Some Statistical Methods for the Analysis of Survival Data in Cancer Clinical Trials

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy
by
Richard J. Jackson

14th August 2015

Abstract

Randomised Clinical Trials (RCT) are one of the most powerful tools of medical research and provide the basis for changing clinical practice. In oncology, the RCT is of particular importance in searching for new therapies and treatment approaches for patients diagnosed with cancer. Many of these trials have overall survival as a primary endpoint and are often designed with marginal effects being of clinical interest. As a result trials are typically large and are expensive in both time and money.

Given the substantial cost involved in running clinical trials, it is an ethical imperative that statisticians endeavour to make the most efficient use of any data obtained. A number of methods are explored in this thesis for the analysis of survival data from clinical trials with this efficiency in mind. Statistical methods of analysis which take account of extreme values of covariates are proposed as well as a method for the analysis of survival data where the assumption of proportionality cannot be assumed. Beyond this, Bayesian theory applied to oncology studies is explored with examples of Bayesian survival models used in a study of pancreatic cancer. Also using a Bayesian approach, methodology for the design and analysis of trial data is proposed whereby trial data are supplemented by the information taken from previous trials. Arguments are made towards unequal allocation ratios for future trials with informative prior distributions.

Contents

List of Figures

Acknowledgement

I would like to take this opportunity to acknowledge the tremendous support by both my parents and my supervisors, Dr. Trevor Cox and Dr. Catrin Tudur-Smith, without whose encouragement this would not have been possible. Further thanks are due to the Liverpool Cancer Trials Unit who gave me the opportunity to carry out this work.

I would further like to extend my heartfelt thanks to my flatmate Fray and my wife-to-be Cleo whose love, support and above all patience gave me the freedom to pursue my research.

"Measure what is measurable,
and make measurable what is not so."

GALILEO

Chapter 1

Introduction

1.1 Background

Cancer will effect more than one in every three people with 331,000 new diagnoses in the UK in 2011 alone [Source `http://www.cancerresearchuk.org`]. On average, an adult patient being diagnosed with cancer will have a 50% chance of surviving 10 years although this prognosis varies widely depending on the patient and the type of cancer.

The search for new treatment strategies is ongoing, with 552 clinical trials listed as open to recruitment by Cancer Research UK (CRUK) at the time of writing. Whilst many of these trials may be early Phase I or Phase II trials, Phase III trials are of the greatest interest with the aim of changing clinical practice and being considered the gold standard of evidence, providing '...one of the most powerful tools of clinical research' [1].

Clinical trials specific to oncology are typically conducted to assess the efficacy of one or more new treatments against the current clinical standard. Often trials are set to search for small or marginal improvements in patient performance and as a consequence can take years to run and may recruit hundreds of patients in order to provide sufficient evidence on which to base a conclusion.

Furthermore, many trials fail in the respects of identifying a new treatment to be superior to a current clinical standard. A recent review by Amiri and Kordestani [2] showed that 62% of a review of 235 published phase III trials failed to demonstrate statistical significance. There is therefore plenty of scope for new methodology to accurately assess therapies at an earlier stage and provide guidance as to the chances of a therapy being effective in a Phase III study.

Due to the severity of the disease, many such trials depend upon the evaluation of time-to-event endpoints, such as time to disease progression or ultimately, time to death. This thesis shall consider methods for the design and analysis of oncology trials with a time-to-event endpoint.

1.2 Bayesian methods in clinical trials

The use of Bayesian methodology in clinical trials is an attractive prospect as it is a natural framework under which greater efficiency may be obtained. In particular, Berry [3, 4] argues that a Bayesian approach can be more ethical and in keeping with scientific principles of accumulating information. A study by Perneger and Courvoiser [5] shows that medical professionals are more inclined to interpret results in a Bayesian fashion, and with the growing interest, direct comparisons of the theoretical and practical differences between Bayesian and Frequentist frameworks have been discussed [6, 7, 8].

However, despite the attractions of the Bayesian approach, as well as expectations of their growing influence (see for example Fleming and Yin [9]), clinical trials to date have been dominated by frequentist methodology. Part of this may be due to the desire to travel the '...path of least resistance' [10] as there is a vast and well established framework of methodology established in terms of sample size, trial design, interim analysis and trial analysis. Furthermore, Whitehead [1] and Howard [11] both argue that clinical trials should remain objective and not influenced by any prior information.

Despite these arguments, the anticipation of greater Bayesian influence has been noted. Lewis and Wears [12] provided an introduction to the benefits of a Bayesian approach with further discussions on the appropriate framework continued with Herson [13] introducing a series of four papers in the Statistics in Medicine Journal [14, 15, 16, 17] to discuss practical Bayesian approached to a multi arm trial. More recently, reviews by Ashby [18] and Grieve [19] detail the increased use of Bayesian methods over the past quarter of a century whilst Berger [20] describes an objective Bayesian approach to counter the perceived subjective nature of this approach compared to frequentist approaches.

With the recognition that Bayesian techniques can be more demanding, Speigelhalter et al. [21] along with Abrams [22] and Abrams and Ashby [23] provide practical applications of Bayesian methods in clinical trials. More recently a series of publication by the Clinical Trials journal [24, 25, 26] give an introduction of Bayesian theory to non statisticians.

With the advancement of Bayesian methodology coupled with advancements in computing power and software, which previously proved an impediment to all but the simplest of Bayesian analyses, there are few practical issues preventing further uptake of Bayesian methods as noted by Moye [27].

1.3 Aim

The aim of this thesis is to explore statistical methods that may be used to improve the accuracy and efficiency of the analysis of survival data in cancer clinical trials with a time-to-event endpoint. The main areas of interest under investigation are

- The analysis of survival data with outlying covariate values
- The analysis of survival data with non-proportional hazards
- The design and analysis of clinical trials with a time-to-event endpoint from a Bayesian perspective
- Efficient use of data through Bayesian clinical trial design

Throughout, the main emphasis will be to maximise the ability of investigators to assess the difference between two treatments through a single efficacy parameter.

1.4 Datasets

A summary is provided with regards to datasets that are to be used throughout this thesis.

1.4.1 ESPAC-3

Pancreatic Cancer is one of the most dangerous forms of cancer. Despite being the 10th most prevalent form of cancer, it is the 5th most common cause of death [Source: CRUK website `http://www.cancerresearchuk.org/cancer-info/cancerstats/mortality/cancerdeaths/#Twenty`]. The overall 5 year survival rate from the time of diagnose is $4-5\%$ [28, 29] which is due in part to pancreatic cancer being asymptomatic in the early stages. Most often, before a patient is diagnosed the cancer has advanced and although surgery can improve prognosis it is only possible for between 10% to 20% of all patients.

ESPAC-3 is an international multi-centre randomised phase III trial set up to investigate the use of chemotherapy as adjuvant, post surgery, therapy for patients with pancreatic cancer. The trial was essentially split into two separate trials dependent upon the type of tumours patients presented with. Patients with ducal pancreatic adenocarcinomas constitute the majority of the dataset (n=1090) with a second group consisting of patients who had ampullary and 'other' types of cancers (n = 431).

Aside from the the effect of the treatment regimen, other variable which are of interest are

- Resection Margins - Classed as negative or positive this determines if any cancerous cells are detected in the margins of a resected tumour following surgery
- Tumour Size - Given as the maximum of two perpendicular measurements
- Tumour Differentiation - Defined as how well a tumour resembles the tissue of origin, categorised as Well, Moderate and Poor

- Involved Lymph Nodes (N stage) - Defined as the presence/absence of cancerous cells in the lymph nodes
- Metastasis (M stage) - Defining whether or not the cancer has spread from the primary site to other part of the body
- Tumour (TNM) Stageing - A composite variable of N stage, M stage and evaluations of the primary tumour (T stage). Full definition given at `http://www.cancer.gov/cancertopics/factsheet/detection/staging`.
- World Health Organisation (WHO) Performance Status - A 5 point scale describing general patient health provided by the World Health Organisation .
- Cancer Antigen 19.9 (CA19.9) - A tumour marker known to be associated with pancreatic cancer.

Survival estimates are obtained via the method of Kaplan and Meier [30] and are shown in Figure 1.1 with confidence intervals obtained via Greenwood's formula [31] for both the 'Ductal' and the 'Ampullary/Other' patients. Figure 1.1 shows the improved survival outlook for the Ampullary/Other patients compared with the Ductal patients with respective median survival (95% confidence intervals) of 39.2 months (32.6, 50.0) and 21.2 (20.3, 23.4).

Analysis of the ductal patients has been previously published [32] and have shown that whilst there was no significant difference between the two chemotherapy regimens in the study, resection margins (included as a stratification factor), lymph node involvement, tumour differentiation, tumour size and WHO performance status are all significant prognostic indicators which affect overall survival. It should be noted that although patients in the Ductal group were randomised to an observational arm, these are not included due to previous results of the ESPAC-1 trial [33] showing that chemotherapy offers a survival benefit over observation only in this tumour group and so recruitment to this arm of the ESPAC-3 trial was stopped.

Analysis of the 'Ampullary' patients has also been published [34]. The main results show that for this tumour group, chemotherapy offers an improvement over observation only, although there was again no evidence of a difference between the two chemotherapy regimens.

The data from the ESPAC -3 trial has also been used to further evaluate the timing of the beginning of therapy following surgery and it's effects on patient prognosis [35]. Here it is shown that delaying the start of therapy may be beneficial to patients as it is more important to ensure that patients to receive the full intended course of planned therapy.

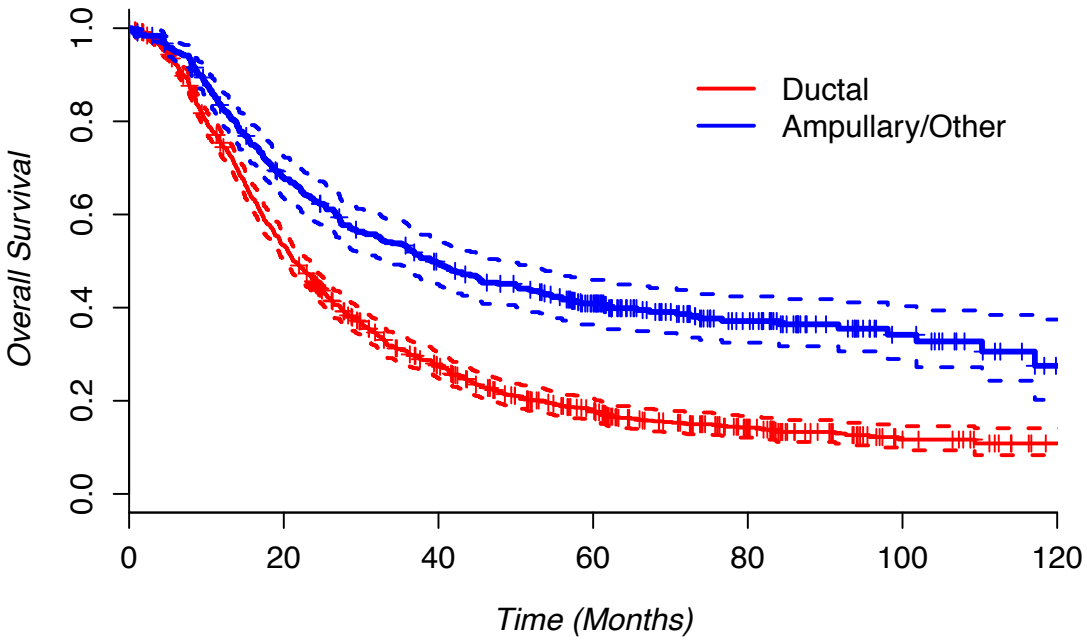

Figure 1.1: Kaplan Meier survival estimates for the 'Ductal' and 'Ampullary/Other' patients of the ESPAC-3 (V2) trial

1.4.2 Gastric cancer dataset

A second dataset used in this thesis is that taken from a trial investigating both chemotherapy and chemotherapy plus radiotherapy for the treatment of gastric cancer. Full details of the trial have been published [36].

Data are available from 90 patients and consist of survival time, censoring indicator and treatment arm only. Figure 1.2 shows a Kaplan Meier graph showing the survival estimates. The results from the trial are of particular interest statistically due to the crossing survival curves as this will violate one of the most common assumptions of proportional hazards in the analysis of survival data.

1.5 Discussion

This chapter serves as background to the research that is to be presented in this thesis. The rest of this thesis is structured as follows. Chapter 2 will provide an introduction to the statistical methods for the analysis of survival data that are to be used. Some practical discussion on the differences of the uses of frequentist and Bayesian survival models are also presented.

In Chapter 3, the problem of analysing survival data in the presence of extreme value covariates is explored and a method of altering the linear predictor of a Cox proportional hazards model to allow for more robust estimation is presented. Chapter 4 focusses

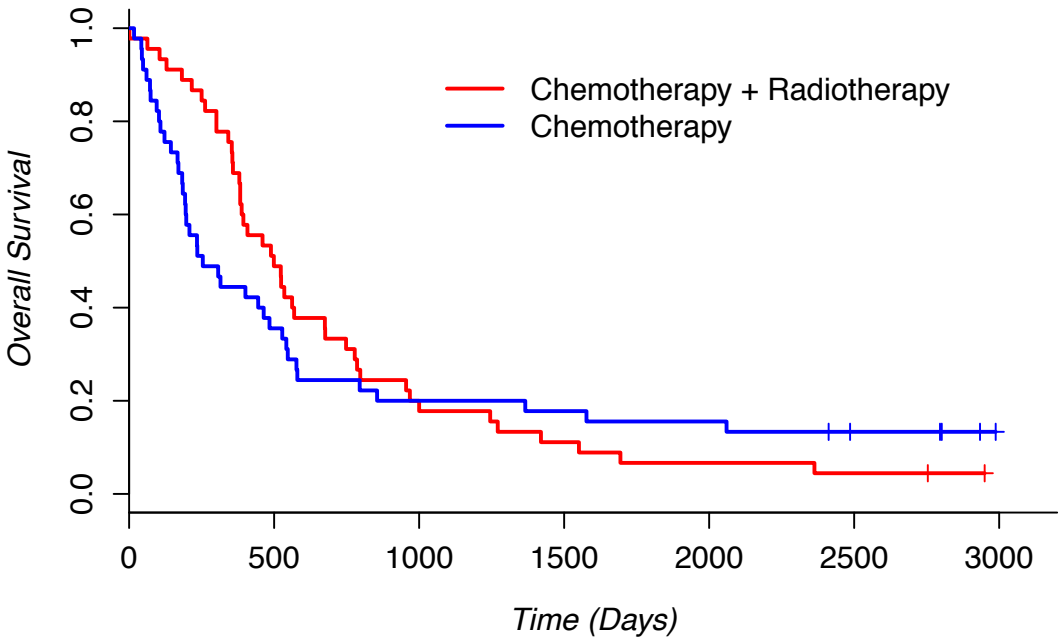

Figure 1.2: Kaplan Meier survival estimates for the chemotherapy and chemotherapy plus radiotherapy arms of a trial for patients suffering from gastric cancer

on the problem of non-proportional hazards. A review of methods for detecting non-proportionality and assessing treatment effects in this scenario are presented. Following this, a new method for modelling non-proportional survival models is proposed.

Bayesian analyses of survival data with applications to the ESPAC-3 trial are considered in Chapter 5 and some of the benefits over frequentist analyses are introduced. Chapter 6 introduces the concepts behind Bayesian sample size calculation and the ViP trial, currently running at the Liverpool Cancer Trials Unit, is considered from a Bayesian perspective. In Chapter 7, a method for deriving prior information from summary information of previously concluded trials is introduced and the effect on the design and analysis of a clinical trial explored. Chapter 8 explores allocation ratios that differ from the standard 1:1 in both a frequentist and Bayesian framework, again with applications to the design of the ViP trial. Some discussion and the scope for further work is given in Chapter 9.

Chapter 2

Analysis of Survival Data in Frequentist and Bayesian Frameworks

2.1 Introduction

In this chapter, a brief overview is provided of the differing viewpoints to analysing time-to-event data that are given by frequentist and Bayesian frameworks. Initially a summary is provided of some of the philosophical differences between the two approaches, following this an exploration is provided into the differing methods for analysing survival data which will be used throughout this thesis.

Primary focus is on proportional hazards models and an overview of the class of fully parametric models as well as the semi-parametric model defined by Cox [37] is provided. The piecewise exponential model first proposed by Friedman [38] and sometimes referred to as a Poisson regression model is explored in further detail as is a class of models that are defined using a counting process notation [39], shown to be related to the Cox model by Anderson and Gill [40].

Following exploration of these methods, a summary of the Gibbs sampling methodology as proposed by Gelfand and Smith [41] is described as a popular tool for estimating required densities for all but the simplest Bayesian models. Lastly, some practical issues for fitting proportional hazards models are discussed. Where appropriate, examples of the methodology are given by fitting models to the Ductal patients from the ESPAC-3 dataset.

2.2 An overview of frequentist and Bayesian methodology

Here a brief description of both the frequentist and Bayesian methodologies is given and some key comparisons highlighted.

2.2.1 Frequentist methodology

Much of the development of frequentist methodology is attributed to the work carried out by R.A. Fisher in the early 20^{th} century. As an example, consider the situation where data $\boldsymbol{x}$ are available which are modelled dependent upon some set of parameters $\boldsymbol{\theta}$. Under a frequentist framework, the data are considered to be produced from some data-generating function dependent upon the 'true' value of $\boldsymbol{\theta}$. An estimates of $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}$, is derived from a single observed realisation of the data from the generating function. Of import to note here is that the 'true' value of $\boldsymbol{\theta}$ is considered to be some fixed but unknown quantity with the data being considered a random variable. Much of frequentist methodology is then based on the theoretical basis of being able to continuously sample data from the same data generating function ad-infinitum and estimating the theoretical error that exists between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$.

Estimation of $\hat{\boldsymbol{\theta}}$ given the realised data is most often obtained using 'likelihood' theory. Generally it is assumed that the data are taken from some known distribution or family of distributions. Fixing $\boldsymbol{\theta}$ at some theoretical value, denoted $\tilde{\boldsymbol{\theta}}$, the probability of observing $\boldsymbol{x}$ is calculated given some assumed distribution. Based on the results, a value of $\tilde{\boldsymbol{\theta}}$ is then searched for that is most likely under the data and assumed likelihood and is denoted as $\hat{\boldsymbol{\theta}}$. Estimates of parameter precision are then estimated from the curvature of the likelihood function at this point.

Under standard notation, denote the likelihood as $L(\tilde{\boldsymbol{\theta}}|\boldsymbol{x})$. It is more common however to work on the log scale and define the log likelihood

$$l(\tilde{\boldsymbol{\theta}}|\boldsymbol{x}) = \log L(\tilde{\boldsymbol{\theta}}|\boldsymbol{x}). \tag{2.1}$$

2.2.2 Bayesian methodology

The introduction of Bayesian methodology can be attributed to a posthumous paper entailed 'An essay towards solving the Problem in the Doctrine of Chances' by the Reverend Thomas Bayes in 1763 [42]. Despite its early inception, frequentist methods remain the dominant methodology in modern statistics. In part this was due to the lack of modern sampling techniques that require substantial computing power (see for example [41]). Indeed for all but the simplest Bayesian models, computation methods were complex and it was often impossible to find any analytical solutions. Whilst advances in theory and computing have made Bayesian methods more accessible, their uses in practice still remain somewhat limited.

Considering again the situation of estimating parameter $\boldsymbol{\theta}$ associated with data $\boldsymbol{x}$, recall that in the frequentist framework, parameters are considered to be a fixed but unknown quantity with the data being a random variable. Here $Pr(\boldsymbol{x}|\boldsymbol{\theta})$ is evaluated where $Pr(.)$ represents some probability density. Bayesian methodology by contrast, does not concern itself with the data that may be observed if resampling were carried

out perpetually but considers the data, once observed, to be a fixed quantity and allows the parameters to be the random variables; thus instead evaluating $Pr(\boldsymbol{\theta}|\boldsymbol{x})$.

Evaluation of $Pr(\boldsymbol{\theta}|\boldsymbol{x})$ is determined via use of Bayes' theorem for conditional probability:

$$Pr(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{Pr(\boldsymbol{x}|\boldsymbol{\theta})Pr(\boldsymbol{\theta})}{Pr(\boldsymbol{x})}. \tag{2.2}$$

Under conventional notation, define

$$\begin{aligned} Pr(\boldsymbol{\theta}|\boldsymbol{x}) &\sim \text{posterior density} \\ Pr(\boldsymbol{x}|\boldsymbol{\theta}) &\sim \text{likelihood} \\ Pr(\boldsymbol{\theta}) &\sim \text{prior density} \\ Pr(\boldsymbol{x}) &\sim \text{marginal density.} \end{aligned}$$

All inferences are made on summaries of the posterior probability density. Note that the marginal density $P(\boldsymbol{x})$ is simply the probability of observing the data which is not dependent upon the parameter $\boldsymbol{\theta}$. Where interest lies only in the evaluation of $\boldsymbol{\theta}$, (2.2) is simplified to

$$Pr(\boldsymbol{\theta}|\boldsymbol{x}) \propto Pr(\boldsymbol{x}|\boldsymbol{\theta})Pr(\boldsymbol{\theta}). \tag{2.3}$$

From (2.3) the dogma of Bayesian methodology is obtained, that the posterior is a product of the information obtained from the data and the information taken from prior knowledge.

2.2.3 Comparisons

There are two key distinctions to be observed when comparing frequentist and Bayesian methodology. Firstly, the frequentist method takes all of the information about $\boldsymbol{\theta}$ only from the observed data. Bayesian methodology by contrast uses information from both the data and prior beliefs. A frequentist may argue that as Bayesian methods are dependent upon subjective beliefs, their analysis can never be truly objective and is therefore open to abuse. It should be noted however that firstly, prior densities are often set that allow only negligible amounts of information to enter an analysis and secondly, that given enough data, the amount of information in the data will far outweigh any information taken from prior densities. Furthermore the claim of objectivity is somewhat misleading as even in a frequentist framework, the results obtained from any model will depend upon the chosen likelihood and any associated assumptions that are required.

A second distinction is in the inference drawn in each framework. Frequentist inferences are typically made on quantities such as parameter estimates with associated

standard errors as well as P-values and 95% confidence intervals. The formal definition of a frequentist P-value is 'the probability of obtaining a value of a test statistic as or more extreme as the one actually observed, given that the null hypothesis is assumed to be true'. Whilst it is a probability statement, it is conditioned on both the 'null hypothesis' and data that were never observed.

Bayesian inferences are based on posterior distributions, these allow direct probability statements such as 'what is the probability that one treatment is better than the other?'. Table 2.1 provides a summary to illustrate some key differences in the interpretation of model parameters

Summary Measure	Frequentist	Bayesian Distribution
Point of central tendency	Parameter estimate	Posterior mean/median
Measure of spread	Parameter standard error	Posterior standard deviation
Interval estimation	Confidence interval	Credibility interval
Method of testing	P-values	Direct probability statements

Table 2.1: Definitions of the key methods of inference under frequentist and Bayesian frameworks

Finally, the point is made that in medical research, the tendency by medical professionals is to interpret statistical analyses as if they have been carried out in a Bayesian framework [5] leading to arguments that all statistical methods in medical research should be Bayesian. Despite this, frequentist approaches still provide the '...path of least resistance' [10] for day-to-day statistical procedures and as with many areas of research, practicalities outweigh any philosophical preference.

