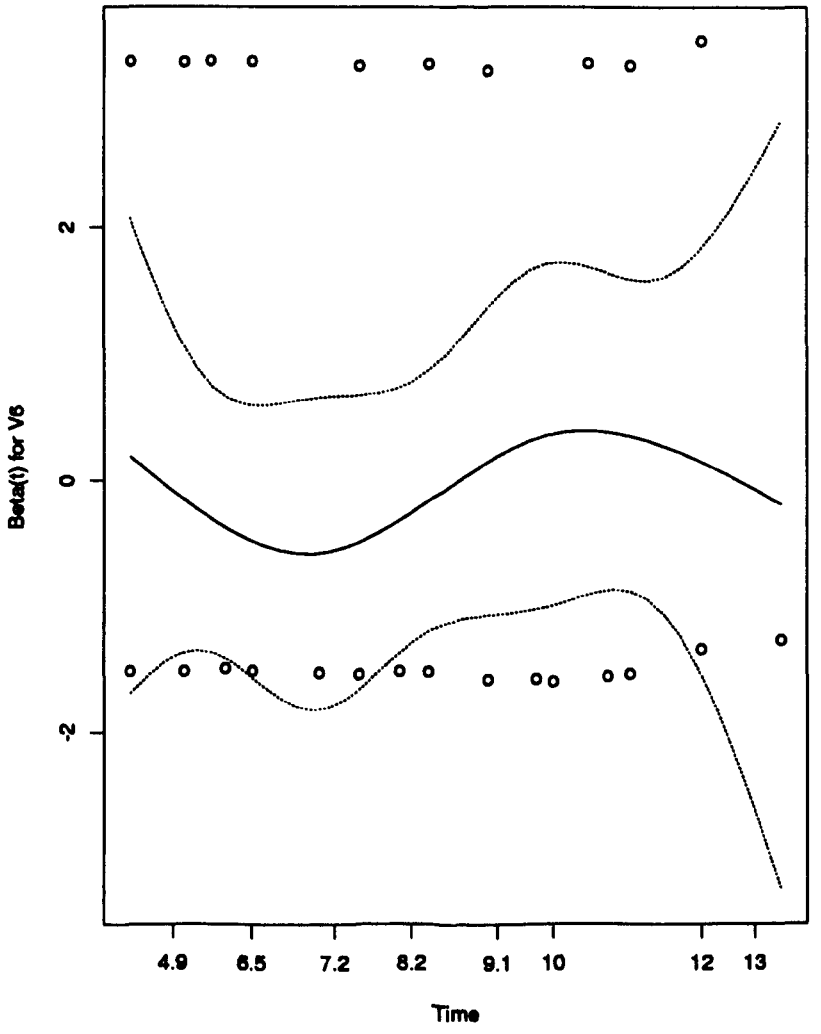

Figure 45: AIDS Data Heavily Treated Group: Test of proportionality for the age covariate, using the Grambsch and Therneau (1994) method yields a chi-squared value of 0.031 on 1 degree of freedom (based on midpoints for both the initiating and terminating events). The effect of age for the most part remains below zero: the younger age group have a poorer prognosis.

Time	β_0	var(β_0)	β_{Treat}	var(β_{Treat})	β_{Age}	var(β_{Age})
2	-5.790575	2.00686	-0.960747	0.39185	-1.64264	1.751489
4	-4.602558	2.46481	-0.817161	0.18623	-1.327341	0.832039
6	-4.129051	0.999091	-0.872241	0.193277	-1.032307	0.561733
8	-3.310349	0.799196	-0.916407	0.176341	-0.820576	0.59865
10	-2.842222	0.389331	-0.901194	0.115074	-0.726836	0.413043
12	-1.427167	0.202624	-0.797409	0.117987	0.028259	0.623024

Table 16: AIDS Data Initiating Event: Estimated parameters for the NDSM, based on 10,000 iterations and using the exact approach for interval censored data.

Time	β_0	var(β_0)	β_{Treat}	var(β_{Treat})	β_{Age}	var(β_{Age})
2	-5.305671	1.087589	-2.234006	3.261825	-0.585549	1.091998
4	-4.128886	0.907251	-1.832218	1.336736	-0.406836	0.625257
6	-3.013641	0.42909	-1.253854	1.510693	-0.223149	0.420917
8	-3.058191	0.228827	-1.080417	0.425423	-0.193311	0.215805
10	-3.146728	0.246401	-0.499555	0.534915	-0.139119	0.226517
12	-3.206084	0.326398	-0.716354	0.514963	-0.156922	0.293463
16	-4.035723	0.869695	-1.469853	1.873698	-0.385959	0.655941
18	-4.7271	1.892896	-1.999056	2.737599	-0.559676	1.03059
20	-5.194874	1.75505	-2.404214	2.723598	-0.732929	0.772102

Table 17: AIDS Data Terminating Event: Estimated parameters for the NDSM, based on 10,000 iterations and using the exact approach for double interval censored data.

Time	β_0	var(β_0)	β_{Treat}	var(β_{Treat})	β_{Age}	var(β_{Age})
2.000000	-5.217024	0.247651	-1.942915	1.521044	-1.243109	1.822406
4.000000	-4.747465	0.288911	-2.328925	0.889899	-0.891388	0.939200
6.000000	-3.018825	0.227894	-2.547969	1.066354	-0.176423	1.992035
8.000000	-2.911131	0.220405	-1.003226	3.947143	-0.508922	0.717651
10.000000	-3.329770	0.120149	-0.596732	0.421432	-0.078895	0.993386
12.000000	-3.140666	0.343819	-0.923506	0.814023	0.164282	0.869636
16.000000	-4.391720	1.050059	-1.871770	4.122241	-0.473107	1.911123
18.000000	-4.940779	1.150206	-2.423250	2.148905	-0.774359	1.308590
20.000000	-5.449628	1.364477	-2.833107	1.471056	-1.167297	0.905398

Table 18: AIDS Data Terminating Event: The numerical output for the estimated covariates and log-baseline hazard using NDSM and the imputation appraoch.

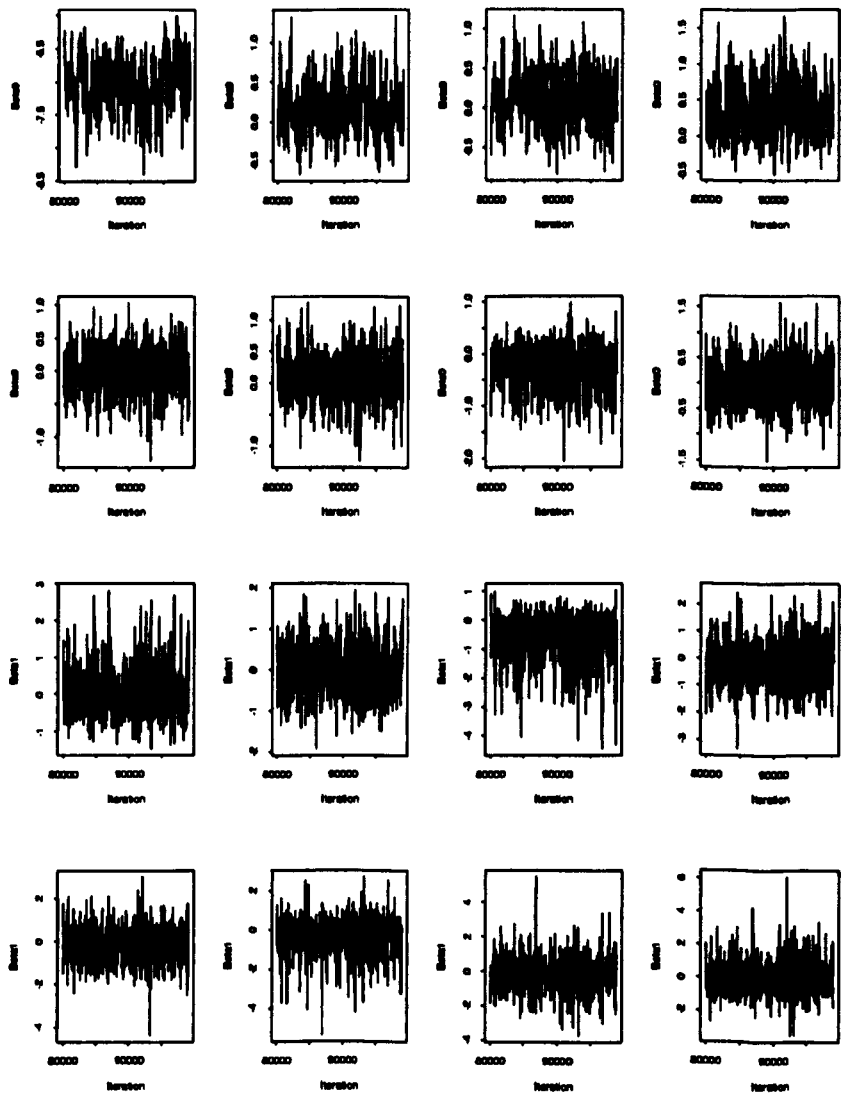

Figure 46: AIDS Data: A selection of the simulated values from the Gibbs sampler for the incubation period. The top half of the plot refers to the baseline parameter, and the latter half to the treatment type and age covariate.

A.5 Simulated Data

Model	Time	β_0	var(β_0)	β_1	var(β_1)
Right Censored (1)	10.00000	-5.685481	0.079080	-0.063452	0.149427
	50.00000	-5.685481	0.079080	-0.063452	0.149427
	100.000000	-5.685481	0.079080	-0.063452	0.149427
	500.000000	-5.894586	0.165527	-1.050403	0.332961
	1000.000000	-6.603894	0.598130	-2.477262	0.721092
	1500.000000	-8.828787	1.042869	-3.931657	0.628030
	1800.000000	-9.527205	0.947608	-4.696644	0.726968
Right Censored (2)	10.000000	-4.969922	0.127922	-0.614938	0.235428
	50.000000	-5.958526	0.120786	-0.437071	0.301176
	100.000000	-6.253285	0.088837	-0.423458	0.324750
	500.000000	-5.647845	0.139456	-1.050906	0.314525
	900.000000	-4.959828	0.158246	-2.735861	0.378840
Interval Censored (1)	100.000000	-5.169090	0.029972	-0.862561	0.107349
	500.000000	-5.880076	0.094768	-0.989358	0.109221
	1000.000000	-5.041160	0.159216	-1.447741	0.126527
	1500.000000	-4.146556	0.203692	-2.090662	0.113536
	1800.000000	-3.267208	0.238393	-1.905883	0.140437
Interval Censored (2)	100.000000	-5.353534	0.063426	-0.749494	0.121556
	500.000000	-5.836803	0.152770	-1.078415	0.168785
	1000.000000	-4.928615	0.144398	-1.297322	0.598187
	1500.000000	-4.213766	0.267698	-2.273114	0.245310
	1800.000000	-3.144088	0.351331	-2.054257	0.355576
Double Interval Censored	100.000000	-5.177161	0.489248	-0.959257	0.384927
	500.000000	-5.992734	0.193816	-1.344065	0.486133
	1000.000000	-4.633619	0.397782	-1.917615	0.273418
	1550.000000	-1.592266	0.359383	-4.449400	0.408326

Table 19: Simulated Data: Summary Output

B Simulation Code

Initialisation

Initial data and priors are stored in text files read in via the following functions.

The following function contains prior parameter estimates.

```
fn_prior(N,n,k,datatype,beta_prmn,beta_prvr,lambda_prmn,
lambda_prvr,alphaW,betaW,alphaV,betaV,alphaF)
int     N,n,k,datatype;
double  beta_prmn[k+2],beta_prvr[k+2];
double  lambda_prmn[k+2],lambda_prvr[k+2];
double  alphaW[k+2],betaW[k+2],alphaV[k+2],betaV[k+2],alphaF[2];
{
int     g,i,l;
for(g=1;g<=k+1;++g){
        beta_prmn[g]=-3.0;
        beta_prvr[g]=15.0;
        alphaW[g]=0.0001;
        betaW[g]=0.0001;
        }
for(g=1;g<=k+1;++g){
        lambda_prmn[g]=-3.0;
        lambda_prvr[g]=15.0;
        alphaV[g]=0.0001;
        betaV[g]=0.0001;
        }
alphaF[1]=7.0;
alphaF[2]=3.0;
}
```

The following function is an initialisation function, which will enable sampled values to be stored and acceptance rates calculated.

```
fn_start_up(N,n,k,datatype,L,NT,store,accpt,storeW,accptW,storeT,
accptT,storeV,accptV,accptF,storeF,my_rand,st_Fparm,
var_store,var_storeW,var_storeV,var_storeT,var_storeF)
int     N,n,k,datatype,L,NT;
int     accptF[L+1],accpt[k+2][N+1],accptW[k+2],accptT[k+2][NT+1],
accptV[k+2],my_rand[3];
double  store[k+2][N+1],storeW[k+2],storeF[L+1],storeT[k+2][NT+1],
storeV[k+2],st_Fparm[L+1];
double  var_store[k+2][N+1],var_storeW[k+2],var_storeF[L+1],
var_storeT[k+2][NT+1],var_storeV[k+2];
{
int     l,g,i;
int     rand();
for(g=1;g<=k+1;++g){
        for(i=1;i<=N;++i){
        store[g][i]=0.0;
```

```
        var_store[g][i]=0.0;
        accpt[g][i]=0;
        }
        storeW[g]=0.0;
        var_storeW[g]=0.0;
        accptW[g]=0;
        }
for(g=1;g<=k+1;++g){
        for(i=1;i<=NT;++i){
        storeT[g][i]=0.0;
        var_storeT[g][i]=0.0;
        accptT[g][i]=0;
        }
        storeV[g]=0.0;
        var_storeV[g]=0.0;
        accptV[g]=0;
        }
for(l=1;l<=L;++l){
        accptF[l]=0;
        storeF[l]=0.0;
        var_storeF[l]=0.0;
        st_Fparm[l]=0.0;
        }
my_rand[1]=rand();
my_rand[2]=rand();
my_rand[3]=rand();
}
```