2.3 Computational methods of the analysis of proportional hazards models

In this section, exploration is given to the formulation of likelihoods for various forms of proportional hazards models. Introducing some notation, let T be a non-negative random variable representing an individual survival time with t being a realisation of that random variable. Define the hazard function, $h(t)$, as being the instantaneous risk of observing an event, so that

$$h(t) = lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{\Delta t}.$$

Let $f(t)$ and $F(t)$ be the density function and cumulative distribution function for T and so

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

The compliment of the distribution function is the survival function $S(T)$, which is often of primary interest and defined as

$$S(t) = Pr(T > t) = 1 - F(t).$$

Lastly define the cumulative hazard function $H(t) = \int_0^t h(u)du$ and note the identities

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

When modelling proportional hazards data, it is generally assumed that the hazard for an individual or group of individuals is related to some baseline hazard function with

$$h(t|x) = h_0(t)G(\boldsymbol{z}, \boldsymbol{\theta}),$$

where $h_0(t)$ is the baseline hazard function and $G(.)$ is some non-negative function of some covariates z and parameters θ. Traditionally set $G(\boldsymbol{z}, \boldsymbol{\theta}) = \exp(\boldsymbol{z}^T\boldsymbol{\beta})$, where $\boldsymbol{\theta} = \boldsymbol{\beta}$ [37]. Here and throughout, β shall be used to represent the log hazard ratio for a covariate z. This is attractive as it allows $\exp(\beta)$ to be expressed as a hazard ratio and defines the multiplicative increase/decrease in the risk of observing an event due to a unit increase in z.

Under the special case of all events being observed, a likelihood for estimating β for $i = 1, ..., n$ patients is defined as

$$L(\beta|z, t) = \prod_{i=1}^{n} f(t_i).$$

This is simply the product of the probabilities that patient i survived up until time t_i. The analysis of survival data is typically complicated by the presence of censored data however. When data are censored, the exact time of an event is not known, but it may be known that an event occurred before some point (left censored), after some point (right censored) or between two points (interval censored) in time. In this thesis, only right censored data are considered and it is assumed that the censoring mechanism is completely independent of the observed event times. Denote C as the random variable for censoring observations and define $\nu_i = I(T_i < C_i)$ where $I(.)$ is the indicator function. Further denote the observed data as $D = \{D_i\}$, $D_i = (t_i, \nu_i, z_i)$ and, allowing for censoring, re-define the likelihood as

$$L(\beta|D) = \prod_{i=1}^{n} f(t_i)^{\nu_i} S(t_i)^{(1-\nu_i)} = \prod_{i=1}^{n} h(t_i)^{\nu_i} S(t_i). \tag{2.4}$$

In the following sections, some of the methods that have been developed to estimate parameters in survival models are evaluated which will be utilised throughout this thesis.

2.3.1 Parametric models

Parametric survival models are characterised by the assumption that the density function follows some known distribution. They have been widely discussed and used in

practice, see for example [43, 44]. A further review is available at `https://files.nyu.edu/mrg217/public/parametric.pdf`. Table 2.2 gives details for some of the most commonly used distributions, though this is by no means exhaustive. Presented are the hazard functions and the survival functions from which full likelihoods are formed.

Distribution	h(t)	S(t)
Exponential	λ	$\exp(-\lambda t)$
Weibull	$\lambda\rho t^{(\rho-1)}$	$\exp(-\lambda t^{\rho})$
Gompertz	$\lambda\exp(\rho t)$	$\exp\left\{\frac{-\lambda}{\rho}(\exp\{\rho t\}-1)\right\}$
Log-Logisitic	$\frac{\lambda\rho t^{\rho-1}}{1+(\lambda t)^{\rho}}$	$\frac{1}{1+(\lambda t)^{\rho}}$
Lognormal	$\frac{\frac{1}{t\rho\sqrt{2\pi}}\exp\left[\frac{-1}{2\rho^2}\{\ln(t)-\ln(\lambda)\}^2\right]}{1-\Phi\left\{\frac{\ln(t)-\ln(\lambda)}{\rho}\right\}}$	$1-\Phi\left\{\frac{\ln(t)-\ln(\lambda)}{\rho}\right\}$

Table 2.2: Definitions of hazard and survival functions for a selection of parametric models; note that Φ represents the standard normal distribution function

From Table 2.2, likelihoods for a wide range of models can be easily defined and routines exist in all statistical packages to obtain parameter estimates. Further, note that the exponential model can be expressed as a special case of the Weibull model with $\rho = 1$. To illustrate the uses of parametric models the likelihoods of the Weibull, Log-Logistic and Lognormal models are fitted for patients from the ESPAC-3 dataset. Models are fit using the 'survreg' function in R and results are presented on the log scale in Table 2.3. Note here that as the ρ parameter for the Weibull model is close to one, then a simpler exponential model may be justified in this case.

Distribution	$\log\lambda$	$\log\rho$
Weibull	-3.78 (0.029)	-0.02 (0.024)
Log-Logisitic	-3.31 (0.031)	0.14 (0.022)
Lognormal	3.28 (0.030)	-0.42 (0.025)

Table 2.3: Model parameters for Weibull, Log-Logistic and Lognormal parametric survival models. Results are presented in the form of means (standard errors)

Each model is assessed graphically via calculation of the survival function. This is compared against the Kaplan Meier [30] estimates and presented in Figure 2.1. It is seen here that the closest fit to the non-parametric survival estimates is obtained by the lognormal model. This model may still not provide an adequate fit however and will produce consistently larger survival estimates between c15 - 60 months than the non-parametric estimate (the converse is true outside of this range). Whilst fully parametric models can be advantageous due to the information they provide about the baseline hazard function, problems can ensue if the observed data do not follow any particular distribution.

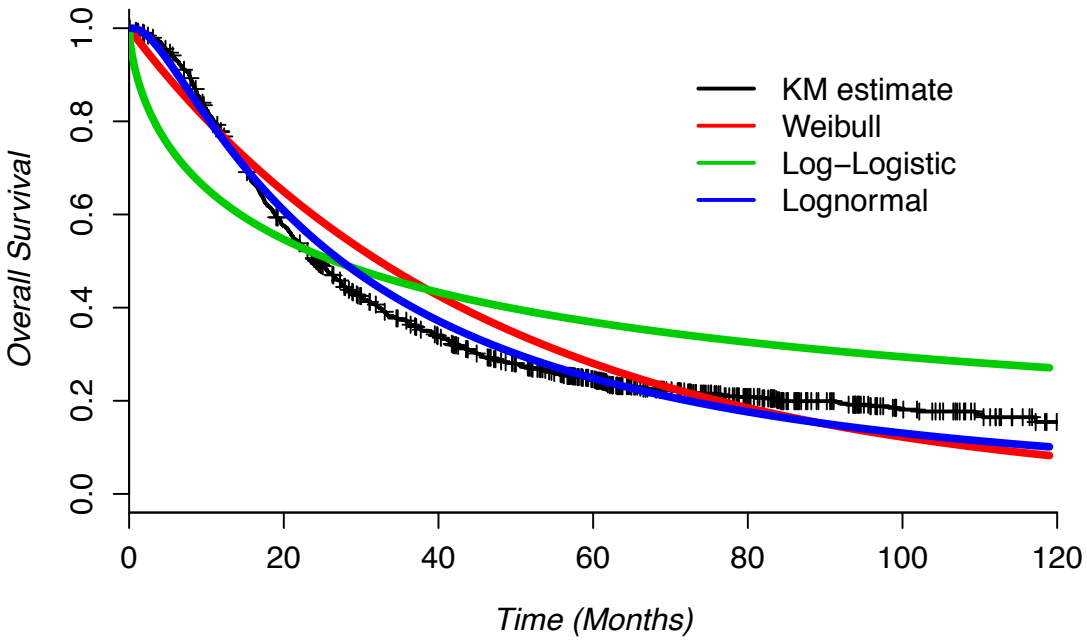

Figure 2.1: Fitted parametric curves to ductal patients from the ESPAC-3 dataset

2.3.2 Cox's semi-parametric model

Cox introduced his proportional hazards model in 1972 [37] and as of 2005, it was the second most widely cited statistical paper (source [45]), second only to the publication by Kaplan and Meier [30]. Under standard proportional hazards modelling, define the hazard for each observation i as

$$h_i(t) = h_0(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{z_i}\},$$

where z_i is a vector of covariates for patient i. Cox demonstrated that estimation of the key parameters of interest, $\boldsymbol{\beta}$, could be carried out without the need to specify a baseline hazard function. Parameter estimation is carried out via a partial likelihood which states that the probability of observing an event for patient i at time t is the ratio of the hazard function for patient i against the sum of the hazards for all other patients at risk of an event at time t. That is for patient i, assuming no tied survival times, the likelihood is defined by

$$\frac{h_i(t)}{\sum_{j \in R} h_j(t)}.$$

Here the summation over R refers to all patients at risk at time t. As the baseline hazard function is considered equal for all observations, this cancels from both the numerator and the denominator. Taking the product over all patients gives the likelihood

$$L(\boldsymbol{\beta}|D) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{z_i})}{\sum_R \exp(\boldsymbol{\beta}^T \boldsymbol{z_i})}. \tag{2.5}$$

Note the likelihood does not explicitly depend upon time, only the ordering of the observed event times. This is referred to as a semi-parametric method because it still utilises asymptotic parametric assumptions to make inferences about $\boldsymbol{\beta}$ but leaves the baseline hazard function completely unspecified. Some alterations of the likelihood have been proposed for the occurrence of tied event times, see for example Efron's method [46].

Cox's model has become extremely popular, especially in medical statistics as often the question of main interest lies in comparing two groups of patients, for example two sets of patients given two different treatments in a clinical trial. In this context, the main question of interest is which group performs best, which is shown by the hazard ratio. Here then there is little need to understand the baseline hazard function so long as the assumption of proportionality is satisfied.

2.3.3 Piecewise exponential models

In this section, the piecewise exponential model (PEM) first proposed by Friedman [38] is explored. Though this is a fully parametric model, it is included separately due to the added complexity.

To understand the basic premise of the PEM, considered initially the simple exponential model with definitions given in Table 2.2. Here it is assumed that the hazard rate is constant and independent of time. The PEM extends the simple exponential model by partitioning the time domain using some 'time-grid' and assuming only that the hazard rate is constant within each partitioned interval.

The extra flexibility in the PEM has made it a popular alternative for modelling survival data when some estimate of the baseline hazard function is required and it has been shown by Breslow [47] to be analytically equivalent to the Cox model when the time-grid is defined by each observed event. The PEM can also be related to a class of models referred to as change point-models (see for example [48]) whereby the underlying hazard of a group of patients is considered to change at given points in time and the aim is usually to estimate the time-points at which a change occurs. The PEM is considered here as a means of estimating an accurate baseline hazard function and retain interest in modelling differences by means of the hazard ratio.

PEMs have become particularly popular in a Bayesian framework as they provide sufficient flexibility to accurately estimate a baseline hazard function whilst allowing the user to limit the number of parameters required to describe the baseline hazard function. In large datasets in particular, this can help to reduce the computational burden required in fitting survival models. A Bayesian approach was proposed by

Gammerman [49] who also explored a dynamic approach to estimating hazard functions [50] as well as Zelterman et al. [51] who explored smooth transitions between time-partitions. Malla and Mukerjee [52] consider an estimator for the PEM which allows for reliable estimation beyond the last observed event. Usually a constant hazard ratio across all partitions is assumed but allowing the hazard ratio to vary with time can be applied as has been shown by Sinha [53]. In practice, Koissi [54] used a PEM in a Bayesian framework with an added frailty component to study child mortality in the Ivory Coast. PEMs have more recently been applied to the analysis of case-control data [55] and meta-analysis [56] in both a frequentist and Bayesian framework.

Whilst the PEM is an attractive option for the analysis of survival data, inferences upon the hazard ratio can not always be considered to be independent from the choice of time-grid. Some popular methods are the aforementioned Breslow [47] method or the approach by Kalbfleisch [57] who argued for the specification of time-grids to be defined prior to any analysis to avoid any chance of bias in ensuing parameter estimates. These and further methods are explored in Chapter 5.

To define the PEM, consider initially the standard parametric exponential model with density function

$$f(t_i|\theta) = \theta \exp(-\theta t_i),$$

and associated survival function

$$S(t_i|\theta) = \exp(-\theta t_i).$$

The likelihood function for θ is written as

$$L(\theta|D) = \prod_{i=1}^{n} f(t_i|\theta)^{\nu_i} S(t_i|\theta)^{1-\nu_i} = \theta^{\eta} \exp\{-\theta\xi\}$$

where $\eta = \sum_{i=1}^{n} \nu_i$ and $\xi = \sum_{i=1}^{n} t_i$. All information regarding covariates and parameter estimates enter the model via $\theta = (\gamma + \boldsymbol{\beta}^T \boldsymbol{z})$ where $\gamma = \log \lambda$. Here γ is the parameter associated with the baseline hazard rate and the parameters $\boldsymbol{\beta}$ explains the hazard ratios associated with covariates $\boldsymbol{z}$.

The standard exponential model therefore has a single parameter γ which incorporates all information relating to the baseline hazard. The PEM extends the standard model by partitioning the time axis into a J smaller intervals. Throughout this thesis, it is assumed that there exists a proportional relationship in which $\boldsymbol{\beta}$ is constant across all partitions, that is $\boldsymbol{\beta_j} = \boldsymbol{\beta}$ for all $j \in J$. Following the definition of Ibrahim et al. [58] the likelihood is defined as:

$$L(\lambda, \beta|D) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left(\lambda_j \exp(\boldsymbol{\beta}^T \boldsymbol{z}_i)\right)^{\delta_{i,j}\nu_i} \exp\left\{ -\delta_{i,j}\Big[\lambda_j(t_i - a_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(a_g - a_{g-1})\Big] \exp(\boldsymbol{\beta}^T \boldsymbol{z}_i)\right\}. \quad (2.6)$$

Here $\delta_{i,j}$ is an indicator variable which takes the value 1 if the i^{th} patient has an event in the j^{th} interval and zero otherwise. The time grid is defined by a_j for $\{j = 1, ..., J\}$ and for completeness define $a_0 = 0$. Further note the special case of $J = 1$ where equation (2.6) reduces to the standard exponential model.

The behaviour of the PEM is illustrated by making use of the ESPAC-3 data. To fit the model, a somewhat arbitrary time grid of $a = \{0, 6, 12, 24, 48, 72, 96\}$ is set. The model is fitted in 'R (Version 3.0)'. Despite its popularity, methods for fitting the PEM are not directly available in all statistical packages. Work carried out by Laird and Oliver [59] however provide an illustration as to how the PEM can be fitted using standard generalised linear model techniques. This is based on the observation that the likelihood formulation can be equivalent to that of a Poisson distribution where the event indicator is the response variable and the logarithm of time is included as a model offset. This formulation also facilitates the fitting of the PEM in a Bayesian framework using the statistical package WinBUGS. See the Appendix for a function written in R which facilitates the organisation of data and fitting of the PEM as well as code for model fitting in WinBUGS.

Plots of the survival functions from both the standard exponential model and the PEM are given in Figure 2.2 for the patients with ductal adenocarcinomas from the ESPAC-3 data along with the set time-grid. Here, the extra flexibility gained by assuming a piecewise constant hazard rate allows the fitted model to closely resemble the Kaplan Meier survival estimates.

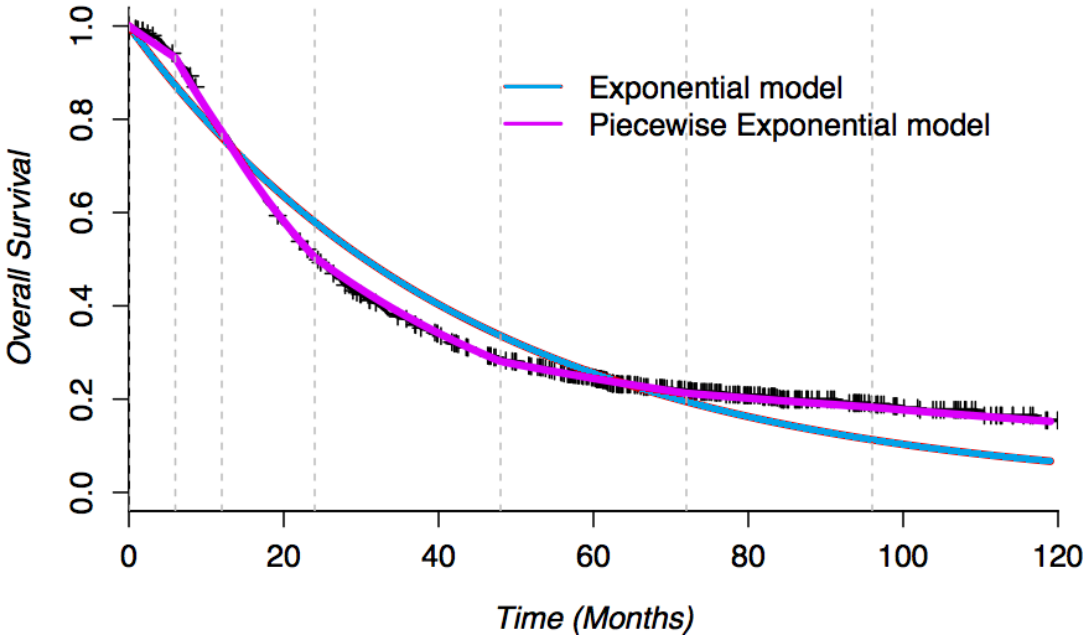

Figure 2.2: Fitted exponential and piecewise exponential survival curves to the ESPAC-3 dataset

Exploration of the PEM for analysing time-to-event data from clinical trials shall be given in Chapters 5-8 when considering the application of Bayesian methods to the design and analysis of clinical trials.

2.3.4 Counting process models

Using counting process notation to model survival data is a popular approach due to the flexibility allowed. Fully parametric, semi-parametric and non-parametric approaches can all be encompassed within this framework. A detailed descriptions of counting processes are provided by Fleming and Harrington [60] with applications to survival data being provided by Anderson et al. [61]. Throughout this thesis, the counting processes used are based on the approach developed by Anderson and Gill [40] where the baseline hazard function is estimated non-parametrically but the parameters associated with patient covariates are estimated parametrically. Anderson and Gill demonstrate that this approach is an analogue of the proportional hazards model defined by Cox [37] and further show that estimated parameters are asymptotically efficient.

Whilst concentrating primarily on counting processes for right censored survival data, they have been used as a platform for more complex models. Anderson [61] gives examples of counting processes for survival data illustrating their use for modelling frailty components, whilst Clayton [62, 63] demonstrated Monte Carlo estimation procedures for Bayesian inference and their use for the analysis of recurrent event data. Further developments of counting process models have also been carried out by Chen [64] and Cheng [65] who consider the survival or hazards function to follow a transformation model where the proportional hazards model is a special case of a wider class of models.

Taking previous notation, assume T_i to be a random variable representing the survival time for observation i and further define C_i to be a random variable representing censoring time. An event is observed for the i^{th} observation whenever $T_i \leq C_i$. Define $t_i = min(T_i, C_i)$ and $\nu_i = I(T_i < C_i)$ and note that for observation i with associated covariates $\boldsymbol{z}_i$, the data $D_i = (t_i, \nu_i, \boldsymbol{z}_i)$ are observed.

Define a counting process, $N(t)$, as a right continuous process with 'jumps' of size 1 which counts the number of events that have occurred up to some time t. Further, simultaneously observe the 'at risk process', $Y(t)$, which defines the number of observations that are candidates for an event at any given point in time. In its general form, the counting process is considered as representing a group of homogeneous observations. Concentration here is on the multivariate counting process and for each $i = 1, 2, ..., n$ observations define

$$N_i(t) = I_{\{t_i \leq t, d_i = 1\}}$$
$$Y_i(t) = I_{\{t_i \geq t\}}.$$

Primarily, interest lies in modelling the 'intensity' function which in a survival context is an analogue to the hazard function. This is essentially the probability of there being a jump in the counting process over some small interval between t and $t+\delta t$ conditional on all information gathered prior to t. The conditional component of this probability is referred to as the 'filtration', labeled $\mathcal{F}_{t-}$. The intensity process for a single observation is then

$$\alpha_i(t)\delta t = Pr\{[N(t+\delta t) - N(t)] = 1|\mathcal{F}_{t-}\}.$$

The cumulative intensity function labelled $A(t)$ is further defined as

$$A(t) = \int_0^t \alpha(u)\delta u.$$

Note the intensity function is a random variable, as is the filtration, and that $Y(t)$ is a fully predictable process. Many of the properties of the counting process are based upon the observation that the process defined by $N_i(t) - A_i(t)$ is a Martingale which is denoted $M_i(t)$. That is to say, the expected value of the process M at some point $t_{(i+1)}$ is the value of the process at the previous point in time, t_i. Formally

$$E[M(t_{i+1})|M(t_i), M(t_{i-1}), ..., M(t_0)] = M(t_i).$$

A non-parametric of estimate of the cumulative intensity function, $\hat{A}(t)$, was proposed by Aalen [66], given by

$$\hat{A}(t) = \int_0^t \frac{I_{\bar{Y}(u)>0}}{\bar{Y}(u)} d\bar{N}(u).$$

For illustration, Figure 2.3 takes a random sample of 10 patients from the ESPAC-3 dataset and illustrates graphically the behaviour of the counting process $N_i(t)$, the at risk process $Y_i(t)$ and the cumulative intensity estimate using the Nelson Aalen estimator $\hat{A}$.

Under the special case of right censored survival data, the intensity process becomes an analogue of the hazard function. From this, it is apparent that a survival function can be defined via $S(t) = \exp\{-A(t)\}$.

An attractive feature of the counting process notation is illustrated by the definition of the multiplicative intensity model in a survival context as given by Aalen [66]. Given the observed data D and parameters θ, define a cumulative intensity/hazard function as

$$A_i(t, \theta) = \int_0^t \alpha_i(u, \theta) Y_i(u) du$$

Here the cumulative intensity function is dependent upon two structures, the first is some unknown function of the baseline hazard function and a possible set of covariates

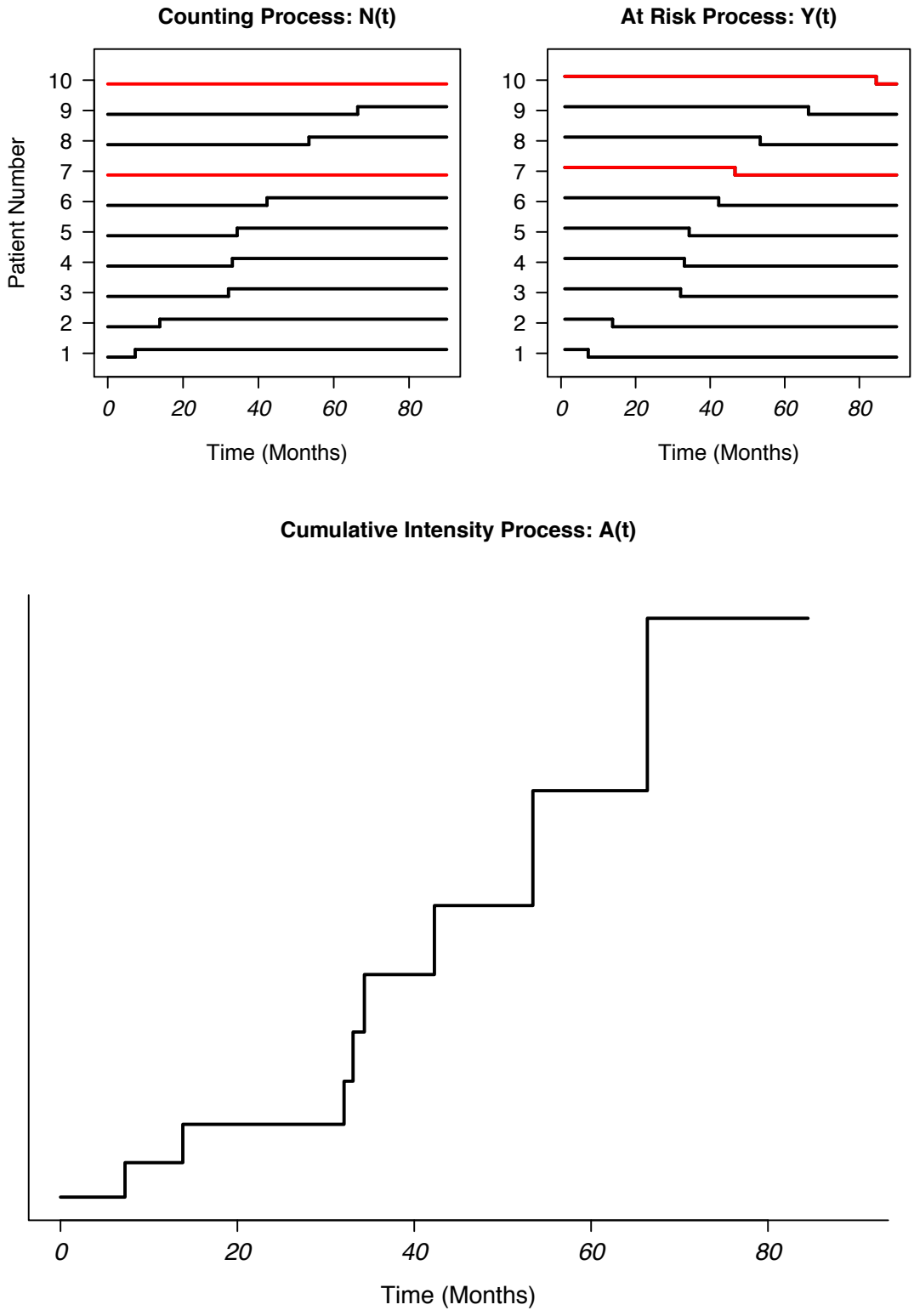

Figure 2.3: Illustrations of the counting process, at risk process and the cumulative intensity process. Red lines indicate censored patients within the trial.

and secondly the at risk process $Y_i(t)$. In the context of survival analysis, $\alpha(t)$ can be defined to incorporate more than one type of event and hence model competing risks. Furthermore by varying definitions of $Y(t)$ (and by extension, $N(t)$) models with complex censoring mechanisms and/or recurrent events can be defined.