The following function reads in the relevent data frames.

```
fn_longdata(N,n,k,datatype,data,covariate,interval,censoring,NT,
intervalT,L,ind_frail,frail)
int     N,n,k,datatype,NT,L,ind_frail;
double  data[datatype+1][n+1],covariate[k+2][n+1],interval[N+1],
censoring[n+1],intervalT[NT+1];
int     frail[n+1];
{
int     i,j,g,gk,dt;
int     covkg;
double  covarj,obsj,inter;
FILE    *fp_int,*fp,*fopen();
fp_int=fopen("interval.txt","r");
for(i=0;i<=N;++i){
fscanf(fp_int,"%lf",&inter);
interval[i]=inter;
}
if(datatype==4){
        for(j=0;j<=NT;++j){
        fscanf(fp_int,"%lf",&covarj);
        intervalT[j]=covarj;
        }}
```

```
close(fp_int);
fp=fopen("data.txt","r");
for(j=1;j<=n;++j){
        fscanf(fp,"%lf",&obsj);
        data[1][j]=obsj;
        if(datatype==2){
                fscanf(fp,"%lf",&obsj);
                data[2][j]=obsj;
        }
        if(datatype==4){
                fscanf(fp,"%lf",&obsj);
                data[2][j]=obsj;
                fscanf(fp,"%lf",&obsj);
                data[3][j]=obsj;
                fscanf(fp,"%lf",&obsj);
                data[4][j]=obsj;
        }
covariate[1][j]=1.0;
for(g=2;g<=k+1;++g){
        fscanf(fp,"%lf",&covarj);
        covariate[g][j]=covarj;
        }
        fscanf(fp,"%lf",&covarj);
        censoring[j]=covarj;
        if(ind_frail==1){
                fscanf(fp,"%d",&covkg);
                frail[j]=covkg;
                }
        if(ind_frail==0){
                frail[j]=1;
                }
}
close(fp);
}
```

The following function reads in starting values for all parameters within the Gibbs sampler.

```
fn_start_values(N,n,k,datatype,beta,Wvar,L,ind_frail,frail_parm)
int     N,n,k,datatype,L,ind_frail;
double  beta[k+2][N+1],Wvar[k+2];
double  frail_parm[L+1];
{
double  beta0,beta1;
int     l,tmp,g,j,indicator;
FILE    *fpinit,*fp_pr_out,*fopen();
fpinit=fopen("initial.txt","r");
for(g=2;g<=k+1;++g){
        fscanf(fpinit,"%lf",&beta0);
```

```
        beta[g][1]=beta0;
        for(j=2;j<=N;++j){
                beta[g][j]=0.0;
                }
        }
        fscanf(fpinit,"%lf",&beta0);
        Wvar[1]=beta0;
for(g=2;g<=k+1;++g){
        fscanf(fpinit,"%lf",&beta1);
        Wvar[g]=beta1;
        }
for(l=1;l<=L;++l){
        frail_parm[l]=1.0;
        }
close(fpinit);
}
```

Output Statistics

The estimated survival and hazard functions are computed via the following two functions.

```
fn_hazard(trans_beta,input_cov,n,N,k,interval,fprun,frail_ind,
frailty,alphaF)
int     n,k,N,frail_ind;
double  trans_beta[k+2][N+1],input_cov[k+2],alphaF[2];
double  interval[N+1],frailty[n+1];
FILE    *fprun;
{
double  exp(),sum,sumtmp,tmplambda,tmp1,tmp2[k+2],Afunction();
int     j,i,g,gb;
double  diff,sum2,total[k+2];
input_cov[1]=1.0;
tmp2[1]=1.0;
total[1]=0.0;
for(g=2;g<=k+1;++g){
        input_cov[g]=0.0;
        tmp2[g]=1.0;
        total[g]=0.0;
}
for(i=1;i<=N;++i){
for(g=1;g<=k+1;++g){
sum=trans_beta[1][i];
for(gb=2;gb<=k+1;++gb){
        input_cov[gb]=0.0;}
        input_cov[g]=1.0;
        for(gb=2;gb<=k+1;++gb){
        sumtmp=trans_beta[gb][i]*input_cov[gb];
        sum=sum+sumtmp;}
```

```
        tmplambda=sum;
        if(frail_ind==1){
                sum2=trans_beta[1][i];
                for(gb=2;gb<=k+1;++gb){
                sumtmp=trans_beta[gb][i]*input_cov[gb];
                sum2=sum+sumtmp;}
                diff=interval[i]-interval[i-1];
                sumtmp=exp(sum2)*diff;
                total[g]=total[g]+sumtmp;
tmp1=exp(tmplambda)*alphaF[1]/(alphaF[1]+total[g]);}
         if(frail_ind!=1){
         tmplambda=sum;
         tmp1=exp(tmplambda);}
         fprintf(fprun,"%f ",tmp1); }
        }
}

fn_survival(trans_beta,input_cov,n,N,k,interval,fprun2,frail_ind,
frailty,alphaF)
int     n,k,N,frail_ind;
double  trans_beta[k+2][N+1],input_cov[k+2];
double  interval[N+1],frailty[n+1],alphaF[2];
FILE    *fprun2;
{
double  exp(),log(),sum,sumtmp,tmplambda,tmp1,tmp2[k+2],diff,tmp3,
tmp2B[k+2],totalB[k+2],tmp4,alp,total[k+2],sum2,frail;
int     tmp,i,g,gb,j;
input_cov[1]=1.0;
tmp2[1]=1.0;
total[1]=0.0;
for(g=1;g<=2;++g){
input_cov[g]=0.0;
tmp2[g]=1.0;
total[g]=0.0;
tmp2B[g]=1.0;
totalB[g]=0.0;
}
for(i=1;i<=N;++i){
        fprintf(fprun2,"\n %f ",interval[i]);
for(g=1;g<=2;++g){
frail=frailty[g];
sum=trans_beta[1][i];
tmplambda=sum;
        tmp1=(-1.0)*exp(tmplambda);
        tmp1=tmp1*frail;
        diff=interval[i]-interval[i-1];
        tmp2[g]=tmp2[g]*exp(tmp1*diff);
        fprintf(fprun2,"%f ",tmp2[g]);
}
}
```

```
}
```

The Gibbs samplers

The following function includes all of the code to carry out a single iteration of the Gibbs sampler.

```
fn_gibbs(M,D,buff,buffT,covariate,Wvar,interval,data,
datatype,beta,N,n,k,store,accpt,storeW,accptW,
my_rand,censoring,MT,DT,Vvar,intervalT,
        beta_prmn,beta_prvr,alphaW,betaW,beta_propvr,
        storeT,accptT,storeV,accptV,lambda_prmn,
        L,ind_frail,frail,frail_parm,F,storeF,accptF,
lambda_prvr,alphaV,betaV,hyper,icaug,dicaugic,
        var_store,var_storeW,var_storeV,var_storeT,var_storeF)
int     M,D,NT,MT,DT,buff,buffT,accptW[k+2],accpt[k+2][N+1],
my_rand[3],frail[n+1],beta_propvr[k+2][N+1];
double  Wvar[k+2],beta[k+2][N+1],store[k+2][N+1],storeW[k+2],
st_Fparm[L+1],censoring[n+1],storeF[L+1];
double  interval[N+1],data[datatype+1][n+1],covariate[k+2][n+1];
double  beta_prmn[k+2],beta_prvr[k+2],alphaW[k+2],betaW[k+2];
int     accptV[k+2],accptT[k+2][NT+1],accptF[L+1];
double  Vvar[k+2],lambda[k+2][NT+1],storeT[k+2][NT+1],storeV[k+2];
double  lambda_prmn[k+2],lambda_prvr[k+2],alphaV[k+2],betaV[k+2];
double  lam_propvr[k+2][NT+1],intervalT[NT+1],frail_parm[L+1],F[2];
double  ,var_storeT[k+2][NT+1],var_storeV[k+2]
double  var_store[k+2][N+1],var_storeW[k+2],var_storeF[L+1];
{
int     m,g,i,j,indicator,typetmp,l;
double  fn_modulas(),tmp1,tmpa,tmpb,log_lik,fn_lik();
double  trans_lambda[k+2][NT+1],mn_lambda[k+2][NT+1],var_lambda[k+2][NT+1]
trans_beta[k+2][N+1],mn_beta[k+2][N+1],var_beta[k+2][N+1];
double  xtmpT,mtmp,tmp,buff2,bufftmp;
double  olddata[datatype+1][n+1],xtmp,fn_value(),fn_valueT();
double  dataS[datatype+1][n+1],log_likA,log_likB;
FILE    *fp_finT,*fp_simT1,*fp_simT2,*fprunT,*fp_final,
*fp_sim1,*fprun,*fp_simT3,*fp_simT4,*fopen();
fp_final=fopen("final_out.txt","w");
fprun=fopen("running_sum.txt","w");
fp_sim1=fopen("sim1.txt","w");

for(j=1;j<=datatype;++j){
                for(i=1;i<=n;++i){
                        olddata[j][i]=0.0;
                }
        }
if(icaug==1){
        for(j=1;j<=datatype;++j){
                for(i=1;i<=n;++i){
                        olddata[j][i]=data[j][i];
                        data[j][i]=0.0;
```

```
                }
        }
}
typetmp=0;
if(datatype==4){
        typetmp=4;
        datatype=2;
}
/*** Augments the data (where necessary) ***/
if(icaug==1){datatype=1;}
for(m=1;m<=M;++m){
printf("This is it no %d %d  (init)\n",m,dicaugrc);
if(icaug==1){
fn_transform(beta,trans_beta,N,k,interval);
for(i=1;i<=n;++i){
        if(olddata[2][i]==0.0){
                data[1][i]=olddata[1][i];
                censoring[i]=0.0;
        }
        if(olddata[1][i]<olddata[2][i]){
                xtmp=fn_value(N,n,k,datatype,i,covariate,interval
olddata,trans_beta,censoring);
                data[1][i]=xtmp;
                censoring[i]=1.0;
                 if(olddata[1][i] > data[1][i] || data[1][i] > olddata[2][i]){
printf("WARNING: DATAPOINT GENERATED OUT OF RANGE %d \n",i);}
        }
        if(olddata[1][i]==olddata[2][i]){
                data[1][i]=olddata[1][i];
                censoring[i]=1.0;
        }


}
}
/*** Carries out a Gibbs sample on the frailty paramaters***/
if(ind_frail==1){
        for(l=1;l<=L;++l){
        frail_gibbs(buff,D,m,l,covariate,interval,data,datatype,
beta,N,n,k,my_rand,censoring,
        NT,intervalT,lambda,L,ind_frail,frail,frail_parm,F,storeF,
accptF,st_Fparm,var_storeF);
        }
        fn_fullfrail(n,k,N,buff,D,L,m,F,frail_parm,accptF,storeF,
my_rand,var_storeF);
}

if(cox==0 & coxinit==0){
indicator=1;
for(g=1;g<=k+1;++g){
        for(i=1;i<=N;++i){
        covar_gibbs(buff,D,1,m,g,i,9,covariate,Wvar,interval,data,
```

```
        datatype,beta,N,n,k,store,accpt,my_rand,censoring,
        beta_prmn,beta_prvr,beta_propvr,NT,intervalT,lambda,fprun,L,
ind_frail,frail,frail_parm,var_store);
        }}
if(hyper==1){
indicator=2;
for(g=1;g<=k+1;++g){
        evolution_gibbs(buff,D,2,m,g,9,covariate,Wvar,interval,data,
        datatype,beta,N,n,k,storeW,accptW,my_rand,censoring,alphaW,
betaW,NT,intervalT,lambda,
        L,ind_frail,frail,frail_parm,var_storeW);
        }
}
}
if(cox==1 || coxinit==1){
indicator=1;
        for(i=1;i<=N;++i){
        covar_gibbs(buff,D,1,m,1,i,9,covariate,Wvar,interval,data,
        datatype,beta,N,n,k,store,accpt,my_rand,censoring,
        beta_prmn,beta_prvr,beta_propvr,NT,intervalT,lambda,fprun,L
ind_frail,frail,frail_parm,var_store);
        }
        for(g=2;g<=k+1;++g){
        covar_gibbs(buff,D,1,m,g,1,9,covariate,Wvar,interval,data,
        datatype,beta,N,n,k,store,accpt,my_rand,censoring,
        beta_prmn,beta_prvr,beta_propvr,NT,intervalT,lambda,fprun,
L,ind_frail,frail,frail_parm,var_store);
        }
        if(hyper==1){
        indicator=2;
        evolution_gibbs(buff,D,2,m,1,9,covariate,Wvar,interval,data,
        datatype,beta,N,n,k,storeW,accptW,my_rand,censoring,alphaW,
betaW,NT,intervalT,lambda,
        L,ind_frail,frail,frail_parm,var_storeW);
        }
}
/*** Stores values which have been sampled ***/
if(m>D){
        mtmp=m+1.0;
        mtmp=mtmp-1.0;
        bufftmp=buff+1;
        buff2=bufftmp-1.0;
        tmp=fn_modulas(mtmp,buff2);
        if(tmp==0.0){
                fprintf(fp_sim1,"\n %d  ",(m-D)/buff);
                for(g=1;g<=k+1;++g){
                        fprintf(fp_sim1,"%lf",Wvar[g]);
                for(i=1;i<=N;++i){
                        fprintf(fp_sim1,"%lf",beta[g][i]);}
                }
                if(ind_frail==1){
                        for(l=1;l<=L;++l){
```

```
                                fprintf(fp_sim1,"%lf",frail_parm[l]);
                        }
                }
        }
}
}

close(fp_sim1);
close(fprun);
/*** Calculates the mean and final likelihood ***/
fn_mean(N,M,D,k,buff,store,mn_beta,var_beta,var_store);
log_likA=fn_lik(N,n,k,datatype,mn_beta,interval,
l,covariate,data,censoring,N,intervalT,
lambda,L,ind_frail,frail,frail_parm);
log_likB=0.0;
if(icaug==1){
log_likB=fn_lik(N,n,k,2,mn_beta,interval,l,covariate,
olddata,censoring,N,intervalT,lambda,L,
ind_frail,frail,frail_parm);
}
fn_summary(n,M,D,k,N,buff,Wvar,accptW,accpt,storeW,
mn_beta,trans_beta,interval,fp_final,L,
ind_frail,accptF,frail_parm,frail,F,storeF,
st_Fparm,var_beta,var_storeW,var_storeF);
fprintf(fp_final,"\n\loglikelihood %lf%lf\n",log_likA,log_likB);
close(fp_final);
}
}
}
```