For the inclusion of covariates, unless otherwise specified, assume that the standard proportional hazards relationship exists and define

$$\alpha_i(t) = \alpha_0(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{z}_i(t)\}.$$

Note that in this form $\boldsymbol{z}$ is allowed to vary over time. Here $\alpha_0(t)$ is some baseline intensity process which can take a parametric form although only non-parametric estimation is considered here. Following the definitions given earlier in this section, let

$$S(t) = \exp\{-A(t)\},$$

and further define $dN_i(t)$ to be

$$dN_i(t) = \begin{cases} 0 & \text{if patient } i \text{ does not experience an event at time } t \\ 1 & \text{if patient } i \text{ does experience an event at time } t. \end{cases}$$

This naturally leads to the likelihood formulation given by

$$L(\theta|D) = \prod_{i=1}^{n} \Big\{ \prod_{t \geq 0} \alpha_i(D_i, \theta)^{dN_i(t)} \Big\} S_i(D_i, \theta).$$

In order that the likelihood is correctly specified, it is ensured that the baseline hazard function is specified in terms of a step function, and $\alpha_0(t)$ is specified by $\Lambda_0\{t\}$. Further specify t^* as the maximum observed event time given by $\sup\{t : \delta N_i(t) = 1, i = 1, ..., n\}$ and take the log to obtain

$$l(\theta|D) = \sum_{i=1}^{n} \int_0^{t^*} \log[\Lambda_0\{u\} \exp\{\boldsymbol{\beta}^T \boldsymbol{z}(u)\}] dN(u) - \int_0^{t^*} \Lambda_0\{u\} \exp\{\boldsymbol{\beta}^T \boldsymbol{z}(u)\} Y_i(u) du. \quad (2.7)$$

Note here that the first integral is only evaluated over patients with an observed event whereas the second integral is evaluated over all observations. Throughout this thesis, the formulation of Anderson and Gill is considered whereby the baseline hazard function is considered to be non-parametric but the parameters associated with the covariates, $z(t)$ are estimated parametrically. In this case, estimation of parameters of the likelihood given by (2.5) can be achieved via the use of Non-Parametric Maximum Likelihood Estimators (NPMLE) where each step in the counting process, $\Lambda_0\{t\}$, is treated as a parameter to be estimated.

In this thesis the counting process will be used in Chapter 4 when some exploration is given to widening the class of models available to explain the relationships between patient hazards in randomised controlled trials.

2.4 Bayesian estimation

Up to this point, the analysis of survival data has been considered in a frequentist framework with the development of likelihood formulations under differing scenarios. In this section, some discussion is provided regarding the fitting of models in a Bayesian framework. Initially, the special case of the parametric exponential model with no covariates and a gamma prior distribution is considered. Though somewhat limited, this special case is used to re-enforce the basic concepts of Bayesian methodology. Following this, an illustration is given regarding the difficulties encountered when the model is extended to include covariates. Finally, an exploration of sampling methods is given, in particular the Gibbs sampler, for providing solutions to more complex models.

2.4.1 Exponential model with gamma priors - no covariates

Considering the parametric Exponential model with no covariate and data $D_i = (t_i, \nu_i)$, the likelihood is defined as

$$L(\lambda|D) = \prod_{i=1}^{n} \lambda^{\nu_i} \exp\{-\lambda t_i\}.$$

In this simple case it is shown that the total observed survival time, $\zeta = \sum_i^n t_i$ and the total number of events $\eta = \sum_i^n \nu_i$ are sufficient for estimating λ with a solution provided by $\hat{\lambda} = \eta/\zeta$.

Consider the special case of a prior distribution for the hazard parameter λ, $Pr(\lambda)$, which follows a gamma distribution with scale and shape parameters (ω, ξ). The density function is given by

$$Gamma(\omega, \xi) \sim \frac{\xi^{\omega} \lambda^{\omega-1} \exp\{-\xi\lambda\}}{\Gamma(\omega)}.$$

Here $\Gamma(.)$ is the gamma function. This prior distribution is shown to be conjugate for the exponential distribution as the posterior distribution itself also has a gamma distribution. To see this, evaluate the posterior distribution as

$$\begin{aligned} Pr(\lambda|D) &\propto Pr(D|\lambda)Pr(\lambda) \\ &= (\lambda^{\eta} \exp\{-\lambda\zeta\})\lambda^{\omega-1} \exp\{-\xi\lambda\} \\ &= \lambda^{\eta+\omega-1} \exp\{-\lambda(\zeta+\xi)\}, \end{aligned}$$

and thus $Pr(\lambda|D)$ follows a gamma distribution given by $Gamma(\eta+\omega, \zeta+\xi)$. The posterior distribution can be directly summarised using

$$E(\lambda|D) = \frac{\eta+\omega}{\zeta+\xi}$$

and

$$Var(\lambda|D) = \frac{\eta+\omega}{(\zeta+\xi)^2}.$$

It is shown therefore that the posterior summaries are a direct combination of the observed data and prior beliefs. As an example, an exponential model is applied to the ductal patients in the ESPAC-3 data and the hypothetical situation where prior information is available which takes the form $Gamma(300, 15000)$ is considered.

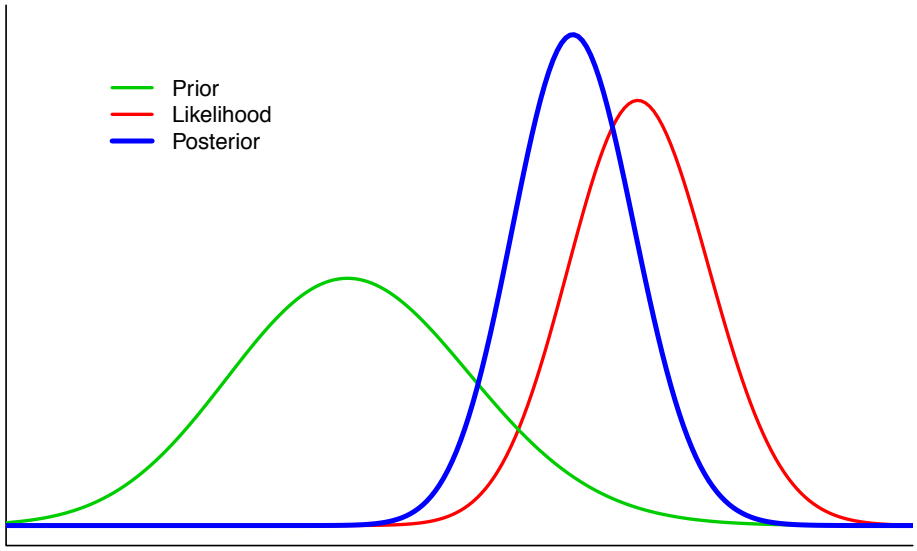

Figure 2.4: Prior, likelihood and posterior densities for an Exponential model fitted to the ESPAC-3 dataset

The fit of the exponential model to the ESPAC-3 data is shown in Figure 2.2. In Figure 2.4 the densities of the hazard parameter λ from the prior, likelihood and the posterior are shown. This shows how the inclusion of an informative prior compliments the information from the likelihood. In this situation, there is a large amount of information in the data and a highly informative prior is required to have any noticeable effect on the posterior distribution. What is notable here is not only the shift in the point estimation but also the increase in the precision when comparing the posterior distribution to the estimated obtained from the likelihood alone.

2.4.2 Exponential model with gamma priors - with covariates

Here it is demonstrated that difficulties ensue when the exponential model is extended to include a single covariate, z.

Formally define the likelihood for parameters $\theta = (\lambda, \beta)$ and data $D = (t_i, \nu_i, z_i)$ as

$$L(\theta|D) = \prod_{i=1}^{n} \left(\lambda \exp\{\beta^T z_i\}\right)^{\nu_i} \exp\left\{ -\lambda \exp\{\beta^T z_i\} t_i \right\}.$$

Following the Bayesian paradigm, prior distributions for both λ and β are required.

Following the laws of conditional probability define

$$\begin{aligned} Pr(\theta|D) &\propto Pr(D|\theta)Pr(\theta) \\ &= Pr(D|\lambda,\beta)Pr(\lambda|\beta)Pr(\beta) \\ &= Pr(D|\lambda,\beta)Pr(\lambda)Pr(\beta), \end{aligned}$$

if it is assumed a-priori that the parameters that describe the baseline hazard and the parameters that describe effects of covariates are independent.

Following the example of Section 2.4.1, set a prior for λ based on a gamma distribution and further set the prior for β to be a normal distribution. Note from (2.3) that to evaluate posterior distributions, only the terms directly dependent upon θ are required and define

$$\begin{aligned} Pr(\lambda) &\propto \lambda^{\omega-1}\exp\{-\xi\lambda\} \\ Pr(\beta) &\propto \exp\Big\{-\frac{(\beta-\mu)^2}{2\sigma^2}\Big\}. \end{aligned}$$

A full form for the posterior distribution is written as

$$P(\theta|D) \propto \prod_{i=1}^{n} \lambda^{\nu_i+\omega-1}\big(\exp\{\beta^T z_i\}\big)^{\nu_i} \exp\Big\{-\Big(\lambda\big(\exp\{\beta^T z_i\}t_i+\xi\big)+\frac{(\beta-\mu)^2}{2\sigma^2}\Big)\Big\}.$$

Of primary interest are the marginal densities for λ and β from which posterior inferences can be made, denote as

$$\begin{aligned} Pr_\lambda(\theta|D) &= \int Pr(\theta|D)d\beta \\ Pr_\beta(\theta|D) &= \int Pr(\theta|D)d\lambda. \end{aligned}$$

Given the form of the likelihood this is not straightforward even in this relatively simple situation. Estimation can be achieved by approximating the posterior distribution for instance via a Laplace transformation or via simulation techniques which shall be explained in the following Section.

2.4.3 Monte Carlo Markov Chain simulation

Markov Chain Monte Carlo (MCMC) simulation is a class of stochastic algorithms for sampling from probability densities. They follow the Monte Carlo property that states that given some probability distribution $\pi(\theta)$, a sequence of random observations $\theta^1, \theta^2...$ can be drawn where at any point m, the distribution of θ^m depends only upon θ^{m-1}. Using techniques such as this in a Bayesian framework facilitated the construction of marginal posterior distributions using a large number of samples taken directly from the joint posterior distribution. These samples can then be directly used to obtain inferences upon key parameters of interest.

Whilst there are a number of differing MCMC routines that can be applied, consideration is given here to the Gibbs sampler first proposed by Gemen and Gemen [67] and applied in a Bayesian setting by Gelfand and Smith [41] as well as the Metropolis Hastings rejection sampling routine [68, 69]. Greater details of both these methods are described by Gelmen et al. [70]. A direct application of the Gibbs sampler for proportional hazards models is given by Dellaportas and Smith [71].

With respect to the Gibbs sampler, suppose there exists a posterior distribution $Pr(\mathbf{\Theta}|D)$ where the parameter vector can be divided into P components $\mathbf{\Theta} = (\Theta_1, ..., \Theta_P)$. The Gibbs sampler is a routine which, at each iteration, draws a sample from each component of $\mathbf{\Theta}$ conditional on the values of all other components. Allow the sequence of iterations m to be denoted as superscripts and the parameter components P to be denoted as subscripts. At each iteration a sample is randomly generated from the probability distribution given by

$$Pr(\theta_p|\Theta_{-p}^{m-1}, D).$$

where Θ_{-p}^{m-1} gives the most recent state of each component other than p at iteration m such that

$$\Theta_{-p}^{m-1} = (\Theta_1^m, ..., \Theta_{p-1}^m, \Theta_{p+1}^{m-1}, ..., \Theta_P^{m-1}).$$

Under this routine each draw of component Θ_p^m is updated dependent upon both the previous value Θ^{m-1} and the most recent states of all other components of Θ_{-p}^{m-1}.

It is sometimes the case that for a given posterior density, marginal distributions for each Θ_p will be known analytically in which case the sampling procedure can be used directly. When it is not, rejection algorithms are required, such as the Metropolis Hastings algorithm of which the Gibbs sampler is a special case as shown by [70].

The Metropolis Hasting algorithm, [68, 69] works on the basis that an evaluation of the state of Θ_p^m with respect to Θ_{-p}^{m-1} can be made by drawing a proposed future state Θ_p^* from a 'jumping' distribution $J_m(\Theta^*|\Theta^{m-1})$. Note that the jumping distribution is conditional on Θ_p^{m-1} only and not on the previous state for the other parameters being evaluated Θ_{-p}^{m-1}.

Taking a draw from the jumping distribution, the posterior distribution is evaluated under the proposal state and the previous state for Θ_p. Formally evaluate

$$r = \frac{Pr(\Theta^*|D)/J_m(\Theta^*|\Theta^{m-1})}{Pr(\Theta^{m-1}|D)/J_m(\Theta^{m-1}|\Theta^*)}.$$

The inclusion of the jumping distribution in the numerator and denominator is required only if they are asymmetric. In the case of a symmetric distribution, that is $J_m(\theta_a|\theta_b) \equiv J_m(\theta_b|\theta_a)$, simplification to

$$r = \frac{Pr(\theta^*|D)}{Pr(\theta^{m-1}|D)},$$

can be used. The estimate r is used to evaluate the proposal state. Under the Metropolis Hastings routine, if Θ^* is a state which provides a more likely solution (i.e. is closer to some local maximum within the neighbourhood of Θ^{m-1}) a value of $r \geq 1$ is obtained and it is accepted with probability 1, otherwise it is accepted with probability r. It is defined as

$$\theta_p^m = \begin{cases} \theta_p^* & \text{with probability } \min(r, 1) \\ \theta_p^{m-1} & \text{otherwise.} \end{cases}$$

Computationally, for each draw of each parameter in the sampling procedure, the following routine is defined:

1. Draw a single value from the jumping distribution $\Theta_p^* \sim J_m(\Theta^*|\Theta^{m-1})$
2. Calculate r based on Θ_p^* and Θ_p^{m-1}
3. Draw a single value 'r_u' from a uniform distribution with limits (0,1)
   - if $r >$ 'r_u' then set $\Theta_p^m = \Theta_p^*$
   - otherwise set $\Theta_p^m = \Theta_p^{m-1}$

An illustration of the use of the Gibbs sampler is given for the exponential model with a single covariate as described in Section 2.4.2. For notation purposes, set $\Theta = \theta = (\gamma, \beta)$. To start the procedure, starting values are required and set to $\theta^0 = (\gamma^0 = -3.4, \beta^0 = 0.1)$. Figure 2.5 gives a graphical representation of the Gibbs sampler illustrating how the estimates from each parameter converge upon some solution. Also illustrated are the marginal distributions for γ and β.

Once the process has converged, samples are continued to be drawn in order to construct the joint and marginal posterior densities from which inferences can be made. Results in Table 2.4 give the posterior summaries that are obtained.

Parameter	Mean	Median	Std. Dev	95% Cred. Int
γ	-3.52	-3.52	0.03	(-3.57, -3.46)
β	-0.09	-0.09	0.06	(-0.20, 0.02)

Table 2.4: Summaries of posterior distribution obtained via Gibbs sampling

Algorithms for the Metropolis Hasting algorithm can be easily coded in many statistical packages and examples for the piecewise exponential model are included in the Appendix. In many situations however, the BUGS (Bayesian inference Under Gibbs Sampling) suite of packages such as WinBUGS and OpenBUGS are utilised. Issues on convergence and model fit are covered by Gelman et al. [70] and are discussed as necessary within the thesis.

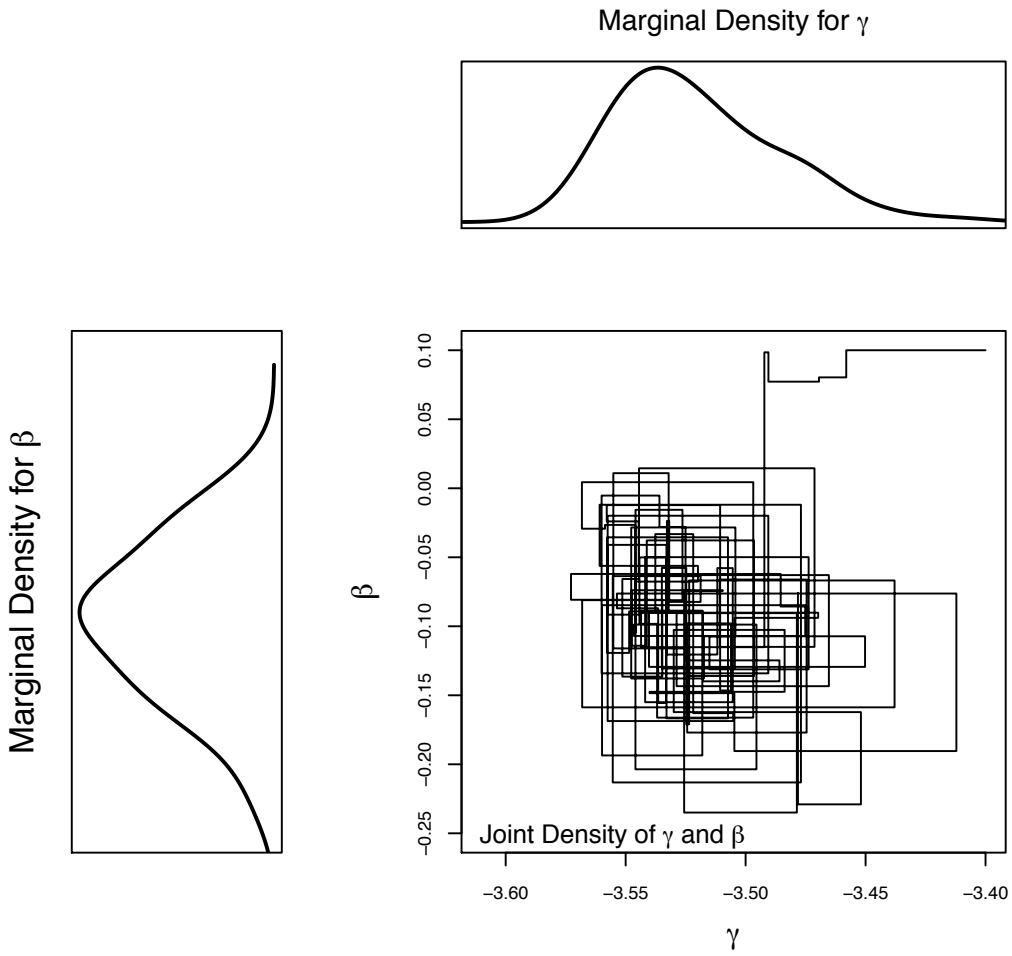

Figure 2.5: An illustration of the Gibbs sampler for the exponential model with a single covariate

2.5 Practical issues for fitting proportional hazards models

Ideally, the decision of which model to fit, and whether or not to fit it in a frequentist or Bayesian framework should depend upon philosophical beliefs and the nature of the data that are being analysed. Often however, the decision will be based on more practical issues.

Typically, Bayesian models will be reliant on some sampling routine. Whilst these are very flexible they can be computationally intensive and difficult to code. In particular, the counting process model will require an extra 'node' to be estimated for each step in the counting process and sampling over a large number of parameters can increase the time to simulate across each iteration. Furthermore, highly correlated nodes can lead to issues such as slow convergence rates and high auto-correlation, both of which require extra samples to be taken.

In a frequentist framework, the counting process method can also cause computational issues due to the empirical likelihood method which effectively requires an extra parameter for each step in the counting process to be estimated. If there is only a single parameter of interest and κ steps in the counting process therefore, estimation of parameters and their standard errors may require the inversion of a matrix with dimensions $(\kappa + 1, \kappa + 1)$.

When large datasets are observed therefore, it may be worthwhile to consider parametric alternatives to the counting process. In particular, a piecewise model will still retain a good deal of flexibility required to accurately estimate a hazard function whilst reducing the number of parameters that are required.

2.6 Discussion

In this chapter a brief overview of the different philosophical differences between frequentist and Bayesian methodology was provided with some indication as to how survival models are fitted in both frameworks.

In particular, an introduction to both the piecewise exponential model and the counting process notation are provided. In both cases connections to the popular Cox model are noted. Following this, some background to the computational issues inherent in Bayesian analysis are discussed and estimation via the Gibbs sampler was introduced. Provided throughout were some basic examples of how these methods can be used to analyse clinical trial data. It should be noted however, that the methods listed here are by no means exhaustive, no comment is made for example on the flexible parametric models proposed by Royston and Parmar (see for example [72, 73]), or other methods, such as accelerated failure time models, which may be applied to survival data.

In the chapters that follow, an exploration of methods for analysing time-to-event data in a clinical trial context is given using the methods introduced in this chapter. In particular, methods of analysis that improve the efficiency of clinical trial data are investigated with the aim of extracting more information from the time consuming and costly process of running a clinical trial.

Chapter 3

Analysis of Survival Data with Unbounded Covariates

3.1 Introduction

In this chapter, some investigation is given to the method by which continuous covariates enter a survival model. In particular, covariates which are prone to outliers or extreme value observations are investigated and a new approach is proposed which gives a scientific rational for an amended form.

The chapter is structured as follows, Section 3.2 gives an overview of robust methods for handling covariates. In Section 3.3 a new parameterisation is proposed with justifications. Section 3.4 provides a simulation study to demonstrate reliable estimation of model parameters. Section 3.5 explores model diagnostics in the form of model residuals and a generalised influence function. The new parameterisation is applied to the ESPAC-3 dataset in Section 3.6 paying particular attention to the effect of the biomarker post operative Cancer Antigen 19.9 (CA19.9). Discussion is given in Section 3.7.