The Metroplis-Hastings Samplers

This is the Metropolis-Hastings sampler for the covariate effect parameters.

```
covar_gibbs(buff,D,indicator,it_no,g,i,l,covariate,
Wvar,interval,data,datatype,beta,N,n,k,store,
accpt,my_rand,censoring,beta_prmn,beta_prvr,
propvar,NT,intervalT,lambda,fp_covar,L,ind_frail,
        frail,frail_parm,var_store)
int     buff,datatype,N,n,k,NT,L,l,ind_frail, frail[n+1];
int     D,indicator,it_no,g,i,accpt[k+2][N+1],my_rand[3];
double  Wvar[k+2],beta[k+2][N+1],store[k+2][N+1];
double  interval[N+1],data[datatype+1][n+1],covariate[k+2][n+1];
double  beta_prmn[k+2],beta_prvr[k+2];
double  propvar[k+2][N+1],censoring[n+1],frail_parm[L+1];
double  intervalT[NT+1],lambda[k+2][NT+1],var_store[k+2][N+1];
FILE    *fp_covar;
{
double  current,new1,new2,r1,r2,weight,tmp6,uni;
double  exp(),full_cond(),rnormal(),runiform(),fn_modulas();
double  tmp,mtmp,bufftmp,buff2;
```

```
current=beta[g][i];
r1=full_cond(indicator,g,i,1,covariate,Wvar,interval,
data,datatype,beta,N,n,k,beta_prmn,beta_prvr,
censoring,NT,intervalT,lambda,L,ind_frail,
frail,frail_parm);
new1=rnormal(current,propvar[g][i],my_rand);
beta[g][i]=new1;
r2=full_cond(indicator,g,i,1,covariate,Wvar,interval,
data,datatype,beta,N,n,k,beta_prmn,beta_prvr,
censoring,NT,intervalT,lambda,L,ind_frail,frail,frail_parm);
weight=r2-r1;
tmp6=exp(weight);
uni=runiform(my_rand);
if(tmp6<=uni){
        beta[g][i]=current;
        if(it_no>D){
        mtmp=it_no+1.0;
        mtmp=mtmp-1.0;
        bufftmp=buff+1.0;
        buff2=bufftmp-1.0;
        tmp=fn_modulas(mtmp,buff2);
        if(tmp==0.0){
        store[g][i]=store[g][i]+current;
        var_store[g][i]=var_store[g][i]+(current*current);
        }}
        }
}
```

The following function samples the evolution variance from the relevent full conditional.

```
evolution_gibbs(buff,D,indicator,it_no,g,l,covariate,Wvar,
interval,data,datatype,beta,N,n,k,storeW,accptW,
my_rand,censoring,alphaW,betaW,NT,intervalT,lambda,
L,ind_frail,frail,frail_parm,var_storeW)
int     buff,datatype,N,n,k,NT,L,ind_frail,l,frail[n+1];
int     D,indicator,it_no,g,accptW[k+2],my_rand[3];
double  Wvar[k+2],beta[k+2][N+1],storeW[k+2];
double  censoring[n+1],covariate[k+2][n+1],interval[N+1];
double  alphaW[k+2],betaW[k+2],data[datatype+1][n+1];
double  intervalT[NT+1],lambda[k+2][NT+1],var_storeW[k+2];
double  frail_parm[L+1];
{
int     i;
double  exp(),pgamma(),rgamma(),fn_modulas();
double  tmp,mtmp,buff2,bufftmp;
double  ith_part,newa,newb,sample,sampleW,my_cond;
FILE    *fp_evol,*fopen();
my_cond=0.0;
for(i=2;i<=N;++i){
        ith_part=beta[g][i]*beta[g][i];
        ith_part=ith_part /2*(interval[i]-interval[i-1]);
```

```
        my_cond=my_cond+ith_part;
        }
newa=alphaW[g]+((N-1.0)/2.0);
newb=betaW[g]+my_cond;
sample=rgamma(newa,1.0/newb,my_rand);
sampleW=1.0/sample;
Wvar[g]=sampleW;
}
```

The following function samples the frailty parameters from the relevent gamma distribution.

```
frail_gibbs(buff,D,it_no,l,covariate,interval,data,
datatype,beta,N,n,k,my_rand,censoring,
        NT,intervalT,lambda,L,ind_frail,frail,frail_parm,
alphaF,storeF,accptF,st_Fparm,var_storeF)
int     buff,datatype,N,n,k,NT,L,ind_frail,l,frail[n+1];
int     D,it_no,accptF[L+1],my_rand[3];
double  beta[k+2][N+1],censoring[n+1],var_storeF[L+1];
double  interval[N+1],data[datatype+1][n+1],covariate[k+2][n+1];
double  intervalT[NT+1],lambda[k+2][NT+1];
double  frail_parm[L+1],alphaF[2],storeF[L+1],st_Fparm[L+1];
{
int     i;
double  aF,bF,lik,bdash;
double  mtmp,bufftmp,buff2,tmp,tmp1,tmpaf;
double  exp(),fn_modulas(),rgamma(),fn_frail_lik(),fn_hyperfrail();
double  newalpha,newbeta,hyperalpha,hyperbeta,betadash,delta;
aF=alphaF[1];
delta=0.0;
for(i=1;i<=n;++i){
        if(frail[i]==1){
                delta=delta+censoring[i];
}
}
lik=fn_frail_lik(k,n,N,L,datatype,1,data,interval,
covariate,beta,frail,frail_parm);
tmpaf=aF;
tmp1=rgamma(aF+delta,1.0/(tmpaf+lik),my_rand);
frail_parm[1]=tmp1;
}
```

This function calculates the full conditional for the covariate effect parameters.

```
double  full_cond(indicator,g,int_index,l,covariate,Wvar,interval,
data,datatype,beta,N,n,k,beta_prmn,beta_prvr,censoring,NT,
intervalT,lambda,L,ind_frail,frail,frail_parm)
int     datatype,N,n,k,NT,l,ind_frail,L,frail[n+1];
double  beta[k+2][N+1],censoring[n+1],data[datatype+1][n+1];
double  beta_prmn[k+2],beta_prvr[k+2],Wvar[k+2],covariate[k+2][n+1];
int     indicator,g,int_index;
```

```
double  lambda[k+2][NT+1],interval[N+1],intervalT[NT+1],frail_parm[L+1];
{
double  lik2,lik,mult_by,ith_part,my_cond,tmp6,tmp1,tmp2,tmp3;
int     i,index1;
double  log(),log_norm(),fn_lik();
lik=fn_lik(N,n,k,datatype,beta,interval,1,covariate,data,censoring,
NT,intervalT,lambda,L,ind_frail,frail,frail_parm);
if(int_index==1){
        mult_by=log_norm(beta[g][1],beta_prmn[g],beta_prvr[g]);
        }
if(int_index!=1){
        mult_by=log_norm(beta[g][int_index],0.0,Wvar[g]);
        }
return(lik+mult_by);

}
```

The following function evaluates the full conditional for the frailty hyper-parameter.

```
double  fn_hyperfrail(current,L,frail_parm)
int     L;
double  current,frail_parm[L+1];
{
double  tmp1,tmp2,tmp3,tmp4,tmp5,tmp6,tmp7,P,S,
hyperalpha,hyperbeta,invcurrent;
int     l;
double  gammafn(),exp(),log();
/*************************************/
P=0.0;
S=1.0;
hyperalpha=0.001;
hyperbeta=1.001;
/*************************************/
for(l=1;l<=L;++l){
        P=P+frail_parm[l];
        S=S*frail_parm[l];
        }
tmp1=gammafn(current);
tmp2=L*log(tmp1);
tmp3=(L*current)+hyperalpha-1.0;
tmp4=tmp3*log(current);
tmp5=current*hyperbeta;
tmp6=(current-1.0)*log(S);
tmp7=(current)*P;
return(tmp4+tmp6-tmp5-tmp7-tmp2);
}
```

The following functions carry out the Metropolis-Hastings step for the frailty hyper-parameter.

```
fn_fullfrail(n,k,N,buff,D,L,it_no,alphaF,frail_parm,
accptF,storeF,my_rand,var_storeF)
```

```
int     n,k,N,buff,D,L,it_no,accptF[L+1],my_rand[3];
double  alphaF[2],storeF[L+1],frail_parm[L+1],var_storeF[L+1];
{
double  current,new1,new2,r1,r2,weight,tmp6,uni,
p1,p2,alpha,beta;
double  exp(),fn_hyperfrail(),rgamma(),runiform(),
fn_modulas(),pgamma();
double  tmp,mtmp,bufftmp,buff2;
/*****************************************************************/
current=alphaF[1];
r1=fn_hyperfrail(current,L,frail_parm);
p1=pgamma(current,alpha,beta);
new1=rgamma(alpha,1.0/beta,my_rand);
r2=fn_hyperfrail(new1,L,frail_parm);
p2=pgamma(new1,alpha,beta);
weight=r2-r1-p2+p1;
tmp6=exp(weight);
uni=runiform(my_rand);
if(tmp6<=uni){
        alphaF[1]=current;
        if(it_no>D){
        mtmp=it_no+1.0;
        mtmp=mtmp-1.0;
        bufftmp=buff+1.0;
        buff2=bufftmp-1.0;
        tmp=fn_modulas(mtmp,buff2);
        if(tmp==0.0){
        storeF[1]=storeF[1]+current;
storeF[2]=storeF[2]+(current*current);
        }}
        }

}
```

Afunction refers to $A(\cdot)$ used within the calculation of the likelihood.

```
double  Afunction(start,end,obsindex,k,n,N,datatype,trans_beta,
interval,covariate,frailty)
int     k,n,N,datatype;
int     start,end,obsindex;
double  frailty,trans_beta[k+2][N+1],interval[N+1],covariate[k+2][n+1];
{
/*****************************************************************/
double  sum,diff;
double  tmp1,tmp2,tmpc,exp();
int     i,g;
/*****************************************************************/
sum=0.0;
for(i=start;i<end+1;++i){
diff=interval[i]-interval[i-1];
tmp2=trans_beta[1][i];
for(g=2;g<=k+1;++g){
```

```
        tmpc=trans_beta[g][i]*covariate[g][obsindex];
        tmp2=tmp2+tmpc;
        }
tmp1=exp(tmp2)*diff*frailty;
sum=sum+tmp1;
}
return((-1.0)*sum);
}
```

Cfunction refers to the function $C(\cdot)$, also used within the calculation of the likelihood.

```
double  Cfunction(begin,finish,intindex,obsindex,k,n,N,datatype,
trans_beta,interval,covariate,frailty)
int     k,n,N,datatype;
int     intindex,obsindex;
double  begin,finish,frailty;
double  trans_beta[k+2][N+1],interval[N+1],covariate[k+2][n+1];
{
/*****************************************************************/
double sum,diff,tmpc;
double tmp1,tmp2,tmp3;
int j,g;
double  exp(),log();
/*****************************************************************/
diff=finish-begin;
sum=trans_beta[1][intindex];
for(g=2;g<=k+1;++g){
        tmpc=trans_beta[g][intindex]*covariate[g][obsindex];
        sum=sum+tmpc;
        }
tmp2=sum;
tmp2=exp(tmp2);
tmp3= (-1.0)*tmp2*diff*frailty;
return(tmp3);
}
```