3.2 Robust estimation in proportional hazards modelling

Under standard proportional hazards modelling, covariates enter the model through a link function given by

$$\exp(\boldsymbol{\beta}^T \boldsymbol{z}).$$

This is convenient as $\exp(\beta)$ gives the hazard ratio which gives the multiplicative increase (decrease) in hazard for each unit increase (decrease) in a covariate z. For categorical covariates this is particularly useful as individual hazard rates for each group of observations can be obtained. For some continuous covariates however, assuming this functional form may not be appropriate and can lead to misleading interpretations. This is particularly true for continuous covariates that are prone to extreme values.

Note here that large covariate values are referred to as extreme covariate values as opposed to outliers as outliers may refer to data points that are observed or measured in error and by extension may be considered for removal or given reduced influence in any model estimation procedure.

It has been noted by Viviania et al. [74] that the presence of only a single extreme value observation can be enough to violate any model assumptions of proportionality. By extension, any covariate that violates this assumption is difficult to interpret in a meaningful fashion through a hazard ratio. A common method to curb the influence of extreme value observations is to apply some transformation, $g(z)$ and model this instead. Common suggestions are $g(z) = log(z)$ and $g(z) = z^{-1}$. Often this is sufficient, but when it is not the user is confronted with the difficulty of either accepting a model with evidence of non-proportionality, searching for further transformations until evidence of non-proportionality disappears, or applying more complex statistical methods such as a fractional polynomial approach. Each method can have adverse effects on both model fit and interpretation, especially if there is any interaction/confounding between the covariate with extreme value observations and other covariates of interest.

Some previous methods to account for extreme value observations have concentrated on amendments to the likelihood formulation. A good overview is given by Farcomeni and Ventura [75] with two approaches in particular given specific attention; an approach based on a weighted likelihood formulation for the Cox model, most notably proposed by Bednarski and Sensiani [76] and Minder et al. [77] and secondly the method of 'trimmed' likelihoods given by Viviani and Farcomeni [74].

For weighted Cox regression with N observations, a log likelihood is proposed in the form:

$$l(\beta) = \log(L(\beta)) = \sum_{i=1}^{N} A(t_i, z_i)\Big[z_i - \frac{\sum_{j\in R} A(t_i, z_i) z_j \exp(\beta^T z_j)}{\sum_{j\in R} A(t_i, z_i) \exp(\beta^T z_j)}\Big].$$

Here $A(t_i, z_i)$ is a smooth non-negative function which has a limit of zero which is obtained for either large values of t or $\beta^T z$. This method then down-weights or completely ignores patients who either have large covariate values or who live longer than may be expected.

A second approach introduced by Viviani [74] and explored by the same author [78] is a trimmed likelihood. This is based on the idea that any likelihood formulation is trimmed by excluding the observations which give the smallest contributions to the likelihood. Specifically, it is considered that the data consist of $[n(1-\varrho)]$ 'clean' observations and $[n\varrho]$ contaminated observations. The hazard rate for each observation is given by

$$\lambda(t, z_i) = \begin{cases} \lambda_0(t)\exp(\beta^T z_i) & \text{if } i \in I^* \\ \lambda_i(t) & \text{otherwise.} \end{cases}$$

Here I^* is an indicator function denoting whether patients belong to the 'clean' or

'contaminated' dataset. Under a trimmed likelihood, $H(\varrho)$ is defined as a class of all subsets of cardinality $[n(1-\varrho)]$ from the vector of integers $(1,...,n)$. The likelihood is then given by

$$L_{TRIM}(\beta) = max_{I\in(H(\varrho))} \prod_{i\in I} \left(\frac{\exp(\beta^T z_i)}{\sum_{j\in R} \exp(\beta^T z_i)} \right).$$

Whilst either procedure may produce more robust hazard ratios they do not solve the problem of non-proportionality. More troublesome may be that the model is explicitly treating some data as less valuable than others, possibly without any predilection for doing so. In certain situations therefore, these methods can be criticised as trying to amend the data to fit a model whereby it should be the goal of a statistician to produce a model to fit the data.

3.3 New parameterisations for proportional hazards modelling

An alternative approach is proposed whereby the function by which covariate enter a model is altered to explicitly allow for extreme value observations. Methods of altering the function by which covariates enter a model are not new. As an example, take the fractional polynomial methods as proposed by Royston and Altman [79] which explore a range of possible transformations of z to search for the best model fit. This approach may still result in unreliable estimates at the extremities of the observed covariates values and cause difficulties with extrapolation. Furthermore, the final functional form that is obtained, whilst being a good fit of the data to the model, may not be cohesive with scientific rational between the underlying relationship between a covariate and the response.

The model proposed here allows for initial exponential growth in the relationship between a covariate and survival function but also with the inclusion of an upper asymptote. This is chosen to allow for the scientific rational whereby there may be an important difference between a 'small' and a 'big' value of a covariate in explaining patient survival but little or no difference between a 'big' and 'very big' value. Reasoning such as this often leads to the desire to dichotomise a covariate, an approach which comes at a cost as shown by Altman [80] and Royston et al. [81].

Define the parameters required to estimate the effect of a covariate as $(\boldsymbol{\theta})$. For the standard parameterisation $(\boldsymbol{\theta_S}) = \beta$ and $f(\boldsymbol{\theta_S}) = \exp(\beta^T z)$ are obtained. The proposed alternative function has two parameters $(\boldsymbol{\theta_R}) = \{\varphi, \beta\}$ and is given by

$$f(\boldsymbol{\theta_R}) = \frac{1+\varphi\exp(\beta z)}{\varphi+\exp(\beta z)}. \tag{3.1}$$

The relationship between an observed covariate and the effect on some baseline

hazard function is shown in Figure 3.1. Here models can be shown to have a similar effect for smaller values for z but diverge wildly as z increases. The role of β in (3.1) is to control the initial rate of growth with φ providing the asymptote. Interpretation as β as a hazard ratio is no longer valid, instead it is defined as the maximum rate of unit increase. It is the unit rate of increase that is observed as $z \to 0$. Further, φ is interpreted as a maximum hazard rate.

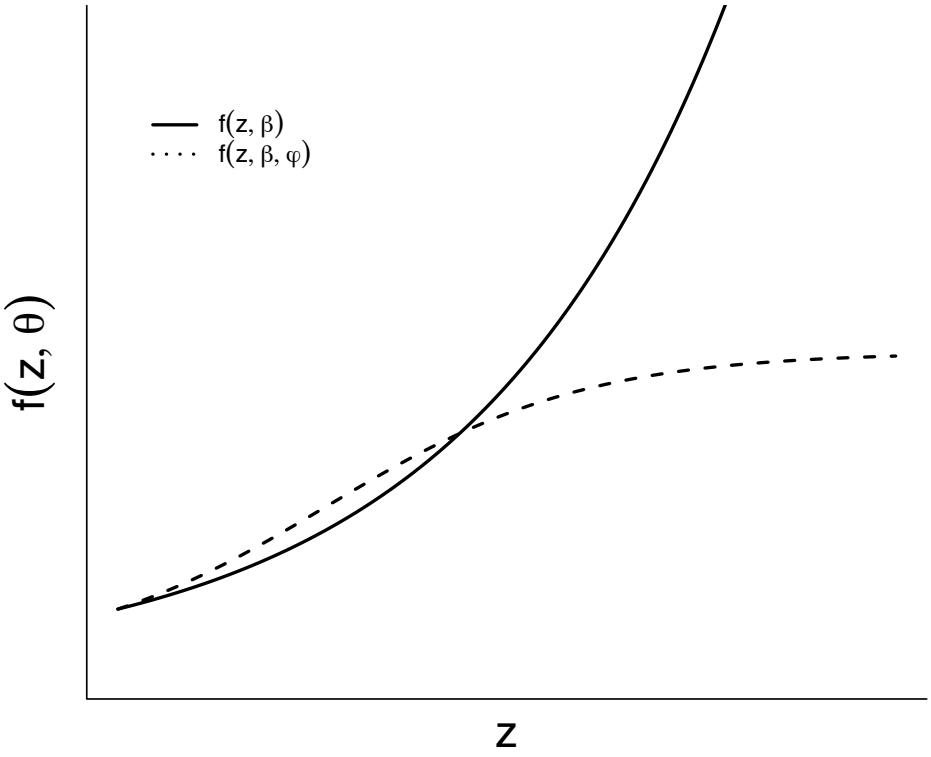

Figure 3.1: Figure to show the functional representation of the standard and new parameterisation for the linear prediction.

Justification for $f(\boldsymbol{\theta_R})$ is given by the fact that it is possible to bound a function $f(\theta)$, above and below, by using a family of transformations with the the form

$$\frac{\delta + \varphi \exp(\beta z)}{\vartheta + \exp(\beta z)}. \tag{3.2}$$

This is an adaption of the logistic function and has asymptotes φ and δ/ϑ. Restrictions are placed on the parameters: $\delta \geq 0$, $\varphi \geq 0$ and $\vartheta \geq 0$ in order for $f(\boldsymbol{\theta_R})$ to be non-negative. The first derivative of $f(\boldsymbol{\theta}, z)$ is $(\vartheta\varphi - \delta)\beta \exp\{\beta z\}/[\vartheta + \exp\{\beta z\}]^2$ and in order for a positive β to have positive slope for $f(.)$, and correspondingly, a negative β to have a negative slope for $f(.)$, then $\delta < \vartheta\varphi$. Also, $f(.)$ is monotonically increasing in z which is usually a useful property in practice. A particular fractional polynomial might not possess this property. Model 3.2 has the property that the hazard function is symmetric regarding the baseline hazard, i.e. $f(\boldsymbol{\theta}, z)$ and $f'(\boldsymbol{\theta}', z)$

have the same functional form for the two models $h(\boldsymbol{\theta}, t, z) = h(\boldsymbol{\theta}, t, z)_0 f(\boldsymbol{\theta}, z)$, and $h(\boldsymbol{\theta}, t, z)_0 = h(\boldsymbol{\theta}, t, z)/f(\boldsymbol{\theta}, z)$.

A desirable property for $f(\boldsymbol{\theta}, z)$ is that when $\beta = 0$, implying that the covariate has no effect on survival, then$f(\boldsymbol{\theta}, z)$ should have the value unity. This implies $\vartheta = \delta + \varphi - 1$ and leads to

$$f(\boldsymbol{\theta}, z) = \frac{\delta + \varphi \exp\{\beta z\}}{\delta + \varphi - 1 + \exp\{\beta z\}} \tag{3.3}$$

The asymptotes for model 3.3 are φ and $\delta/(\delta + \varphi - 1)$ and it still retains baseline hazard symmetry. For φ>1, positive β will give a positive slope for $f(.)$ and negative β a negative slope. The value of z which has no effect on the baseline hazard is $z = 0$. If this should be a different value then the variable z should be adjusted accordingly with a linear transformation. Note if a parameter is entered into the model, replacing z by $z - \Upsilon$ with estimation of Υ, then it can be shown that, by rearranging parameters, model 3.3 reverts back to model 3.2.

The slopes of the logistic function at $+z$ and $-z$ are identical. For model 3.3 to have this property, then $\vartheta = 2 - \varphi$ and hence

$$f(\boldsymbol{\theta}, z) = \frac{2 - \varphi + \varphi \exp\{\beta z\}}{1 + \exp\{\beta z\}} \tag{3.4}$$

where the asymptotes are φ and $2 - \varphi$. This model loses its baseline hazard symmetry.

For model 3.3 to have reciprocal asymptotes, φ and $1/\varphi$, then $\vartheta = 1$ giving

$$f(\boldsymbol{\theta}, z) = \frac{1 + \varphi \exp\{\beta z\}}{\varphi + \exp\{\beta z\}}. \tag{3.5}$$

This model retains baseline hazard symmetry.

Lastly the standard Cox model is obtained by letting $\varphi \to \infty$ in 3.5, or by letting $\varphi = 0$ which will negate the β coefficient. If $\varphi = 1$ then $f(\boldsymbol{\theta}, z) = 1$ with z having no effect on the hazard function. Throughout the rest of this chapter, concentration in on model 3.5 although other models could be used and fitted to data in a similar manner.

3.4 Simulation study

A simulation study is carried out to demonstrate that parameters in the robust model can be accurately estimated.

Data are simulated for both a covariate and a two-level factor to replicate the scenario of a single covariate with extreme value observations measured as part of a two arm clinical trial. Denote the log hazard ratio for the treatment factor as β_{trt} and the parameters for the continuous covariate as φ, β_{cov}. Parameters for the simulation are set arbitrarily as:

$$\beta_{trt} = 0.15, \quad \beta_{cov} = 0.05, \quad \varphi = 5.$$

The treatment factor, z_{trt}, is simulated to ensure that there are 50% of patients assigned to each arm. The covariate is simulated to satisfy $log(z_{cov}) \sim N(3.5, 1.5)$. Survival times are simulated from an exponential distribution with a hazard parameter set as $\lambda = 0.5$. Of the simulated data, 5% of survival times randomly selected as censored to replicate the data that are observed in practice. This approach makes is no assumption over the form of the censoring distribution, but any bias due to informative censoring will be small due to the small amount of censoring that is induced. Data sets of size $100, 250, 500$ and 1000 are considered. For each set of parameters, 1000 datasets are simulated resulting in 4000 datasets in total.

To each sampled dataset, a standard model with linear predictor defined by $\exp(\beta_{trt} z_{trt} + \beta_{cov} z_{cov})$ is fit along with the new parameterisation given by

$$\exp(\beta_{trt} z_{trt}) \frac{1 + \varphi \exp(\beta_{cov} z_{cov})}{\varphi + \exp(\beta_{cov} z_{cov})}.$$

using the partial likelihood given by (2.5). Each model is assessed in terms of parameter bias, accuracy, coverage and average confidence interval length (ACIL) following the recommendation of Burton et al. [82]. The results are given in Table 3.1. Evaluations of model performance for β_{trt} are considered across both models. Evaluations of β_{cov} and φ are only provided for the new parameterisation.

N	Parm.	Standard Model					New Model				
		Est. (se.)	Bias	Acc.	Cov.	ACIL	Est. (se.)	Bias ($\times 10^{-3}$)	Acc.	Cov.	ACIL
100	β_{trt}	0.13 (0.20)	0.02	0.04	0.95	0.81	0.15 (0.20)	4.1	0.04	0.95	0.82
	β_{cov}	0.003 (0.002)					0.06 (0.03)	6.3	0.001	0.92	0.09
	φ						6.35 (3.68)	1356	15.38	0.97	13.20
250	β_{trt}	0.130 (0.13)	0.02	0.02	0.95	0.51	0.15 (0.14)	1.9	0.02	0.94	0.51
	β_{cov}	0.002 (0.001)					0.05 (0.02)	2.3	5×10^{-5}	0.93	0.06
	φ						5.31 (1.12)	310	1.35	0.97	4.33
500	β_{trt}	0.14 (0.09)	0.02	0.01	0.95	0.36	0.15 (0.09)	4.4	0.01	0.94	0.36
	β_{cov}	0.002 (0.003)					0.05 (0.01)	1.0	4×10^{-5}	0.93	0.04
	φ						5.20 (0.73)	203	0.57	0.97	2.91
1000	β_{trt}	0.13 (0.06)	0.02	0.004	0.95	0.25	0.15 (0.06)	0.04	0.004	0.96	0.25
	β_{cov}	0.001 (0.001)					0.05 (7e-3)	0.6	5×10^{-5}	0.91	0.03
	φ						5.07 (0.49)	70	0.25	0.96	1.98

Table 3.1: Results of the simulation study to compare performance of the standard model and the new parameterisation

Considering initially the new parameterisation, it is observed that for datasets of 250 or larger, acceptable levels of bias, accuracy, coverage and average confidence interval length are obtained for each parameter. For datasets of size 100, reasonable estimates of β_{trt} and β_{cov} are obtained but there is a reasonably large bias obtained for φ. It should be observed that this parameter, estimating the upper asymptote, is going to be dependent upon the number of extreme value observations that are observed. For smaller datasets, the number of extreme covariate values will be small and as a consequence estimates of φ will be less reliable. It is also worth noting that the

standard errors associated with φ are relatively large compared to the standard errors associated with other parameters, this is again due to the variability induced by this parameter being driven by data in the extremes of the distribution for z_{cov}.

Observing the estimates of coverage, by chance it is reasonable to expect the 'true' parameter values to be included in the confidence interval 95% of the time. There is some small but consistent divergence with φ being constantly larger than 0.95 which may be explained in part due to the relatively large ACIL. Considering β_{cov} the coverage estimate is consistently lower than 0.95 however. This may be of little concern due to the small bias observed and may be a result of correlation between β_{cov} and φ.

Of further interest is the estimates of β_{trt} for both the standard and new parameterisations. Estimates of bias for the new parameterisation are consistently small. For the standard parameterisation however, there is a consistent negative bias despite the data being simulated without any correlation structure between the covariates. This illustrates that incorrectly specifying the relationship between a continuous covariate and the underlying hazard function can have adverse effects other covariates in the model and may result in biased estimates. This point is reinforced by Figure 3.2. This shows kernel density plots of all estimates of β_{trt} from the datasets with a sample size of 1000 and demonstrates a consistent shift for the standard model. This shift between distributions demonstrates a negative bias when models are incorrectly specified.

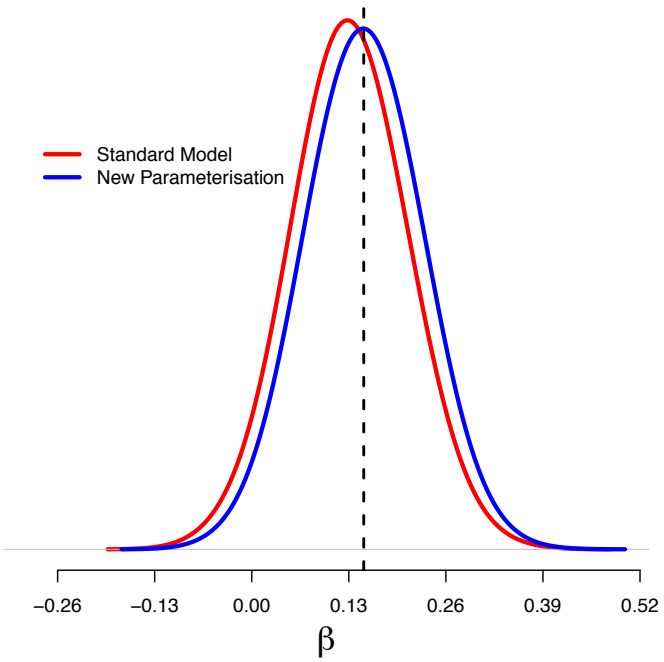

Figure 3.2: Figure to show the distribution of estimated β_{trt} for standard and new parameterisations.

3.5 Model diagnostics

In this section, the new parameterisation is explored with respect to model diagnostics. In particular, a form of model residuals based on the those first proposed by Schoenfeld [83] are explored as well as an analytical form for an influence function following the method Reid and Crapeau [84].

3.5.1 Model residuals

Taking Cox's partial likelihood, Schoenfeld's methods for obtaining residuals is based on noting that the score function of the likelihood for a single covariate, is obtained from

$$\sum_{i=1}^{N} \big[\log\{f(z_i,\theta)\}\big]' - \big[\log\{\sum_{j\in R} f(z_j,\theta)\}\big]' = 0,$$

where $'$ represents a first derivative with respect to θ. Residuals are obtained by observing

$$\sum_{i=1}^{N} \big[\big[\log\{f(z_i,\hat{\theta})\}\big]\big]' - E\big[\log\{f(z_j,\hat{\theta})\}'|R\big] = 0.$$

Here R represent the group of patients at risk at time t_i. From this, it is seen that

$$E\big[\log\{f(z_j,\theta)\}'|R\big] = \big[\log\{\sum_{j\in R} f(z_j,\theta)\}\big]'.$$

For any given individual covariate z_i at time t_i therefore, an expected value can be derived from all other covariate values still at risk at time t_i. Note that this condition only holds so long as there is no relationship between model parameters and time. Considering the model defined by (3.1), a score function satisfies

$$\sum_{i=1}^{N} \nu_i \Big[\log\{\varphi\exp(\beta z)\} - \log\{\varphi - 1 + \exp(\beta z)\} - \log\Big\{\sum_{j\in R} \frac{\varphi\exp(\beta z)}{\varphi - 1 + \exp(\beta z)}\Big\}\Big] = 0.$$

Analytical forms for the residuals are then obtained by differentiating with respect to β and φ and equating to zero. For β the following is obtained:

$$\sum_{i=1}^{N} \nu_i \Big[\frac{(\varphi-1)z}{\varphi - 1 + \exp(\beta z)} - \frac{\sum_{j\in R} \frac{\varphi z(\varphi-1)\exp(\beta z)}{[\varphi-1+\exp(\beta z)]^2}}{\sum_{j\in R} \frac{\varphi\exp(\beta z)}{\varphi-1+\exp(\beta z)}}\Big] = 0, \tag{3.6}$$

and respectively for φ:

$$\sum_{i=1}^{N} \nu_i \Big[\frac{\exp(\beta z) - 1}{\varphi(\varphi - 1 + \exp(\beta z))} - \frac{\sum_{j\in R} \frac{\exp(\beta z)(\exp(\beta z)-1)}{[\varphi-1+\exp(\beta z)]^2}}{\sum_{j\in R} \frac{\varphi\exp(\beta z)}{\varphi-1+\exp(\beta z)}}\Big] = 0. \tag{3.7}$$

Model residuals obtained using Shcoenfeld's method have been used as a means of assessing the assumption of proportionality. A widely used example is that by Grambsh and Thernau [85] who develop a test based on the correlation between residuals and the rank of time. More details on the uses of residuals to assess proportionality are provided in Chapter 4.

3.5.2 Influence function

Here an analytical form of an influence function is presented. The algebra that follows is extremely involved and complex and the interested reader is advised to follow the paper by Reid and Crepeau [84] upon which the methodology presented here is based.