The following function calclates each individual contribution to the likelihood.

```
double  log_fn(data,cen,obsindex,k,n,N,datatype,trans_beta,interval,
covariate,L,ind_frail,frail,frail_parm)
int     k,n,N,datatype,L,ind_frail;
int     obsindex;
double  data[datatype+1][n+1],cen,trans_beta[k+2][N+1];
int     frail[n+1];
double  frail_parm[L+1],interval[N+1],covariate[k+2][n+1];
{
double  sum,tmpc,tmp1,tmp2,tmp3,tmp4,tmp5,diff,dataR,dataL,frailty;
int     intR,intL,i,g,frailty_ind;
int     locationR();
double  Afunction(),Cfunction(),exp(),log();
```

```
frailty=1.0;
if(ind_frail==1){
        frailty_ind=frail[obsindex];
        frailty=frail_parm[frailty_ind];
}
dataR=data[1][obsindex];
intR=locationR(N,dataR,interval);
if(datatype==2){
        dataL=data[2][obsindex];
        intL=locationR(N,dataL,interval);
}
tmp1=0.0;
tmp2=0.0;
if(intR!=1){
tmp1 = Afunction(1,intR-1,obsindex,k,n,N,datatype,trans_beta,
interval,covariate,frailty);
}
tmp2=Cfunction(interval[intR-1],dataR,intR,obsindex,k,n,N,
datatype,trans_beta,interval,covariate,frailty);
if(datatype==1 & cen==1.0 || datatype==2 & dataR==dataL){
sum=trans_beta[1][intR];
for(g=2;g<=k+1;++g){
        tmpc=trans_beta[g][intR]*covariate[g][obsindex];
        sum=sum+tmpc;
        }
        tmp5=tmp1+tmp2+sum+log(frailty);
return(tmp5);
}
tmp3=0.0;
tmp4=0.0;
if(datatype==2 & dataL!=0.0){
if(intL!=1){
tmp3= Afunction(1,intL-1,obsindex,k,n,N,datatype,
trans_beta,interval,covariate,frailty);}
tmp3=exp(tmp3);
tmp4=Cfunction(interval[intL-1],dataL,intL,obsindex,
k,n,N,datatype,trans_beta,interval,covariate,frailty);
tmp4=exp(tmp4);
}
return(log((exp(tmp1)*exp(tmp2))-(tmp3*tmp4)));
}
```

The following function calculates the likelihood for the complete data set, based on current parameter estimates.

```
double  fn_lik(N,n,k,datatype,beta,interval,l,covariate,data,
censoring,NT,intervalT,lambda,L,
        ind_frail,frail,frail_parm)
int     k,n,N,datatype,NT,l,L,ind_frail,frail[n+1];
double  beta[k+2][N+1];
```

```
double  interval[N+1],intervalT[NT+1],lambda[k+2][NT+1],frail_parm[L+1],
covariate[k+2][n+1],data[datatype+1][n+1],censoring[n+1];
{
int     i,j,g;
double  cen,sum,tmp1,tmp2,tmp3,tmp4;
double  log_fn(),log_aids_fn();
double  trans_beta[k+2][N+1];
double  trans_lambda[k+2][NT+1];
for(g=1;g<=k+1;++g){
trans_beta[g][1]=beta[g][1];
for(i=2;i<=N;++i){
        trans_beta[g][i]=trans_beta[g][i-1]+beta[g][i];
}}
if(datatype==4){
for(g=1;g<=k+1;++g){
trans_lambda[g][1]=lambda[g][1];
for(i=2;i<=NT;++i){
        trans_lambda[g][i]=trans_lambda[g][i-1]+lambda[g][i];
}}
sum=0.0;
for(j=1;j<=n;++j){
        cen=censoring[j];
        tmp1=0.0;
        if(data[3][j]!=0.0){
        tmp1=log_aids_fn(data,cen,j,k,n,N,NT,datatype,
trans_beta,trans_lambda,
        interval,intervalT,covariate,L,ind_frail,frail,frail_parm);
        }
return(sum);
}
sum=0.0;
for(j=1;j<=n;++j){
        tmp1=0.0;
        /*** if(data[1][j]!=0.0){ ***/
        cen=censoring[j];
        tmp1=log_fn(data,cen,j,k,n,N,datatype,trans_beta,
interval,covariate,
        L,ind_frail,frail,frail_parm);
        sum=sum+tmp1;
}
return(sum);
}
```

References

[1] O. O. Aalen. Heterogeneity in survival analysis. *Statistics in Medicine*, 7:1121–1137, 1988.

[2] M. Aitkin and D. Clayton. The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics*, 29:156–63, 1980.

[3] A. Alioum and D. Commenges. A proportional hazards model for arbitrarily censored and truncated data. *Biometrics*, 52:512–524, 1996.

[4] S. Anderson, A. Auquier, W. W. Hauck, D. Oakes, W Vandaele, and H. I. Weisberg. *Statistical Methods for Comparative Studies.* Wiley, Chichester, 1980.

[5] E. Arjas and D. Gasbarra. Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, 4:505–524, 1994.

[6] H. Aslanidou and D. K. Dey. Model determination for multivariate survival data. Research Report TR9608, Univarsity of Connecticut, Storrs, CT 06269-3120, 1996.

[7] H. Aslanidou, D. K. Dey, and D. Sinha. Bayesian analysis of multivariate survival data using Monte Carlo methods. *The Canadian Journal of Statistics*, 26:33–48, 1998.

[8] P. Bacchetti. Estimating the incubation period of AIDS by comparing population infection and diagnosis patterns. *Journal of the American Statistical Association*, 85:1002–1008, 1990.

[9] N. Best, M. K. Cowles, and K. Vines. Coda: Convergence diagnostics and output analysis software for gibbs sampling output, version 0.4, 1997.

[10] N. E. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974.

[11] R. Brookmeyer and M. H. Gail. *AIDS Epidemiology: A Quantative Approach.* Oxford University Press, 1994.

[12] R. Brookmeyer and J. J. Goedert. Censoring in an epeidmic with an application to hemophilia-associated AIDS. *Biometrics*, 45:325–335, 1989.

[13] B. P. Carlin and T. A. Lewis. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, 1998.

[14] W. H. Carter, G. L. Wampler, and D. M. Stablein. *Regression Analysis of Survival Data in Cancer Chemotherapy.* Marcel Dekker, New York, 1983.

[15] D. G. Clayton. A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 47:467–485, 1991.

[16] D. G. Clayton and J. Cuzick. Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society Series A*, 148:82–117, 1985.

[17] D. Collett. *Modelling Survival Data in Medical Research.* Chapman & Hall, London, 1994.

[18] M. K. Cowles and B. P. Carlin. Markkov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904, 1996.

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

[20] V. De Gruttola and S. W. Lagakos. Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, 45:1–11, 1989.

[21] P. Dellaportos and A. F. M. Smith. Bayesian inference for generalised linear and proportional hazards models via Gibbs sampling. *Applied Statistics*, 42:443–459, 1993.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B*, 39:1–38, 1977.

[23] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

[24] B. Efron. The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565, 1986.

[25] C. P. Farrington and N. J. Gay. Interval-censored survival data with informative examination times: Parametric models and approximate inference. *Biometrika*, 83:467–471, 1999.

[26] D. M. Finkelstein. A proportional hazards model for interval censored failure time data. *Biometrics*, 42:845–854, 1986.

[27] D. M. Finkelstein, D. F. Moore, and D. A. Schoenfeld. A proportional hazards model for truncated AIDS data. *Biometrics*, 49:731–740, 1993.