Define the score function from a partial likelihood given in a general case as

$$\sum_{i=1}^{N} \nu_i \Big[\boldsymbol{\gamma_1} - \frac{\sum_{j \in R} \boldsymbol{\gamma_2}}{\sum_{j \in R} \boldsymbol{\gamma_3}} \Big]. \tag{3.8}$$

It can be seen that both (3.6) and (3.7) are special cases of (3.8) as indeed is the score function obtained for the standard Cox model. Specifically $\boldsymbol{\gamma_1}$, $\boldsymbol{\gamma_2}$ and $\boldsymbol{\gamma_3}$ are a set of functions whose exact form will depend on the likelihood formed. For the standard parameterisation with a single covariate $f_S(\theta)$, set

$$\boldsymbol{\gamma_1} = \gamma_1^1 = z,$$

$$\boldsymbol{\gamma_2} = \gamma_2^1 = z \exp(\beta z),$$

and

$$\boldsymbol{\gamma_3} = \gamma_3 = \exp(\beta z).$$

For the new parameterisation proposed, definitions are given by $\boldsymbol{\gamma_1} = \{\gamma_1^1, \gamma_1^2\}$ with

$$\gamma_1^1 = \frac{(\varphi - 1)z}{(\varphi - 1 + \exp(\beta z))}$$

and,

$$\gamma_1^2 = \frac{\exp(\beta z) - 1}{\varphi(\varphi - 1 + \exp(\beta z))}.$$

Next $\boldsymbol{\gamma_2} = \{\gamma_1^1, \gamma_1^2\}$ with

$$\gamma_2^1 = \sum_{j \in R} \frac{\varphi z(\varphi - 1)\exp(\beta z)}{[\varphi - 1 + \exp(\beta z)]^2}$$

and

$$\gamma_2^1 = \sum_{j \in R} \frac{\exp(\beta z)\{\exp(\beta z) - 1\}}{[\varphi - 1 + \exp(\beta z)]^2}$$

Lastly

$$\gamma_3 = \frac{\varphi \exp(\beta z)}{\varphi - 1 + \exp(\beta z)}.$$

In this general form, a influence function is evaluated by noting that the data observed are contaminated by some small amount of noise. Following the notation of Reid and Crepeau [84], define the empirical cumulative distribution function for time T, a single covariate Z and a censoring indicator Δ by $H(t, z, \delta)$. Further define the marginal function for T and Z as $H(t, z)$. The empirical function for the Cox partial likelihood formulation places a point mass of magnitude n^{-1} on each uncensored observation. A generalised score function based on the empirical distribution is then obtained by

$$\int_N \delta_i \Big\{ \boldsymbol{\gamma_1}(\boldsymbol{\theta}) - \Big[\frac{\int \boldsymbol{\gamma_2}(\boldsymbol{\theta}) I(\tilde{t} \geq t) \delta H_n(\tilde{t}, \tilde{z})}{\int \boldsymbol{\gamma_3}(\boldsymbol{\theta}) I(\tilde{t} \geq t) \delta H_n(\tilde{t}, \tilde{z})} \Big] \Big\} \delta H(t, z, \delta) = 0.$$

Here $I(.)$ is the indicator function. Observing that the infinite sample defines the parameters as a set of functionals of the empirical distribution, $\boldsymbol{\theta} = \boldsymbol{\theta}(H)$, the score function is defined as:

$$\int_N \delta_i \Big\{ \boldsymbol{\gamma_1}(\boldsymbol{\theta}(H)) - \Big[\frac{\int_{\tilde{t} \geq t} \boldsymbol{\gamma_2}(\boldsymbol{\theta}(H)) \delta H_n(\tilde{t}, \tilde{z})}{\int_{\tilde{t} \geq t} \boldsymbol{\gamma_3}(\boldsymbol{\theta}(H)) \delta H_n(\tilde{t}, \tilde{z})} \Big] \Big\} \delta H(t, z, \delta) = 0$$

For the purposes of obtaining an influence function, it is assumed that data are generated from the empirical distribution function $H(.)$ with some small contamination $\epsilon(G(.))$. A form of influence measure for parameters $\boldsymbol{\theta}$ is evaluated from $\epsilon^{-1}(\boldsymbol{\theta}(H + \epsilon G) - \boldsymbol{\theta}(H)) = 0$ as $\epsilon \to 1$. Defining for the $k = 1, 2, 3$ sets of functions, $\boldsymbol{\gamma_k}(\boldsymbol{\theta} + \epsilon \dot{\boldsymbol{\theta}})$,

$$\int_N \delta_i \Big\{ \boldsymbol{\gamma_1}(\boldsymbol{\theta} + \epsilon \dot{\boldsymbol{\theta}}) - \frac{\int_{\tilde{t} \geq t} \boldsymbol{\gamma_2}(\boldsymbol{\theta} + \epsilon \dot{\boldsymbol{\theta}}) \delta H(\tilde{t}, \tilde{z}) + \epsilon \delta G(\tilde{t}, \tilde{z})}{\int_{\tilde{t} \geq t} \boldsymbol{\gamma_3}(\boldsymbol{\theta} + \epsilon \dot{\boldsymbol{\theta}}) \delta H(\tilde{t}, \tilde{z}) + \epsilon \delta G(\tilde{t}, \tilde{z})} \Big\} \delta H(t, z, \nu) + \epsilon \delta G(t, z, \delta) = 0.$$

A solution is obtained by first solving the inner integrands and then expanding into the outer integrands. Begin by noting that the Taylor series expansion of $\boldsymbol{\gamma_k}(\boldsymbol{\theta} + \epsilon \dot{\boldsymbol{\theta}})$ to the second order is given by

$$\boldsymbol{\gamma_k}(\boldsymbol{\theta} + \epsilon \dot{\boldsymbol{\theta}}) = \boldsymbol{\gamma_k}(\boldsymbol{\theta}) + \nabla \boldsymbol{\gamma_k}(\boldsymbol{\theta})^T \epsilon \dot{\boldsymbol{\theta}} + \frac{1}{2} (\epsilon \dot{\boldsymbol{\theta}})^T \nabla^2 \boldsymbol{\gamma_k}(\boldsymbol{\theta}) (\epsilon \dot{\boldsymbol{\theta}}).$$

Here $\nabla \boldsymbol{\gamma_k}(\boldsymbol{\theta})$ is a vector of first derivatives with respect to $\boldsymbol{\theta}$ and $\nabla^2 \boldsymbol{\gamma_k}(\boldsymbol{\theta})$ is the matrix of second derivatives. Note that

$$\frac{1}{2} (\epsilon \dot{\boldsymbol{\theta}})^T \nabla^2 \boldsymbol{\gamma_k}(\boldsymbol{\theta}) (\epsilon \dot{\boldsymbol{\theta}}) \simeq 0,$$

and evaluate the inner integrand for the term containing $\boldsymbol{\gamma_2(\theta)}$. Re-write as:

$$\int_{\tilde{t}\geq t}\boldsymbol{\gamma_2(\theta)}\delta\tilde{H}+\epsilon\int_{\tilde{t}\geq t}\nabla\boldsymbol{\gamma_2(\theta)}^T\dot{\boldsymbol{\theta}}\delta\tilde{H}+\epsilon\int_{\tilde{t}\geq t}\boldsymbol{\gamma_2(\theta)}\delta\tilde{G}+\epsilon^2\int_{\tilde{t}\geq t}\nabla\boldsymbol{\gamma_k(\theta)}^T\dot{\boldsymbol{\theta}}\delta\tilde{G} \quad (3.9)$$

Note that $\delta\tilde{H}$ is used to abbreviate $\delta H(\tilde{t},\tilde{z})$ and likewise for $\delta\tilde{G}$. For further brevity, define $\int\boldsymbol{\gamma_k(\theta)}\delta\tilde{H}=\boldsymbol{\Gamma_k}$ and note that the last term in (3.10) $\simeq 0$ due to ϵ^2. The inner intergrand containing $\boldsymbol{\gamma_2(\theta)}$ is re-written as

$$\int_{\tilde{t}\geq t}\boldsymbol{\gamma_2(\theta)}\delta\tilde{H}+\epsilon\delta\tilde{G}=\tilde{\boldsymbol{\Gamma_2}}+\epsilon\tilde{\boldsymbol{\Gamma_2}}'^T\dot{\boldsymbol{\theta}}-\epsilon\boldsymbol{\gamma(\theta)_2}$$

where $\tilde{\boldsymbol{\Gamma_2}}'$ is the integral of the vector of first derivatives, $\int_{\tilde{t}\geq t}\nabla\boldsymbol{\gamma_k(\theta)}\delta\tilde{H}$. Using the same process, re-write the integrand including $\boldsymbol{\gamma_3(\theta)}$ as

$$\frac{1}{\int_{\tilde{t}\geq t}\boldsymbol{\gamma_3(\theta)}\delta\tilde{H}+\epsilon\delta\tilde{G}}=[\tilde{\boldsymbol{\Gamma_3}}]^{-2}[\tilde{\boldsymbol{\Gamma_3}}-\epsilon\tilde{\boldsymbol{\Gamma_3}}'^T\dot{\boldsymbol{\theta}}-\epsilon\boldsymbol{\gamma(\theta)_3}]$$

In this form, evaluate both inner integrands as

$$\frac{\int_{\tilde{t}\geq t}\boldsymbol{\gamma_2(\theta)}\delta\tilde{H}+\epsilon\delta\tilde{G}}{\int_{\tilde{t}\geq t}\boldsymbol{\gamma_3(\theta)}\delta\tilde{H}+\epsilon\delta\tilde{G}}=\left[\frac{\tilde{\boldsymbol{\Gamma_2}}}{\tilde{\boldsymbol{\Gamma_3}}}\right]+\epsilon\left[\frac{\tilde{\boldsymbol{\Gamma_2}}'\tilde{\boldsymbol{\Gamma_3}}-\tilde{\boldsymbol{\Gamma_2}}\tilde{\boldsymbol{\Gamma_3}}'}{(\tilde{\boldsymbol{\Gamma_3}})^2}\right]^T\dot{\boldsymbol{\theta}}-\epsilon\left[\frac{\tilde{\boldsymbol{\Gamma_2}}\gamma(\theta)_3^{x_0}+\tilde{\boldsymbol{\Gamma_3}}\gamma(\theta)_2^{x_0}}{(\tilde{\boldsymbol{\Gamma_3}})^2}\right]$$

Applying a Taylor series expansion to all terms in the outer integrand, a full form for $\boldsymbol{\theta}(H+\epsilon G)$ is expressed as:

$$\int_N\nu_i\Bigg\{\boldsymbol{\gamma_1(\theta)}+\epsilon\nabla\boldsymbol{\gamma_1(\theta)}^T\dot{\boldsymbol{\theta}}+\left[\frac{\tilde{\boldsymbol{\Gamma_2}}}{\tilde{\boldsymbol{\Gamma_3}}}\right]+\epsilon\left[\frac{\tilde{\boldsymbol{\Gamma_2}}'\tilde{\boldsymbol{\Gamma_3}}-\tilde{\boldsymbol{\Gamma_2}}\tilde{\boldsymbol{\Gamma_3}}'}{(\tilde{\boldsymbol{\Gamma_3}})^2}\right]^T\dot{\boldsymbol{\theta}}-\epsilon\left[\frac{\tilde{\boldsymbol{\Gamma_2}}\gamma(\theta)_3^{x_0}+\tilde{\boldsymbol{\Gamma_3}}\gamma(\theta)_2^{x_0}}{(\tilde{\boldsymbol{\Gamma_3}})^2}\right]\Bigg\}\delta H+\epsilon\delta G=0. \quad (3.10)$$

Expanding out the integrands, the following is obtained:

$$\int_N\nu_i\left\{\boldsymbol{\gamma_1(\theta)}+\left[\frac{\tilde{\boldsymbol{\Gamma_2}}}{\tilde{\boldsymbol{\Gamma_3}}}\right]\right\}\delta H+\epsilon\left[\int_N\nu_i\left\{\nabla\boldsymbol{\gamma_1(\theta)}+\left[\frac{\tilde{\boldsymbol{\Gamma_2}}'\tilde{\boldsymbol{\Gamma_3}}-\tilde{\boldsymbol{\Gamma_2}}\tilde{\boldsymbol{\Gamma_3}}'}{(\tilde{\boldsymbol{\Gamma_3}})^2}\right]\right\}\delta H\right]^T\dot{\boldsymbol{\theta}}+$$
$$\epsilon\int_N\left[\frac{\tilde{\boldsymbol{\Gamma_2}}\gamma(\theta)_3^{x_0}+\tilde{\boldsymbol{\Gamma_3}}\gamma(\theta)_2^{x_0}}{(\tilde{\boldsymbol{\Gamma_3}})^2}\right]\delta H+\epsilon\int_N\nu_i\left\{\boldsymbol{\gamma_1(\theta)}+\left[\frac{\tilde{\boldsymbol{\Gamma_2}}}{\tilde{\boldsymbol{\Gamma_3}}}\right]\right\}\delta G=0 \quad (3.11)$$

Note that to obtain influence measures, we evaluate $\epsilon^{-1}(\boldsymbol{\theta}(H+\epsilon G)-\boldsymbol{\theta}(H))=0$ and it is clear that first term in (3.12) disappears. A solution for the influence measures, $\dot{\boldsymbol{\theta}}$ can be given in the form

$$\epsilon\boldsymbol{Q}\dot{\boldsymbol{\theta}}=\epsilon\boldsymbol{R}.$$

Here Q is a matrix of dimension $\rho \times \rho$ where ρ is the number of parameters in $\boldsymbol{\theta}$. Individual elements of $\boldsymbol{Q}$ are given by

$$q_{rs} = \int_N \nu_i \Big\{ \nabla \gamma_1^r(\boldsymbol{\theta}) + \Big[\frac{\tilde{\Gamma}_{2,s}^{r'} \tilde{\Gamma}_3 - \tilde{\Gamma}_2^r \tilde{\Gamma}_{3,s}'}{(\tilde{\Gamma}_3)^2} \Big] \Big\} \delta H$$

where $\tilde{\Gamma}_{2,s}^{r'}$ is the r^{th} element of $\tilde{\boldsymbol{\Gamma}}_{\mathbf{2}}$ differentiated with respect to the parameter s and similarly for $\tilde{\Gamma}_{3,s}'$. $\boldsymbol{R}$ is a vector of length ρ with elements

$$R_r = \nu_i \Big\{ \gamma_1^r(\boldsymbol{\theta}) + \Big[\frac{\tilde{\Gamma}_2^r}{\tilde{\Gamma}_3} \Big] \Big\} + \int_N \Big[\frac{\tilde{\Gamma}_2^r \gamma(\theta)_3^{x_0} + \tilde{\Gamma}_3 \gamma(\theta)_2^{r,x_0}}{(\tilde{\Gamma}_3)^2} \Big] \delta H.$$

Solving for $\dot{\boldsymbol{\theta}}$ is obtained via

$$\dot{\boldsymbol{\theta}} = \boldsymbol{R}\boldsymbol{Q}^{-1}.$$

3.6 Application to ESPAC 3 data

The methodology introduced here is applied to the ESPAC-3 trial. Of particular interest are the group of patients who had pancreatic ductal adenocarcinomas (PDAC) and for whom a value CA19.9 was recorded post operatively (n=759). It is considered that data for this covariate are missing completely at random and no bias is introduced by considering a complete case analysis. Previously published analyses [32] are followed 'forcing' into the model the terms 'Resection Margin' (Negative vs. Positive) as a stratification factor and 'Treatment Arm' (5FU vs. Gemcitabine) as the key covariate of interest. Also identified as important are 'Lymph Nodes' (Negative vs. Positive), 'Tumour Differentiation' (Poor vs. Moderate vs. Well) and 'Smoking Status' (Never vs. Past vs. Present vs. Missing).

The behaviour of post operative CA19.9 is given by a histogram in Figure 3.3 and is shown to be prone to extreme value observations. This is further illustrated by a median (inter quartile range) of $24(10, 63)$ but with a number of observations greater than $1,000$. Only values up to $2,000$ are displayed, the largest recorded value is $37,000$.

A number of differing modelling approaches are considered. As a reference, analysing post operative CA19.9 directly is considered. Two further models, applying a log transformation and a fractional polynomial approach are also considered. Lastly the new parameterisation is applied.

Models are assessed directly using the log likelihood estimates and Akaike's Information Criterion (AIC) [86]. For each model, residuals are investigated and compared using the methods of Grambsh and Thernau [85]. Influence measures are used to assess the effect of each individual observation.

Note that for the new parameterisation, a definition of the key covariate of interest (CA19.9) is given as z_{cov} and all other covariates as $\boldsymbol{z}$. Further consider the parameter

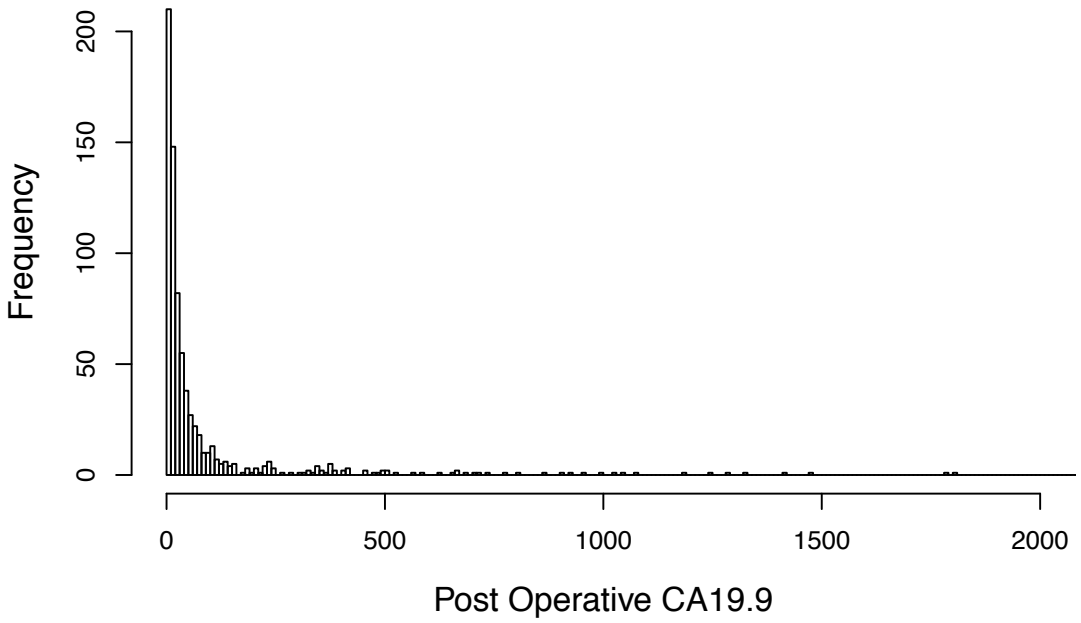

Figure 3.3: Histogram showing the behaviour

estimates associated with $\boldsymbol{z}$ as $\boldsymbol{B}$. Residual estimation for both φ and β in light of other terms is defined as

$$\sum_{i=1}^{N} \nu_i \Bigg[\frac{(\varphi-1) z_{cov}}{\varphi - 1 + \exp(\beta z_{cov})} - \frac{\sum_{j \in R} \frac{\varphi z_{cov}(\varphi-1)\exp(\beta z_{cov} + \boldsymbol{B}^T \boldsymbol{z})}{[\varphi - 1 + \exp(\beta z_{cov})]^2}}{\sum_{j \in R} \frac{\varphi \exp(\beta z_{cov})}{\varphi - 1 + \exp(\beta z_{cov})}} \Bigg] = 0,$$

and

$$\sum_{i=1}^{N} \nu_i \Bigg[\frac{\exp(\beta z_{cov}) - 1}{\varphi(\varphi - 1 + \exp(\beta z_{cov}))} - \frac{\sum_{j \in R} \frac{\exp(\beta z_{cov} + \boldsymbol{B}^T \boldsymbol{z})(\exp(\beta z_{cov}) - 1)}{[\varphi - 1 + \exp(\beta z_{cov})]^2}}{\sum_{j \in R} \frac{\varphi \exp(\beta z_{cov})}{\varphi - 1 + \exp(\beta z_{cov})}} \Bigg] = 0.$$

The results of the reference model, the log transformed model and the new parameterisation are given in Table 3.2. Considering initially the reference model, small values for both $\beta_{CA19.9}$ and its standard error are observed. This is a consequence of the large extreme values observed. Taking as an example a median value for Post Operative CA19.9 as 24, a hazard ratio of 1.02 is obtained showing very modest increases in the baseline hazard. For extreme values of $2,000$, $5,000$ and $37,000$, hazard ratios of 1.17, 1.49 and 18.9 are obtained. A clinician however may find it difficult to believe that a patient with Post Operative CA19.9 value of $37,000$ has an instantaneous risk of death of almost 20 times that of than a similar patient with a zero value.

A log-transformed approach gives an improved model fit as shown by an AIC of 6881.80 (Compared to 6911.94 for the reference model). Considering reference values of post Operative CA19.9 ($24, 2,000, 15,000$ and $37,000$), hazard ratios of $1.95, 4.93, 7.53$ are 9.11 are obtained. Here, extreme hazard ratios are avoided to some extent. Patients

Factor	Level	Reference		Log transf.		New Param.	
		coef	se(coef)	coef	se(coef)	coef	se(coef)
Resec. Margin	Negative						
	Positive	0.21	0.09	0.19	0.09	0.18	0.09
Treatment	5FU						
	Gem.	-0.12	0.08	-0.10	0.08	-0.09	0.08
Lymph Nodes	Negative						
	Positive	0.55	0.10	0.48	0.10	0.46	0.10
Tumor Diff.	Poor						
	Moderate	-0.29	0.10	-0.29	0.10	-0.30	0.10
	Well	-0.64	0.15	-0.62	0.15	-0.63	0.15
Smoke. Status	Never						
	Past	0.09	0.10	0.08	0.10	0.08	0.10
	Present	0.24	0.12	0.26	0.12	0.27	0.12
	Missing	0.22	0.18	0.22	0.18	0.17	0.18
Post Op CA19.9	$\beta_{CA19.9}$	8×10^{-5}	2×10^{-5}	0.21	0.03	0.01	4×10^{-4}
	$\varphi_{CA19.9}$					3.77	0.70
likelihood (AIC)		-3446.97 (6911.94)		-3431.90 (6881.80)		-3420.40 (6860.80)	

Table 3.2: Results for models fitted to the ESPAC-3 data with different strategies for post operative CA19.9

with a median value of Post Operative CA19.9 are almost twice as likely to die at any given time point as those with a zero value. A nine-fold increase in risk is observed for the most extreme patient.

The new parameterisation offers further improvement over the log-transformed model with an AIC of 6860.80. The parameter that represents the upper asymptote, $\varphi_{CA19.9}$ shows that an estimated maximum hazard of 3.77 is obtained. A hazard ratio of 1.19 is obtained for a median Post Operative of CA19.9 value of 24. All other reference values gives a hazard ratio of 3.77 obtained from the upper asymptote. From a clinical perspective, this is the most attractive model with modest small increases in the Post Operative CA19.9 resulting in modest increases in the hazard ratio and larger values curtailed to ensure that unrealistic hazard ratios are not obtained.

Results for the model fitted using a fractional polynomial approach are included in Table 3.3. Judging the model based on AIC alone, this gives the best fit (AIC = 6847.35). The transformations chosen by the fractional polynomial approach are $100/(x+1)$ and $\log(x+1)/100$. These might not be immediately obvious to clinicians and the factor of 100 in each may suggest that changing the units with which Post Operative CA19.9 is measured may improve modelling. Continuing to use the reference points of $24, 2,000, 5,000$ and $15,000$ hazard ratios of $1.73, 3.61, 4.50$ and 5.90 are obtained.

An illustration of the fitted relationship for the effect of Post Operative CA19.9 from each model is given on the log scale in Figure 3.4. The standard and log transformed models both show a linear increase. It is shown that the effect of the log transformation is to alter the rate of increase. Considering the new parameterisation, the 'growth' for

Factor	Level	coef	se(coef)
Resec. Margin	Negative		
	Positive	0.20	0.09
Treatment	5FU		
	Gem.	-0.11	0.08
Lymph Nodes	Negative		
	Positive	0.46	0.10
Tumor Diff.	Poor		
	Moderate	-0.27	0.10
	Well	-0.69	0.15
Smoke. Status	Never		
	Past	0.07	0.10
	Present	0.27	0.12
	Missing	0.21	0.18
Post Op CA19.9	$100/(CA19.9+1)$	0.02	4×10^{-3}
	$\log((CA19.9+1)/100)$	0.32	0.03
	likelihood (AIC)	-3411.86 (6847.35)	

Table 3.3: Results of models fit to the ESPAC-3 data using a fractional polynomial approach to model post operative CA19.9

the function occurs at a Post Operative CA19.9 value of $\exp(4) \approx 50$ with an upper asymptote at around $\exp(6) \approx 400$.

Whilst the fractional polynomial model provides the best fit to the data, Figure 3.4 shows no value of Post Operative CA19.9 will have zero effect on the baseline hazard function within the observed range of data. This can make clinical interpretations troublesome. Furthermore, confounding between post operative CA19.9 ant other covariates may occur as a baseline hazard function is amended to account for this. This can be seen somewhat in the analysis of the ESPAC-3 dataset with some amendments in the point estimates, especially for the Tumour Differentiation covariates.

It is worth noting at this point that there are small adjustments in parameters estimates for covariates other than post operative CA19.9. Whilst any shift is generally small, it is still shown that bias can enter into a model if continuous terms with extreme value observations are not modelled correctly.

3.6.1 Model diagnostics

In this section, the four models fitted are considered in terms of their model diagnostics. Initially, models are considered in terms of their residuals, following this, influence measures are compared.

Residual measures are obtained for the β coefficient only for the standard (reference) and log transformed models. For the fractional polynomial model, two sets of residuals are obtained, one for each transformation applied. For the new parameterisation, residuals are provided for each parameter, φ and β.

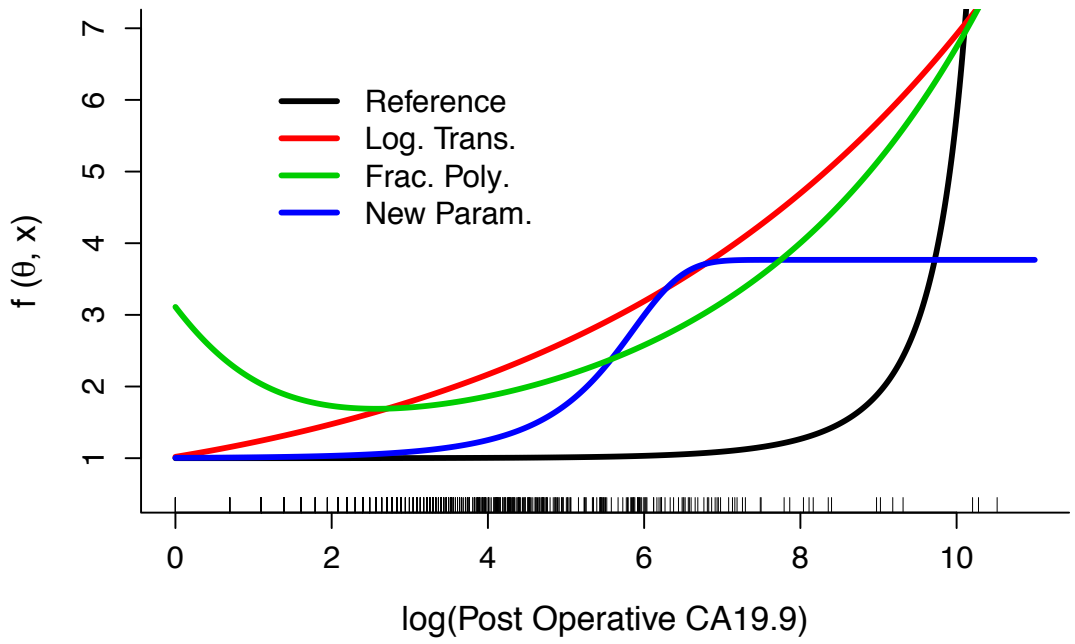

Figure 3.4: Figure showing the effect that various parameterisations have on the baseline hazard function

All residuals are plotted against time. Due to the differing relationships that are observed, direct evaluations against time, such as plotting regression models and assessing the slope parameter to assess proportionality (as proposed by Grambsch and Therneau [85]) are not provided and all assessments are visual. Further justification for this is also given by the lack of standard relationship that would meet the assumptions of simple linear regression, most notably that of all observations being drawn from identically drawn distributions which is shown by an apparent mean-variance relationship between the residuals and time for log transformed components.

Considering initially the reference model, there are some obvious extreme, positive outliers. These all occur before 40 months and distort the figure to make further evaluations difficult. There is also evidence of variability in the residuals that reduces in time. Considering the log transformed residuals, it is observed that the extreme outliers have been accounted for. However, there is still a relationship which shows that variability in the residuals decreases with time.

Residuals from the fractional polynomial models are considered for each transformation individually. For the transformation given by $((z+1)/100)^{-1}$ there are again some outliers, though none as extreme as for the reference model. For the transformation given by $\log((z+1)/100)$, a similar relationship to the log transformed model is observed whereby outliers appear to have been accounted for although the relationship between the residuals and time still persists.

For the new parameterisation, both φ and β are considered collectively. Here, the

residuals show the effect of the upper asymptote as there are upper and lower bounds that appear on both plots. These plots show that this is the functional form that is most successful in controlling the relationship between time and residual variability, especially for the residuals associated with β. The effect of the residuals for φ is less pronounced - however it should be noted here that φ is the term associated with the upper asymptote and that small values of Post Operative CA19.9 will have little effect upon the estimation of this parameter.

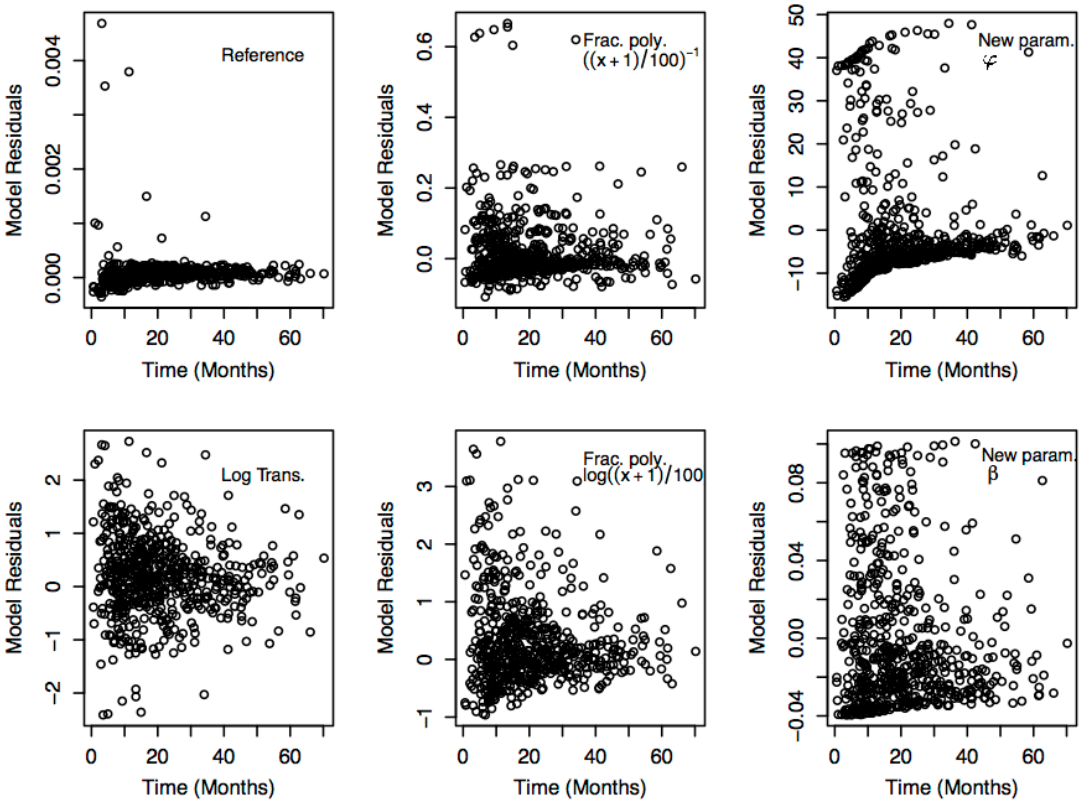

Figure 3.5: Residual measures for models fit to the ESPAC-3 dataset

Turning attention to the influence measures, these are obtained for both censored and uncensored observations. In the Figure 3.6 presented, crosses mark the observed events and circles the censored times. Here the scale range on the y-axis is not provided as it is the relative difference between influence measures that are of interest. The scale of the y-axis differs between parameters due to the parameter estimate and comparing between models is of little use. Influence measures are plotted against Post Operative CA19.9, on the log scale, to observe any functional relationships. Initially considering the reference model, there is no obvious relationship with Post Operative CA19.9 with no observation having particularly large effects on the parameter estimation. Considering the log-transformed model, it is observed that there is some central point of Post Operative CA19.9 just short of 4. Either side of this point, there is a

general divergence with both small and large values of Post Operative CA199 having relatively large effects upon parameter estimation.

For the fractional polynomial model, the influence measures associated with the $((z+1)/100)^{-1}$ term show that very small values of Post Operative CA19.9 can have a disproportionally large influence upon parameter estimation. There are also some large positive influence measures associated with Post Operative CA19.9 values greater than 6. For the term given by $\log((z+1)/100)$, there is a similar relationship to the standard log transformed model although here the divergence from some central point is less pronounced. Large values of Post Operative CA19.9 are again associated with typically large, positive influence measures.

Considering the new parameterisation, there is neither the divergence away from some central point, nor any large influence measures associated with small values of Post Operative CA19.9. The φ parameter is associated with some large positive influence measures. This is to be expected as this is the parameter associated with setting the upper asymptote. This parameter is driven by the amount of 'large' data that are observed and any single one can have a relatively large effect on the parameter estimate. It should be noted here that this parameter is effectively bounded below by zero and there is a limit on the negative effect any observation can have on the parameter estimation therefore making large positive influence values more pronounced. Considering β, upon first inspection, aside from two large influence measures there is a fairly flat relationship between Post Operative CA19.9 and the influence measures, even at large values. Upon closer inspection, there is some change in the relationship between log Post Operative CA19.9 values of 4 and 6, and again between 6 and 8. This change, although small, can be seen to correspond to the points in the function that are chiefly concerned with the growth of the functional relationship and immediately afterwards as shown by Figure 3.4.

3.7 Discussion

In this chapter, some attention has been given to the functional form by which covariates enter a survival model. It has been shown that the linear relationship typically given by $\exp(\beta z)$ may not always be appropriate. Typical strategies to account for this depend upon finding some satisfactory transformation of z to obtain either a model that satisfies all diagnostic assessments or provide an optimal fit of the data to the model.

It has been shown that this approach can firstly give misleading results and secondly, results in models that are difficult to interpret in a meaningful fashion. It has also been shown that failure to correctly model covariates can introduce some level of bias through confounding into the estimation of other parameters within a model.

As an example, data are taken from the ESPAC-3 trial to investigate the effect

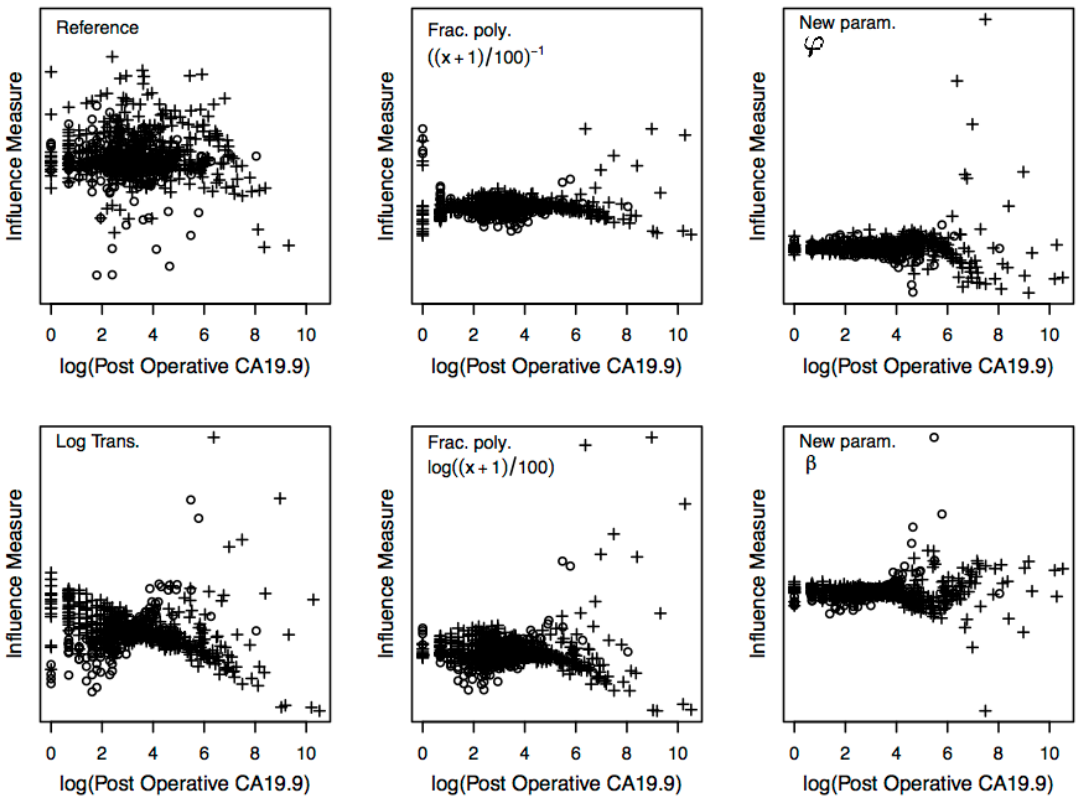

Figure 3.6: Influence measures for models fit to the ESPAC-3 dataset, observed events are represented by a cross, censored events by a circle.

that Post Operative CA19.9 has upon the survival of patients undergoing resection for pancreatic ductal adenocarcinomas. Strategies investigated included a standard 'reference' approach, using a basic log transformation and using more complete set of transformations as given by a fractional polynomial model. Compared to these approaches was a new parameterisation which was developed based on providing a functional form that would be in-keeping with clinical thinking based on the effect that Post Operative CA19.9 has with patient survival.

The results show that, as may be expected, the worst model in terms of model AIC was the reference model. Interestingly however, aside from a few outliers within the residual plots, this form performed reasonably well in terms of the model diagnostics.

The best performing model was the model that was obtained using fractional polynomial techniques. Whilst this may have the best AIC however, there are issues with this model as no value of Post Operative CA19.9 will have a zero effect upon the baseline hazard function. Furthermore, as the covariate enters the model through two different transformations, direct interpretations are difficult without the aid of graphical representations. Neither a standard log transformation nor the fractional polynomial approach solve all of the issues with regard to model diagnostics either.

The new parameterisation performs better than the standard log transformed model, although not as well as the fractional polynomial model in terms of model AIC. Whilst the new parameterisation may not perform as well in terms of model fit, it is more acceptable from a clinical perspective.

For the analysis of the ESPAC-3 data, a maximum hazard ratio of 3.77 is obtained for the new formulation which is not achievable with any other linear predictor. Based on this new formulation, clinicians can state that any patient who presents with a large Post Operative CA19.9 will have a 4 fold increase on the baseline hazard function relative to a similar patient with a low value. Under other formulations a similar patient may be given an unrealistically poor prognosis due to the limitations of the model assumed. This is still the case even under a fractional polynomial approach.

Another benefit over the fractional polynomial approach is that by assuming some known functional form (as we do with the standard notation as well as the new parameterisation), extrapolation beyond the observed range of the data is possible. Predictions under the fractional polynomial approach, more so than other approaches, depend on the location of the data on which the model are fit and model performance in the extremities can often be compromised to provide the best fit of the model. This is unfortunate as it is often the patients who lie in the extremities upon who clinicians may find it more difficult to provide a prognosis for. By assuming that the functional form has an upper asymptote it can also be stated that if a patient presents with a covariate value larger than anything that has yet been observed, the effect will still be a 3.77 fold increase in the baseline hazard function relative to a similar patient.

Another advantage of the new parameterisation is that the parameter that controls the growth rate of the function, β, can be estimated without any undue influence from 'large' covariate values. This is illustrated by the residual plots with the plots given by β for the new parameterisation giving the most satisfactory results in terms of model residuals and influence measures.

Drawbacks of this approach are, like the standard form, a functional form is assumed which may be wrong. Furthermore, the estimation of the upper asymptote is dependent on the number of 'large' covariate values that are observed and estimation may become unreliable for covariates that do not have a large number of outliers or in small datasets as was demonstrated in the simulation study. It should also be noted that the use of a more complex functional form may depend on the form of the covariate. Whilst in general, the model was developed for biological covariates which have the natural bounds $[0, \infty)$, interpretation of model parameters may be altered to transformed covariates which are subject to some shift in location and some care should be taken.

Whilst there was generally little difficulty in using standard partial likelihood approaches in estimating model parameters, model residuals could not be obtained in the

standard fashion due to the non standard functional form. Analytical forms for the residuals, based on the methodology behind Schoenfeld residuals are produced along with an analytical form for the influence measures based on the methodology of Reid and Crepeau.

In the following chapter, some investigation of the assumptions inherent to survival models are carried out, with particular reference to the assumption of proportionality which dominate current statistical practice.

Chapter 4

Use of an Asymmetry Parameter in the Analysis of Survival data

4.1 Introduction

In this chapter, some exploration of the analysis of time-to-event data is carried out where the assumption of proportional hazards is not met. A review of current available methods is explored and applied to a historical gastric cancer dataset. An overview of an alternative method to the Kaplan Meier plot for visualising survival data, the PP plot, is given and its uses for assessing model fit investigated. Lastly a new method allowing for direct modelling without proportional hazards is proposed.

4.2 Non-proportional hazards

By far the most popular method of modelling the association between a set of covariates and a time-to-event outcome is via the proportional hazards model. Despite its popularity however, the suitability of this approach is dependent upon the assumption of proportionality. Indeed, the hazard ratio to which many have become familiar is only a valid measure under this assumption. Moreover, non-proportional hazards do occur in practice, a notable example being Mok et al.[87]. In light of this, much work has been carried out both in assessing the proportionality assumption and in analysing data in which this assumption is violated.

The reasons for the occurrence of non proportional hazards can be wide and varying. For example, considering a simple two arm trial, non proportional hazards can occur if the effect of a single treatment is seen to decay over time at a greater rate than another treatment. Conversely a new treatment may be associated with severe early toxicities, but with the survival outcome of these patients improving following the initial dangerous phase. Both of these situation are examples of time-dependencies in the modelling of patient performance and result in notable biases if modelled assuming proportional hazards as noted by Schemper [88]. In other situations, inclusion of time

dependencies may not be appropriate and alternate methods such as the inclusion of 'cure fractions' or modelling on an odds scale may be attempted.

In this Sections that follow, a brief review is provided on some of the methods that have been applied to assessing the assumption of proportional hazards. Following this, there is some exploration of various methods that are applied to modelling survival data in this context. Lastly, a case study of a gastric cancer dataset in which non-proportional hazards are observed is used to illustrate the suitability of some of the different methods proposed.

4.2.1 Assessing proportionality

The first step in assessing proportionality typically occurs via visual inspection. In many cases, such as with crossing survival curves, Kaplan Meier estimates will quickly identify where non-proportional hazards exist. Other scenarios are less obvious. Of further use may be the 'log negative log' plot. Formally, take the survival function estimated by the method of Kaplan and Meier as $S(t)$, the transformed quantities $S_l(t)$ are obtained such that

$$S_l(t) = \log[-\log\{S(t)\}].$$

Under the assumption of proportional hazards, plotting $S_l(t)$ against the log of time across different treatment groups should produce plots of parallel lines. Any departure from this can be taken as evidence against proportional hazards although there is a lack of any formal test to assess this. A further approach is that taken by Arjas [89] where the cumulative number of events for different levels of a covariate are plotted against cumulative hazard estimates. Under proportional hazards, different levels of a categorical variable should show a similar relationship to one another, any divergence is then evidence of non-proportionality.

Expanding on visual assessments, a number of more formal approaches for testing the assumption of proportionality have been developed. An empirical comparison is provided by Ng'andu [90] and much of the rational for testing proportionality is based on assessing the relationship of a model against time. The remainder of Section 4.2.1 shall be concerned with providing a brief overview of a number of approaches available.

Linear correlation test

A test proposed by Harrell [91] is obtained from Fisher's z (not to be confused with the notation for patient covariates) transformation of Pearson's correlation between rank time and model residuals as obtained by Schoenfeld's method. Given the observed correlation, $\hat{\rho}$, a formal test statistic, z, is obtained as

$$z = \hat{\rho}\sqrt{(n_\mu - 2)/(1 - \rho^2)},$$

where n_μ is the number of uncensored observations. Comparing z against a standard normal distribution tests the null hypothesis that $\rho = 0$. A significant p-value gives evidence for a model fit which is not time independent and therefore not proportional.

Weighted residuals test

This is the test as proposed by Grambsch and Therneau [85]. Briefly, under a model where the coefficients are allowed to vary with time, that is $\beta(t) = \beta + \gamma g(t)$, it is shown that the expected value of the Schoenfeld residuals for the k^{th} death time, where β is assumed to be known can be expressed as

$$E[r_k(\beta)] \approx V(\beta, t_k)G(t_k)\gamma.$$

Here $V(\beta, t_k)$ is the variance of β at the k^{th} death time and $G(t_k)$ is a diagonal matrix with elements, $g(t_k)$ which are some functions of time. Grambsch and Therneau show that a test for $\beta(t) = \beta$ (or $\gamma = 0$) is equivalent to a generalised least squares test on the Schoenfeld residuals.

Score process test

A test based on the score process was proposed by Therneau et al. [92] and is based on martingale residuals. Here, under the counting process notation, martingale residuals denoted $\hat{M}_i(t)$ and defined in Section 2.3.4, can be interpreted as the difference between the observed number of events and the expected number of events at time t. A test for non proportionality for each of J covariates can be obtained by summing over each patient's score process at time t. A test is derived from $\sup_t \sum_i L_{ij}(\hat{\beta}, t)$ where

$$L_{ij}(\hat{\beta}, t) = \int_0^t \{X_{ih}(s) - \bar{X}_j(b, s)\} d\hat{M}_i(s),$$

with 'large' value providing evidence against proportional hazards.

Omnibus test

The most popular omnibus test is that provided by Moreau [93] and Quigley [94]. The method explicitly allows for interaction between time and covariates. Specifically, the time domain is partitioned into k disjoint intervals and the predictor is specified by

$$\exp\{(\beta + \gamma_k)z\}.$$

A model is proposed here whereby the hazard ratio for a covariate is only piecewise constant. Moreau proposed a test for assessing $H_0 : \gamma_1 = \gamma_2 = ... = \gamma_k = 0$. Similar tests can be obtained by assuming models such as Freidman's Piecewise Exponential Model [38], or a flexible parametric approach [73] and allowing for the interaction between the covariates and the partitions in the time domain.

Time dependent covariate test

The inclusion of a time dependent covariate to check the PH assumption was proposed in the original paper by Cox [37]. Considering a single two level covariate, the predictor of a model is defined as

$$\exp\{\beta z + \gamma z g(t)\},$$

where $g(t)$ is some function of time. Commonly $g(t) = t$ and $g(t) = \log(t)$ are used. If it is assumed true that $\gamma = 0$, a test for $\gamma = 0$ can be obtained from a likelihood ratio test statistic using $-2\ln\{L(\hat{\beta}, 0)/L(\hat{\beta}, \hat{\gamma})\} \sim \chi^2_{\Delta p}$ where Δp is the change in the degrees of freedom between the null model and the time dependent model.

4.2.2 Modelling non-proportional hazards

This section explores various methods of assessing the effects of a covariate on a time-to-event response when proportionality can not be assumed.

Time dependent covariates

As introduced by Cox [37] and defined in Section 4.2.1, this method of modelling explicitly allows for the effect of a covariate to vary with some function of time $g(t)$. This method is particularly useful in comparing two treatments from a randomised trial where the effect of a treatment may decay with time. However, if the nature of the covariate has no obvious relationship with time, or the relationship is not adequately explained by some linear function, then this approach may be insufficient. Furthermore, the model itself is dependent on the user deciding the function of time to be used.

Historically, Stablein et al. [95] consider both t and t^2 as functions of time to model the treatment effect in the gastric carcinoma dataset. The process of model estimation is also associated with some difficulties, as noted by Fisher and Lin [96]. Zucker and Karr [97] consider a non-parametric approach, using a penalised likelihood to fit a model with time varying covariates.

Further attempts have been made to model in a time-dependent fashion [95, 98] which have been applied to modelling cancer data [99, 100]. Lustbader [101] show that the Wilcoxon test can be obtained from the Cox model with time dependent covariates under certain conditions.