[28] D. M. Finkelstein and R. A. Wolfe. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, 41:933–946, 1985.

[29] H. Frydman. A note on nonparmametric estimation of the distribution function from interval censored and truncated observations. *Journal of the Royal Statistical Society Series B*, 56:71–74, 1994.

[30] D. Gamerman. *Dynamic Analysis of Survival Models and Related Processes.* PhD thesis, University of Warwick, Coventry, U.K., 1987.

[31] D. Gamerman. Dynamic Bayesian models of survival. *Applied Statistics*, 40:63–79, 1991.

[32] D. Gamerman. Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, 85:215–227, 1998.

[33] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the Royal Statistical Society Series A*, pages 398–409, 1990.

[34] A. E. Gelfand and A. F. M. Smith. *Bayesian Computation.* John Wiley & Sons, New York, 1997.

[35] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7:457–511, 1992.

[36] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[37] P. K. Ghosh and D. Sinha. Analysis of interval-censored survival data using posterior likelihood. Research Report TR9525, University of Connecticut, Storrs, CT 06269-3120, 1995.

[38] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 1995.

[39] W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41:337–348, 1992.

[40] W. B. Goggins, D. M. Finkelstein, D. A. Schoenfeld, and A. M. Zaslavasky. A Markov Chain Monte Carlo EM algorithm for analyzing interval censored data under Cox proportional hazards model. *Biometrics*, 54:1498–1507, 1998.

[41] S. M. Gore, S. J. Pocock, and G. R. Kerr. Regression models and non-proportional hazards in the analysis of breast cancer survival. *Applied Statistics*, 33:176–195, 1984.

[42] P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81:515–526, 1994.

[43] R. J. Gray. A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, 50:244–253, 1994.

[44] Robert J. Gray. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of American Statistical Association*, 87:942–951, 1992.

[45] A. P. Grieve. Applications of Bayesian software: two examples. *The Statistician*, 36:283–288, 1987.

[46] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[47] N. A. Hessol, N. R. Lifson, P. M. O'Malley, L. S. Doll, H. W. Jaffe, and G. W. Rutherford. Prevelance, incidence, and progression of HIV infection in homosexual and bisexual men in hepatitis B vaccine trials. 1978-1988. *American Journal of Epidemiology*, 130:1167–1175, 1989.

[48] P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:387–396, 1986.

[49] J. D. Kalbfleisch. Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society Series B*, 40:214–221, 1978.

[50] J. D. Kalbfleisch and J. F. Lawless. Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association*, 84:360–372, 1989.

[51] J. D. Kalbfleisch and R. L. Prentice. Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60:267–278, 1973.

[52] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

[53] M. Y. Kim, V. G. De Gruttola, and S. W. Lagakos. Analyzing doubly censored data with covaraites with application to AIDS. *Biometrics*, 49:13–23, 1993.

[54] S. W. Lagakos, L. M. Barraj, and V. De Gruttola. Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika*, 75:515–523, 1988.

[55] C. A. McGilchrist and C. W. Aisbett. Regression with frailty in survival analysis. *Biometrics*, 47:461–466, 1991.

[56] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chemical Physics*, 21:1087–1091, 1953.

[57] J. C. Naylor and A. F. M. Smith. Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, 3:214–225, 1982.

[58] P. M. Odell, K. M. Anderson, and R. B. D'Agostino. Maximum likelihood estimation for interval censored data using a Weibull-based accelerated failure time model. *Biometrics*, 48:951–959, 1992.

[59] M. C. Paik, W. Y. Tsai, and R. Ottman. Multivariate survival analysis using piecewise gamma frailty. *Biometrics*, 50:975–988, 1994.

[60] W. Pan and R. Chappell. Estimating survival curves with left truncated and interval censored data under monotone hazards. *Biometrics*, 54:1053–1060, 1998.

[61] R. Peto. An experimental survival curve for interval censored data. *Applied Statistics*, 22:86–91, 1973.

[62] R. L. Prentice and L. A. Gloecker. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34:57–67, 1978.

[63] Z. Qiou. Multivariate survival data with positive stable frailties. Research Report TR9707, University of Connecticut, Storrs, CT 06269, 1997.

[64] A. E. Raftery and S. Lewis. How many iterations in the Gibbs sampler? pages 763–773. Oxford University Press, Oxford, U.K., 1992.

[65] A. E. Raftery, D. Madigan, and C. T. Volinsky. Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). pages 323–349. Oxford University Press, Oxford, U.K., 1996.

[66] P. M. Reilly. The numerical computation of posterior distributions. *Applied Statistics*, 25:201–209, 1976.

[67] G. O. Roberts and A. F. M. Smith. Simple conditions for the converegence of the Gibbs sampler and Metropolis-Hastings. *Stochastic Processes and their Applications*, 49:207–216, 1996.

[68] D. J. Sargent. A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics*, 54:1486–1497, 1998.

[69] G. A. Satten. Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, 83:355–370, 1996.

[70] D. Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69:239–241, 1982.

[71] D. Sinha. Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 88:979–983, 1993.

[72] D. Sinha. Semiparamtric Bayesian analysis of survival analysis. *Journal of the American Statistical Association*, 92:1195, 1997.

[73] D. Sinha. Time-discrete beta process model for interval censored survival data. *Canadian Journal of Statistics*, 25:445–456, 1997.

[74] D. Sinha. Posterior likelihood methods for multivariate survival data. *Biometrics*, 54:1463–1474, 1998.

[75] D. Sinha and D. K. Dey. Semiparametric Bayesian analysis of survival data. Technical Report TR9606, Department of Statistics, University of Connecticut, Storrs, United States, 1996.

[76] D. Spiegelhalter, A. Thomas, N. Best, and W. Gilks. BUGS 0.5: Bayesian inference Using Gibbs Sampling. manual (version *ii*). Downloadable from `http://www.mrc-bsu.cam.ac.uk/bugs/` or by anonymous ftp from `ftp.mrc-bsu.cam.ac.uk`, subdirectory `pub/methodology/bugs`, 1996.

[77] T. J. Sweeting. Apprximate Bayesian analysis of censored survival data. *Biometrika*, 74:809–816, 1987.

[78] M. A. Tanner. *Tools for statistical Inference.* Springer, New York, 1996.

[79] M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation (with comments). *Journal of the American Statistical Association*, 82:528–550, 1987.

[80] L. Tierney. Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, 22:1701–1762, 1994.

[81] A. A. Tsiatis. A large sample study of cox's regression model. *Annals of Statistics*, 9:93–108, 1981.

[82] B. W. Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society Series B*, 38:290–295, 1976.

[83] J. W. Vaupel, K. G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16,3:439–454, 1979.

[84] G. C. G. Wei and M. A. Tanner. Posterior computations for censored regression data. *Journal of the Royal Statistical Society Series B*, 85:829–839, 1990.