Time dependent coefficients

Time dependent coefficients have a linear predictor $\lambda(t) = f\{\beta(t), z\}$ which is not to be confused with time dependent covariates where $\lambda(t) = f\{\beta, z(t)\}$. Examples of this may be a biological variable in a clinical trial which changes with time and whose influence may be important in describing a patient's response to treatment. Treating

such a variable as fixed throughout time can lead to misleading results and the violation of the proportional hazards assumption.

Models such as these have been considered by Murphey [102] as well as Cai and Sun [103] who consider local linear procedures for estimating parameters. In both cases, time dependent covariates are included directly through inclusion of observed $z(t)$. Alternatively, $z(t)$ may itself be expressed as a function of time, for example a drug whose concentration may decay over time may have $z(t) = \exp(-zt)$. Models with some time relationship may be sensitive to the assumed dependency on time.

As an alternative to this, Giorgi et al. [104] investigate a relative survival model using B-splines to model both baseline mortality rates and the relationship between coefficient estimates and time. More popular recently is the method of joint modelling of survival data with longitudinal data first introduced by Henderson et al. [105]. Here a time dependent covariate is modelled longitudinally and the smoothed relationship with time is allowed to affect a survival model. A further overview of this method is given by [106] whilst Tseng et al. [107] apply the methodology to accelerated failure time models. Chi and Ibrahim [108] bring together joint modelling and cure fraction modelling. An advantage of the joint modelling procedure over more standard methods is that they allow for errors in measurement and a smoothed estimate of the effect of a covariate with time to be incorporated into a survival model.

Piecewise models

A method which may be seen as an extension of Moreau's Omnibus test [93] is the use of piecewise models, the most popular being the Piecewise Exponential Model (PEM) proposed by Friedman [38]. Here, non-proportional hazards can be accounted for by allowing for the interaction between the covariate estimates and the time-grid used to partition the time domain. Whilst these methods can be extremely useful in explaining the nature of any relationship with time, they can be costly to fit in the number of parameters that are required and are themselves dependent upon the points at which the time domain is partitioned.

Flexible parametric models given by [73] can be considered in a similar fashion to piecewise models. Here a spline function with knots κ is fitted to the cumulative hazard function. Parameter estimates β can be interacted with the spline function in the same fashion as they can with the PEM.

Accelerated failure time models

An alternative to proportional hazards modelling is the accelerated failure time approach. For full description see Kalbfleisch and Prentice [109]. Here, under parametric forms of the proportional hazards model, define

$$S_\beta(t, z) = [S_0\{g(t, \boldsymbol{\kappa})\}]^{\exp(\beta z)}$$

where $g(t, \boldsymbol{\kappa})$ is some parametric family set to describe the behaviour of a baseline survival (or hazard) function. Under an accelerated failure time model,

$$S_\beta(t, z) = [S_0\{g(t, \boldsymbol{\kappa}, \exp(\beta z))\}]$$

is obtained. Here the predictor acts directly in a multiplicative fashion on time (or in an additive fashion on the log of time). Specifically

$$\log(t_i) = -z_i\beta + \log(\nu_i)$$

where $\log(\nu_i)$ is the error term that is set from some parametric distribution. Note here that ν_i is not to be confused with the notation in Chapter 2. Typically, unlike PH models, AFT models are predominantly dependent upon some distributional assumption. AFT models however do have the advantage that they do not depend upon the assumption of proportional hazards.

Restricted mean survival

A method proposed by Royston and Parmar [110] is that of restricted mean survival (RMST). The RMST is calculated up to time point t^*, denoted by $\mu(t^*)$, and is evaluated as the area under the survival curve to time t^*

$$\mu(t^*) = \int_0^{t^*} S(u)du.$$

Royston and Parmar show that estimates can either be taken directly from the Kaplan Meier estimates, using a pseudo value approach to obtain standard errors or using a parametric modelling approach. The difference in mean survival times $\mu_2(t^*) - \mu_1(t^*)$ can be used to assess differences between treatment groups irrespective of the form of the relationship. In this way direct comparison between groups can be calculated when non-PH are observed. Formal tests based on the difference are only shown to be powerful under the PH assumption.

Averaged hazard ratios

Averaged hazard ratios were proposed by Schemper [88] and later by Xu and O'quigley [111]. Here the score function for Cox's partial likelihood is supplemented by a weight function, $f_r(t)$, such that

$$U = \sum_{i=1}^{N} f_r(t)\Big[z - \frac{\sum_R z\exp\{\beta z\}}{\sum_R \exp\{\beta z\}}\Big].$$

Options for the weight function suggested are the number of patients at risk or the scaled Kaplan Meier estimates. Schemper suggests use of the scaled Kaplan Meier estimates but does not state whether these should be stratified by treatment arm. It is shown that this approach can help produce hazard ratios which are less susceptible to the systematic bias introduced from a non-proportional relationship.

Cure fraction

The analysis of survival data where a proportion of the population under inspection are cured in the sense that their survival profile converges to that of the overall population can be a cause of non proportionality. This is because under proportional hazards, survival functions only converge theoretically at values of zero and one. Take for example, the mixture cure model as explored by Lambert [112], the survival function is redefined as

$$S(t) = \eta + (1 - \eta)S(u)$$

where η is the cure fraction to be estimated. They have been further applied to a general class of semi-transformational survival models by Yin and Ibrahim [113].

Proportional odds

Proportional odds models [114, 115] have gained some interest as they assume a different underlying relationship between a covariate and a time-to-event outcome. These have been explored further by Chen et al. [116], looking at the inclusion of external time-varying covariates. Murphy [117] considered a maximum likelihood estimator. These models were made more accessible by Cheng et al [64] who describe a semi-parametric transformational model based on either the logarithmic or Box Cox family of transformations. Both the proportional hazards and the proportional odds models are special cases of this model.

Outlying covariates

As has been detailed in Chapter 3, a single outlying covariate value can be enough for a non-proportional hazard relationship to be observed. Here, a usual approach is to transform the covariate or to apply some functional form to explain the relationship a covariate has with the baseline hazard function.

Additive models

Some use of additive models as described by Aalen [118] has been explored as this model does not rely on a proportional hazards assumption [119]. Here a covariate structure acts on a baseline hazard function in an additive fashion as opposed to a relative scale as for proportional hazards models such that

$$h(t|z) = h_0 + \beta^T x.$$

Details for fitting a semi-parametric additive model have been provided by [120]. This approach has been explored with applications to breast cancer [121] and gastric cancer [122] datasets.

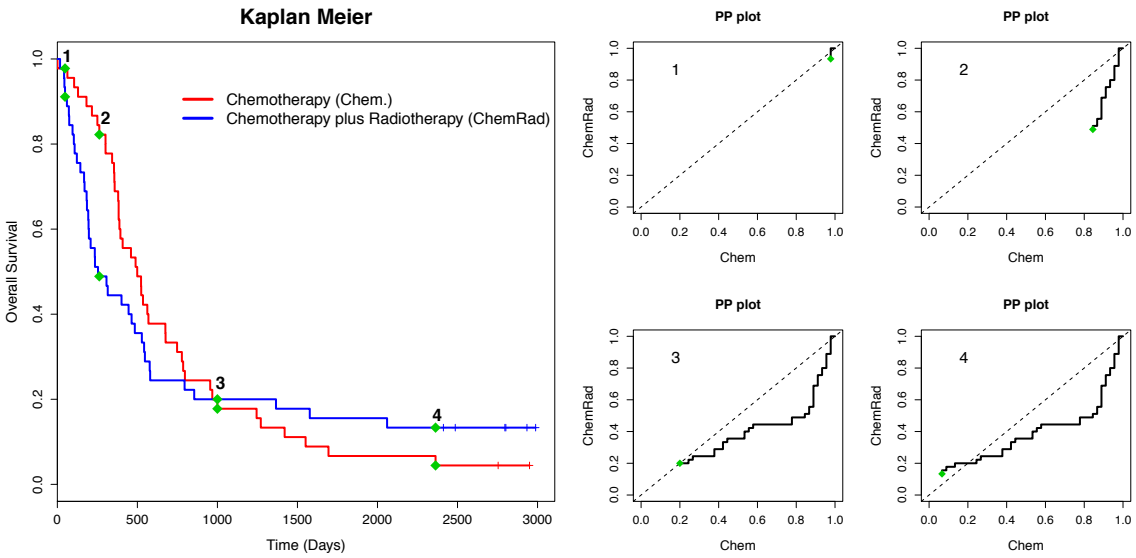

Figure 4.1: Figure to illustrate the process of obtaining a PP-plot from Kaplan Meier survival estimates

4.3 The PP-plot

PP-plots have been used and developed by Cox [123] as a method of visualising and measuring the relationship between two survival distributions. It is given special attention here as they are a recently developed approach that will be used throughout this chapter. Whilst originally used to test for differences between survival distributions, here that are used only to visualise survival data as a time independent alternative to the Kaplan Meier plot.

The PP-plot offers an alternative to the widely used Kaplan Meier plot. Whereas the Kaplan Meier plots shows survival estimates against time, the PP-plot plots the estimated survival of two (or more) survival functions against one another. Where more than one survival function is visualised, one survival function is chosen as a reference level against which other survival functions are measured.

Figure 4.1 gives an illustration of the formation of the PP plot. Here the Kaplan Meier estimates from both arms of the gastric cancer dataset are displayed along with the formation of the PP plot as it develops over time.

Importantly, the form of the PP-plot allows comparison between two groups of patients which are independent of time. As an example here, comparing survival estimates against $\log(t)$ would change the visual inspection of the Kaplan Meier plot but not the PP- plot. This gives the PP-plot an advantage over the Kaplan Meier plot in many situations as the Kaplan Meier estiates may be prone to large shifts at time points where data are sparse, a point which is further noted by Schemper [88].

A further advantage of the PP-plot is that it facilitates the visualising of fitted

models without the need to specify some fitted baseline hazard (or survival) function. For example, visualising the fit of a Cox model would require that first some non-parametric baseline survival estimates are obtained. This may be achieved by ensuring any covariates are coded within the design matrix as (-0.5, 0.5) for each factor level as opposed to (0,1). A pooled Kaplan Meier estimate over all covariates can provide a baseline survival function with which to adjust. Nevertheless Kaplan Meier plots can quickly become 'busy'.

For a PP-plot, under the proportional hazards model,

$$S_1(t) = S_0(t)^\theta \tag{4.1}$$

where θ is a hazard ratio and $S_k(t)$, $k = 0, 1$ are survival functions. The relationship between two survival function can then be directly plotted without the need for a baseline survival function. Furthermore as two survival functions that obey the PH assumptions must satisfy equation 4.1. the PP plot can be used to assess PH visually.

In the development of the PP-plot, Cox utilised the fact that the plot area is unity and that two equivalent survival curves will follow the diagonal from (1, 1) to (0, 0). Two tests are then developed based upon the area between the curve and the diagonal and the arc length of a fitted curve. Cox goes on to show that these tests retain good levels of power irrespective of PH assumptions.

Throughout the remainder of this chapter, PP-plots shall be used, predominantly as a tool for illustrating model fit under differing assumptions.

4.4 Case study - gastric cancer dataset

The data used in this section are from the gastric cancer dataset [36]. Initially, proportional hazards are assessed visually and via the time dependent covariate test, the weighted residual score test and the linear correlation test following the recommendations of Ng'andu [90]. Following this, modelling of the data are carried out using a time dependent covariate and piecewise and restricted mean survival approaches.

4.4.1 Assessing non proportional hazards

A Kaplan Meier plots is shown in Figure 4.2 (repeated from Figure 4.1) and shows clear crossing of the survival curves. The log negative log plot is also included, from which it is clear that the proportional hazard assumption is not met since the two curves are not parallel.

Of the three formal testing approaches taken, both the weighted residual and linear correlation depend upon the calculation of the Schoenfeld [83] residuals, $\{r_i\}$, information for which is given in Chapter 3. Figure 4.3 shows the residuals plotted against time along with a fitted spline function to give some indication of the correlation the time.

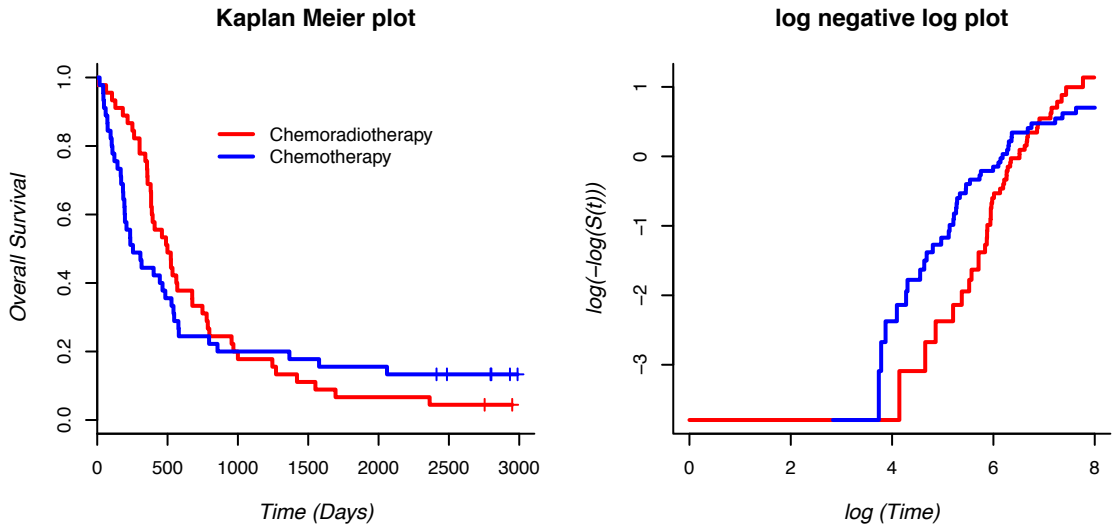

Figure 4.2: Survival estimates illustrated by means of a Kaplan Meier and log negative log plots

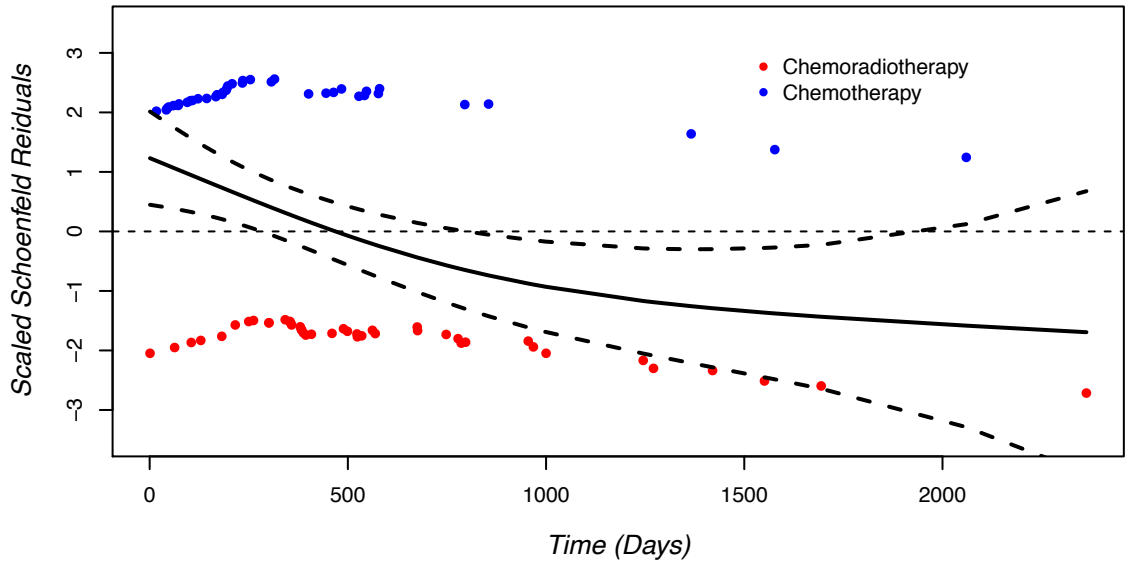

Figure 4.3: Scaled Schoenfeld residuals plotted against time for the Gastric Cancer dataset

Both the weighted residuals and the linear correlation approaches depend upon the correlation ρ between time (in its absolute and rank form respectively) and the residuals. These correlations are given in Table 4.1 along with tests for their significance based on the appropriate test statistic. Note that the test statistic for the linear correlation test is usually presented on the standard normal scale; here the test statistic is squared to be presented on the χ^2 scale to remain consistent.

Also shown in Table 4.1 are the results for fitting a standard Cox model with a time-dependent covariate. Here the test is based on twice the difference in the model deviance between the models with and without the time-dependent covariates. All three methods show high levels of significance as is to be expected given the visual inspections.

Test	ρ	χ^2	p
Weighted Resid.	-0.35	10.29	<0.001
Linear Correlation	-0.40	15.51*	<0.001
Time dep cov.		9.20	<0.001

Table 4.1: Table to show the results of testing the PH assumptions via the weighted residuals, linear correlation and time dependent covariates method. *linear correlation test statistic is expressed as a Chi-square test statistic to remain consistent.

4.4.2 Modelling non proportional hazards

Given that non-proportional hazards have been established, some consideration is given to the modelling techniques for non PH models. Specifically, the time-dependent covariate, restricted mean survival and piecewise models are considered. Where appropriate model fit is illustrated using the PP-plot. Some discussion on the model results are presented.

Time dependent covariate

The time dependent covariate method is used in Section 4.1.1 in order to test the PH assumption. Under the basis of time dependent covariates, the effect of the treatment is assumed to change constantly with time. Some functional form of time is required and here it is assumed that the treatment effect changes in a linear fashion with the natural logarithm of time although other functions of time may be considered.

Table 4.2 details the model fit. Initially, the standard model with no time-dependent covariate is fitted. Here a hazard ratio (95% confidence interval) of 1.11 $(0.73, 1.72)$ shows no significant difference. A test of the PH assumption in Section 4.1.1 shows the time-dependent model to be an improvement over the standard model. The results in Table 4.2 show that both coefficients are now statistically significant at the 0.05 level.

Interpretation of the time-dependent model however cannot be made independent of time. Here for example, the hazard ratio at 30 days is given by $\exp(4.15 - 0.69 \times$

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Arm	0.11	1.11	0.22	0.48	0.63
Arm	4.15	63.64	1.51	2.76	0.01
Arm × log (time)	-0.69	0.50	0.25	-2.75	0.01

Table 4.2: Results of a standard and time dependent proportional hazard model fitted to the gastric cancer dataset.

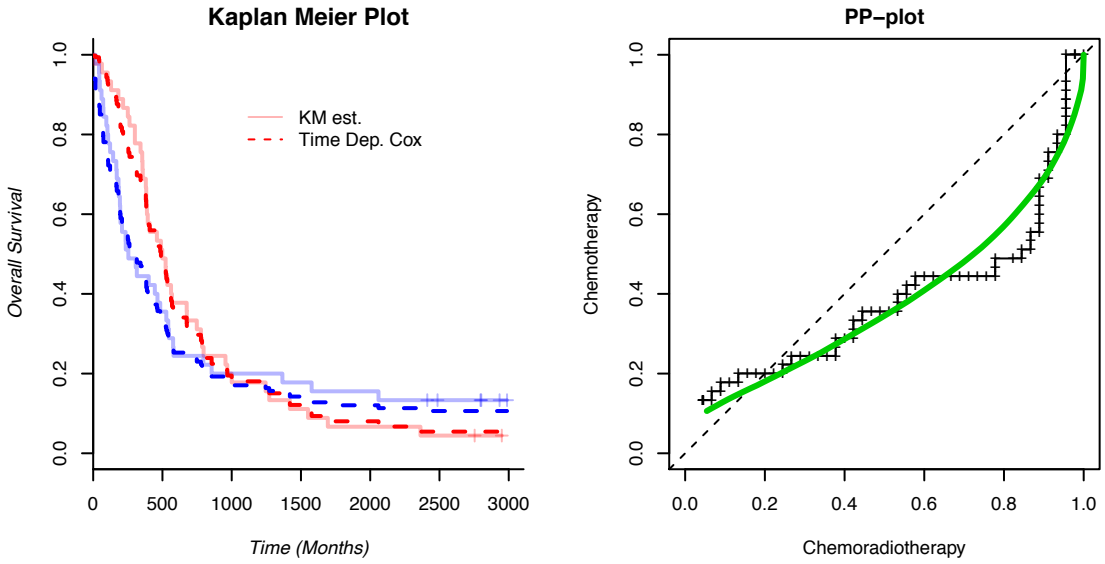

Figure 4.4: Figure to illustrate the fit of a time dependent covariate model using a Kaplan Meier and PP-plot.

log(30)) = 6.05. At 30 days therefore, the hazard of an event is roughly six times that for chemotherapy group as it is for the chemotherapy and radiotherapy group. At 365 days, a hazard ratio of 1.08 is obtained. Whilst still a greater hazard in the chemotherapy group, the effect is greatly reduced. By 500 days a hazard ratio of 0.87 is obtained showing that at this point, there is a greater hazard on the Chemotherapy plus Radiotherapy arm.

Figure 4.4 gives an illustration of the model fit both on the Kaplan Meier plot and through the PP-plot. Visual inspection of both graphs show that incorporation of a time-dependent effect gives a reasonable fit to the data.

Piecewise model

A second modelling approach is to fit a piecewise model. The example given here is the Piecewise Exponential Model (PEM), full details of which can be found in Chapter 2. Similar approaches can be taken by assuming the flexible parametric approach of Royston and Parmar [73].

Under a standard PEM model, the baseline hazard function would be estimated to be piecewise constant hazard rate but with a single hazard ratio estimated across

all partitions. The alternative under non-PH is to allow separate hazard ratios to be estimated in each partition.

The results of both the standard model and the non-PH model are included in Table 4.3. Here, the time-grid is partitioned arbitrarily at $a = (0, 182, 365, 730, 1460, 3044)$ where time is measured in days. There is some clear increase in the hazard ratio over time however the decrease from 1.26 to 1.14 in the last two partitions may be taken as some evidence that the monotonic relationship assumed when fitting a time-dependent covariate model is not valid.

	Prop. Hazards		Non Prop. Hazards	
Parameter	Estimate	Std. Error	Estimate	Std. Error
λ_1	-6.70	0.25	-6.20	0.27
λ_2	-6.34	0.24	-5.93	0.30
λ_3	-6.28	0.22	-6.38	0.33
λ_4	-6.96	0.30	-7.71	0.58
λ_5	-7.88	0.47	-8.34	0.71
β_1	0.10	0.22	-1.15	0.52
β_2			-0.69	0.45
β_3			0.23	0.42
β_4			1.26	0.66
β_5			1.14	0.91

Table 4.3: Table to show results from a PEM for proportional and non proportional hazards.

Interpretation is aided by plots of the hazard ratio over time and a PP-plot given in Figure 4.5. Here plots of the hazard ratio show clearly the increase of the hazard through time. Confidence intervals of the time dependent hazard ratios are illustrated by the shaded areas. These increase due to the smaller amount of information (i.e. events) available in the later time partitions. The standard model is also represented as a horizontal line for comparison.

The PP-plot also shows a good fit of the model to the data. The step-wise nature of the model is also clearly visible.

The PP-plot also illustrates how the nature of the hazard function is dependent on how the time-grid is defined. In some situations, definitions of the time-grid may be guided by knowledge of the treatments involved. Usually however, the decision will either be pragmatic or data-driven. Lastly, it should be noted that the piecewise model is a relatively computer intensive approach, with ten parameters needed to be estimated compared to just two with a time-dependent approach.

Restricted mean survival

Here the restricted mean survival times (RMSTs) are applied using a flexible parametric modelling approach. Here flexible parametric models are fit to each treatment arm

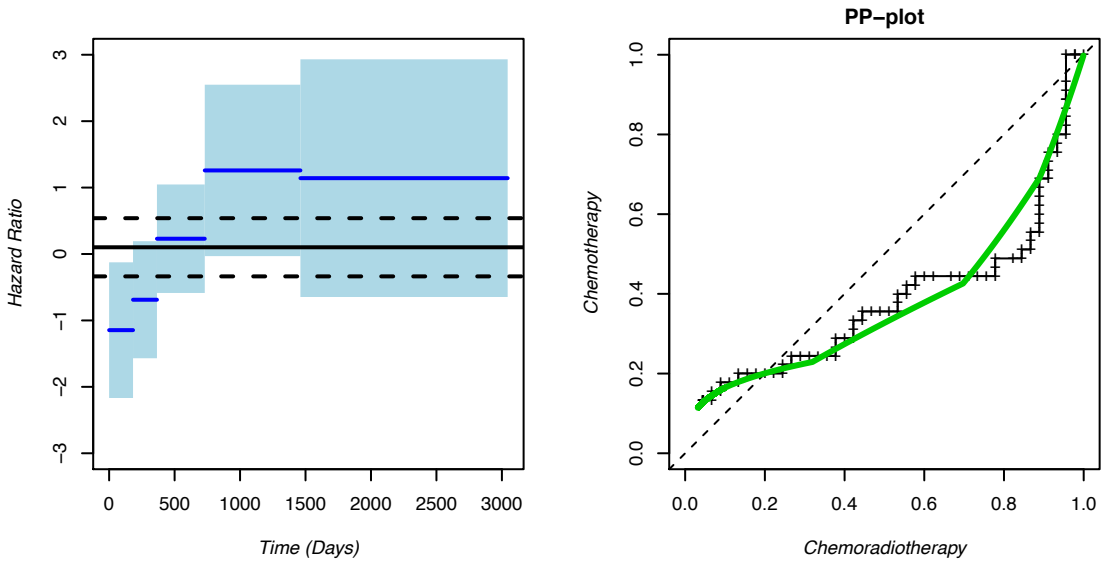

Figure 4.5: Figure to illustrate the fit go the piecewise exponential model

separately. Given a parametric form for the hazard function, estimated RMST can be directly obtained as the area under the each survival curve. Standard errors for each estimate are robustly estimated using a bootstrap approach. Analyses are carried out using the 'postestimation' command in Stata version 12.

The definition of the RMST as the area under the survival curve between times 0 and the 'maximum time' t^* means that estimates are somewhat dependent upon the definition of t^*. Here, four values, $t^* = 1500$, 2000, 2500 and 3000 are used.

Analyses are carried out fitting spline functions for each treatment arm with five degrees of freedom. This can be thought of as similar to the piecewise approach with five partitions in the time-grid. Results are given in Table 4.4 in the form of RMSTs with associated standard errors and 95% confidence intervals. Also presented are the difference in RMST between treatment arms. The results show that for all definitions of t^* the RMST is larger in the chemoradiotherapy arm. For $t^* = 3000$ however the difference has become almost negligible. The largest difference is observed for a value of $t^* = 1500$. As t^* increases, the standard errors of the RMST estimates increase and the difference between treatments decrease. For none of the estimates is there significant evidence that the two survival profiles are different from one another.

Overall, from using RMST estimates, there is no conclusive evidence of a difference in the two treatment arms. It is clearly highlighted however that interpretations are somewhat dependent on t^*. In the context of a randomised clinical trial, t^* may be defined a-priori to fit in with a design feature of the trial. It is a shortcoming of this procedure however that model interpretations may be directly dependent on this user set parameter.

t^*	Treatment Arm	Restricted Mean Survival		Diff. between means	
		est (se)	95% CI	est (se)	95% CI
1500	Chemo.	517.7 (76.1)	(368.5, 666.8)		
	Chemoradio.	630.4 (64.5)	(503.6, 756.4)	112.8 (99.7)	(-82.7, 308.3)
2000	Chemo.	593.4 (99.4)	(398.5, 788.2)		
	Chemoradio.	677.18 (80.46)	(519.5, 834.88)	83.81 (127.89)	(-166.9, 334.5)
2500	Chemo.	662.6 (121.9)	(423.7, 901.5)		
	Chemoradio.	708.2 (91.2)	(529.4, 887.0)	45.56 (152.3)	(-252.9, 344.0)
3000	Chemo.	727.34 (144.58)	(444.0, 1010.7)		
	Chemoradio.	729.6 (103.7)	(526.3, 932.9)	2.3 (178.0)	(-346.5, 351.1)

Table 4.4: Estimates of restricted mean survival times for each arm and their difference for various 'maximum time' points t^*

4.4.3 Discussion

Three approaches were explored for the modelling of a gastric cancer dataset with non-proportional hazards, the time dependent covariate approach, a piecewise approach and an approach based on restricted mean survival times. Whilst there are few difficulties in applying each method, model interpretations can vary and addressing the key question of interest, is one treatment better than another, is fraught with difficulty.

Both the time dependent covariate and piecewise approaches provide reasonable fits to the data but also have their shortcomings. For the time dependent covariate approach the user must specify the nature of the relationship with time. Here $\beta(t) = \beta_0 + \beta_1 \log(t)$ is used. This assumes some linear relationship with the natural log of time, an assumption which may not be valid in light of the results from fitting a piecewise model. The piecewise approach is itself dependent on the setting of a reasonable time-grid and is more expensive to fit due to the potentially large number of parameters required. Nevertheless, both approaches demonstrate that firstly there is good evidence that the two treatment arm have responses that are not taken from the same distributions. Secondly, it is shown that initially the hazard for an event is significantly greater for the Chemotherapy compared to the Chemotherapy plus Radiotherapy arm.

Analysis using RMST attempts to directly answer the question of which treatment arm should be preferred. Whilst from a clinical perspective this may be the most useful as it is attempting to identify the group with the best overall survival, it is flawed by the dependence on the setting of t^*. Furthermore, under crossing hazards like the ones seen here, comparing the two groups with only a single parameter may be viewed as oversimplifying the problem.

4.5 Modelling non-proportionality via an asymmetry parameter

In this section, a new approach to the analysis of survival data with non-proportional hazards is proposed. Here, models are re-parameterised so that, as opposed to interpreting in terms of a hazard ratio, models are assessed in terms of a dispersion parameter, measuring the magnitude of divergence due to a covariate and an asymmetry parameter measuring departure from proportionality.

Assume that time-to-event models are expressed in terms of a counting process [40] and consider the semi-transformational models [64, 65] which allow for the form of the survival models to be interpreted on scales other than the hazard scale. Following this notation, each patient is defined via a counting process, $N(t)$, which records the number of events until time t. Also let $Y(t)$ be the at risk process defined as $Y(t) = I(T \geq t)$ where $I(.)$ is the indicator function. Let Z be a $n \times p$ covariates matrix with associated parameters $\boldsymbol{\beta}$. A conditional survival function is defined as

$$S(t|Z) = \Psi\{\exp(\boldsymbol{\beta}^T Z)\Lambda(t)\},$$

where $\Lambda(t)$ is a non-decreasing function, $\int_0^t \lambda(u)du$ with λ as the intensity function to the associated counting process. Note that this conflicts slightly with the notation used in Section 2.3.4 and is not to be confused with the rate parameters used in the exponential and piecewise exponential models. Allowing $\Psi(.)$ to be either the Box-Cox or logarithmic family of transformations allows the estimation of both the proportional hazards or proportional odds models as special cases. Specifically, $\Psi(x) = \exp(-x)$ gives a proportional hazards model and $\Psi(x) = (1 + x)^{-1}$ gives the proportional odds models. In the special case of the proportional hazards function, note that λ is an analogue to the hazard function. This class of models is extended by introducing a second 'asymmetry' parameter which allows for departure away from the assumption of proportionality. The asymmetry parameter, denoted α, acts on a set of covariates U which may contain elements of Z. As an example, when modelling a single two level covariate, the definition would be $U = Z$. The relationship for conditional survival is re-defined as

$$S(t|Z, U) = \Psi\{\exp(\beta^T Z)\Lambda(t)^{\exp(\alpha^T U)}\}. \tag{4.2}$$

In this form, model (4.2) looks similar to the relationship observed by assuming a Weibull distribution where both the scale and shape parameter are allowed to vary dependent on covariates. The similarity ends here however, as this approach requires no dependency on a parametric form for a baseline hazard function. A formulation similar to this was proposed by Quantin et al. [124] who propose a test on α as a means of assessing proportionality. Full details describing the behaviour of the $\theta = (\beta, \alpha)$ are

given in Section 4.5.2. Justification for the derivation of the asymmetry parameter with respect to the proportional odds model are given in Section 4.5.1.

With respect to model estimation, given (4.2), a hazard function is defined as

$$h(t|Z,U) = \frac{\Psi^{'}\{\exp(\beta^T Z)\Lambda_0(t)^{\exp(\alpha^T U)}\}}{\Psi\{\exp(\beta^T Z)\Lambda_0(t)^{\exp(\alpha^T U)}\}}\Big[\exp(\beta^T Z + \alpha^T U)\Lambda_0(t)^{(\exp\{\alpha^T U\}-1)}\lambda_0(t)\Big].$$

Here λ_0 is a baseline intensity process. A log likelihood function is given by

$$l(t|Z,U,\theta) = \sum_{i=1}^{n}\nu_i\Big[\log(\lambda_0(t)) + \beta^T Z + \alpha^T U + (\exp\{\alpha^T U\}-1)\log(\Lambda_0(t)) + \log\big(-\Psi^{'}\{\exp(\beta^T Z)\Lambda_0(t)^{\exp(\alpha^T U)}\}\big)\Big] + (1-\nu_i)\Big[\log\big(\Psi\{\exp(\beta^T Z)\Lambda_0(t)^{\exp(\alpha^T U)}\}\big)\Big], \tag{4.3}$$

where $\nu_i = 1$ represents an observed event and $\nu_i = 0$ represents a censored observation. In order to maximise the likelihood, it is necessary to express the intensity function in terms of a step function and replace $\lambda(t)$ with $\Lambda\{t\}$. Maximisation of (4.3) as a non parametric maximum likelihood estimation (NPMLE) can be carried out using standard maximisation techniques available in statistical packages. Some simplification of (4.3) can be achieved when all event times are unique. Note that an estimate of the cumulative intensity function is obtained by

$$\hat{H}(t|Z,U) = -\log\big(\hat{S}(t|Z,U)\big).$$

As a cumulative intensity process can be defined for all patients at observed time-points, an estimate of the intensity process is obtained via

$$\hat{h}(t|Z,U) = \hat{H}(t|Z,U) - \hat{H}(t-|Z,U) \tag{4.4}$$

where $H(t-|Z,U)$ is the cumulative hazard function at the observed time point immediately prior to t. The log likelihood under this formulation is estimated by

$$\hat{l}(t|Z,U) = \sum_{i}^{n}\nu_i\log\{\hat{h_i}(t|Z,U)\} + \log\{\hat{S}_i(t|Z,U)\}. \tag{4.5}$$

In this form, only a definition of a survival function is required in order to produce parameter estimates. Use of the NPMLE is straightforward for small datasets. For larger datasets however, the routine can be difficult to compute due to the need to invert a large matrix in order to obtain standard errors for the fitted parameters. More attractive in this case may be an approach similar to that taken by Yin and Zeng [125] who provide an efficient algorithm which uses a Lagrange multiplier to allow the step sizes given by $\Lambda\{t\}$ to be calculated via a set of recursive equations. Parameter

estimation can then be reduced to maximising over $\rho + 2$ parameters, where ρ is the number of parameters of interest. Furthermore, it is illustrated that evaluating the model via a profile likelihood - taking the cumulative hazard function to be a nuisance parameter, standard errors of the key parameters of interest can still be obtained. More details are provided by Murphey[126]. As a guide, the model will fit to datasets of size approximately 100 and provide standard error estimates within a few minutes. For larger datasets, greater than 250 observations say, parameter estimates can be found relatively quickly but standard error estimates via a Hessian matrix may take a few hours. Models are fit using the 'optim' functions in R [127]. Code is provided in the Appendix for reference.

4.5.1 Derivation of the asymmetry parameter

Here, derivation of the asymmetry parameter with respect to the proportional odds model is given.

It has been noted by Chen [64] that a transformation of $S(t) = 1/(1+\Psi(t))$ will yield a proportional odds model. Considering only the condition where survival between two groups is compared, the proportional odds model is defined as satisfying the condition:

$$logit\{S_1(t)\} = logit\{S_0(t)\} + \psi$$

where ϕ is the odds ratio between two survival functions. Considering the two survival functions as being naturally bounded by $(0, 1)$ there has been much work on the analysis of parametric ROC curves which include the comparisons on similarly bounded function with the inclusion of asymmetry parameters. Define the following structures

$$V = logit\{S_1(t)\} - logit\{S_0(t)\}$$

$$W = logit\{S_1(t)\} + logit\{S_0(t)\}.$$

The relationship between the survival functions is then estimated via the regression formula

$$V = \varrho + \vartheta W.$$

The above can be rearranged to provided a solution for $S_1(t)$ in terms of $S_0(t)$ such that:

$$S_1(t) = inv.logit\Big\{\frac{\varrho + logit(S_0(t))(1+\vartheta)}{1-\vartheta}\Big\}.$$

Recalling $S(t) = (1+\Psi(t))^{-1}$ and noting that $inv.logit(x) = (1+\exp(-x))^{-1}$, re-write the above as

$$\Psi_1(t) = \exp\Big\{-\frac{\varrho + logit(1/(1+\Psi_0))(1+\vartheta)}{(1-\vartheta)}\Big\}.$$

Lastly note that $logit\{1/(1+\Psi_0)\} = -\log(\Psi_0)$ and rearrange to

$$\Psi(t) = \exp\left\{ -\frac{\varrho}{(1-\vartheta)} \right\} \exp\left\{ \log(\Psi_0(t)) \frac{(1+\vartheta)}{(1-\vartheta)} \right\}. \tag{4.6}$$

From (4.6) define

$$\Psi_1(t) = \omega\{\Psi_0(t)\}^{\xi}$$

where $\omega = \exp\{-\varrho/(1-\vartheta)\}$ and $\xi = \frac{1+\vartheta}{1-\vartheta}$. It follows that when measuring the difference between two levels of a covariate in terms of their relative survival odds, an asymmetric (or non-proportional) model can be formulated in terms of a divergence parameter ω and an asymmetry parameter ξ.

4.5.2 Illustration of the parameter of asymmetry

To simplify the notation, define $\phi = \exp\{\beta^T Z\}$ as a function of covariates that defines the departure away from H_0 due to Z, and $\gamma = \exp\{\alpha^T U\}$ as a function of asymmetry due to U. Model (4.1) becomes

$$S(t|Z,U) = \Psi\{\phi(H_0(t)^{\gamma})\}.$$

Here, consider ϕ to be a parameter which measures the divergence due to parameters for Z. Further, γ acts on the survival/hazard function with values of $\gamma > 1$ resulting in greater divergence at lower probabilities and values of $\gamma < 1$ giving greater divergence at higher probabilities. Here the dispersion parameter can be thought of acting proportionally on an adjusted cumulative hazard function. Under the special case of $\gamma = 1$, ϕ is interpreted as the standard proportional hazards/odds parameter if the transformations as given by [64] are followed.

Illustration of the behaviour of the model parameters when modelling a single two-level covariate is illustrated via the PP-plot. This method has the advantage that it does not require time to be included on the plot, this may be particularly attractive as the Cox proportional hazard method for estimation of covariates does not directly include time either. Figures 4.7, 4.8 and 4.9 demonstrate traditional Kaplan Meier plots of survival functions against the PP-plot so that the reader may make direct comparisons.

Figure 4.6 shows the behaviour of ϕ and γ are illustrated for both the special cases of the proportional hazards and proportional odds model. In each plot, the diagonal is refered to as the null line, as a curve that follows the diagonal would represents two identical survival functions. In both plots, the solid line in the upper triangle represents the standard proportional hazards/odds line. The upper triangle in each plot illustrates the effect of the asymmetry parameter given a fixed value for ϕ. Conversely, the lower triangle shows the different relationships that can be modelled by fixing γ and allowing ϕ to vary. These plots illustrate the wide range of flexible models that can be achieved from the two parameters.

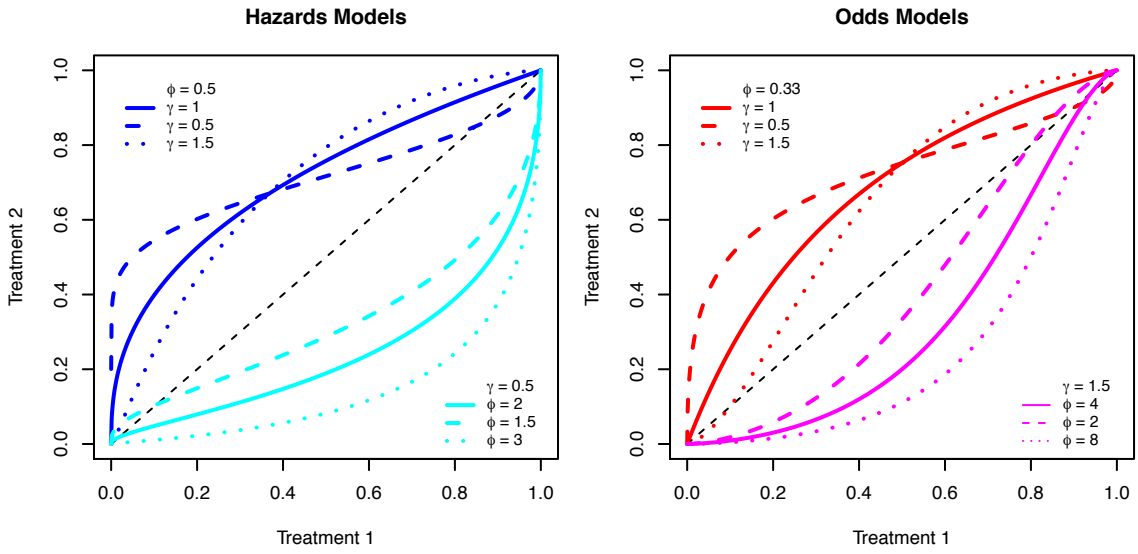

Figure 4.6: Figure to illustrate the flexibility of proportional hazards and proportional odds models with the inclusion of asymmetry parameters.

It is shown from Figure 4.6 that value of $\phi \neq 1$ has the effect of dragging the fitted model away from the null diagonal line, but only with a value of $\gamma = 1$ is proportionality attained. With a value of $\gamma \neq 1$, ϕ can no longer be regarded as a hazard ratio (or an odds ratio) and thus ϕ is referred to as a dispersion parameter. Here in the presence of non-proportionality, ϕ can still be interpreted as a parameter which measures the magnitude of the overall difference between two treatments in a similar fashion to a hazard ratio. The effect of the asymmetry parameter γ is also illustrated with values of $\gamma > 1$ resulting in greater divergence at higher probabilities and values of $\gamma < 1$ giving a greater divergence at lower probabilities.

Note here only the special cases of the proportional hazards and proportional odds models are included. A wider range of models can be achieved by allowing some model between a hazards model or an odds model using either the Box-Cox or logarithmic class of functions as described by Chen [64].

4.6 Simulation study

In this section, simulation studies are carried out for the special cases of the hazards and odds relationships. The purposes of these studies are firstly to demonstrate that parameter estimates can be reliably estimated and secondly to show that assuming proportional relationships where they are not justified can have adverse effects both in parameter bias and precision.

4.6.1 Hazards models

Two sets of survival time data were simulated using similar methods to that described by [128]. Here, for the 'control' group, the cumulative hazard function follows an exponential distribution with

$$H_c(t) = \lambda t.$$

Arbitrarily fix $\lambda = 0.2$. Measuring time in months, this ensures that the baseline survival function approaches 0 at around 12 months (Figure 4.7). In a similar fashion,set up the cumulative hazard function for the experimental group as

$$H_E(t) = \phi\{(\lambda t)^\gamma\}.$$

To simulate survival times, generate two sets of random survival probabilities, denoted S_C and S_E, from a uniform distribution with limits $(0, 1)$. Survival times t_C and t_E are obtained via

$$t_C = \frac{-\log(S_C)}{\lambda} \tag{4.7}$$

and

$$t_E = \frac{\exp\left\{\frac{\log\left(-\log(S_E)\right)-\log\left(\phi\right)}{\gamma}\right\}}{\lambda}.$$

The purpose of this approach is to simulate data from two distributions which are known not to follow proportional hazards assumption. Fix the parameters $\phi = 2, \gamma = 0.75$ ($\beta = 0.69$ and $\alpha = -0.28$ respectively). The value for ϕ is chosen as a value representative of what is observed in published clinical trials whereas γ is chosen to represent a situation where asymmetry may not be directly obvious from a Kaplan Meier plot. Figure 4.7 illustrates the fitted survival function using both a Kaplan Meier and a PP plot. On the PP plot, both the simulated relationship and a proportional hazards relationship are displayed showing how asymmetry may be visualised more clearly This relationship is set up to mimic a scenario whereby there is an apparent treatment effect but that it does not become immediately obvious. This type of situation is not uncommon in medical studies whereby an experimental treatment may have a delayed effect with little or no improvement being observed at the early stages.

To evaluate model performance, the recommendations of Burton[82] are followed. Defining the true parameters to be estimated as θ, estimated parameters, $\hat{\theta}$, are evaluated using measures of bias, accuracy and coverage as defined in Table 4.5.

A total of 250 datasets are simulated for two sample sizes ($n = 100, 250$) where n is the total number of patients in each dataset. Censoring was included by selecting 10%

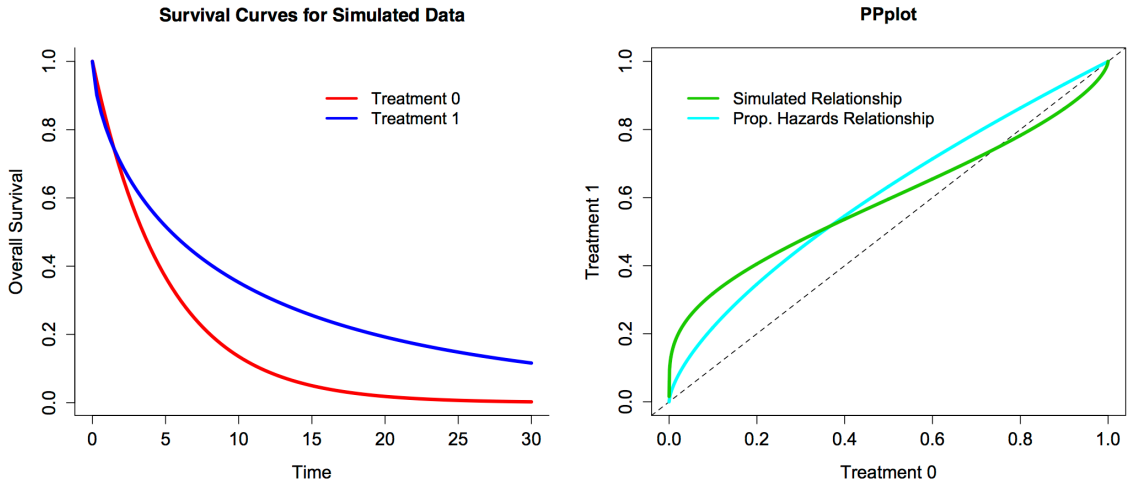

Figure 4.7: Figure to show the behaviour of the proportional hazards models with the inclusion of asymmetry parameter

Evaluation criteria	Formula
Bias	$\bar{\hat{\theta}} - \theta$
Accuracy	$(\bar{\hat{\theta}} - \theta)^2 + (SE(\hat{\theta}))^2$
Coverage	Proportion of times 95% CI for $\hat{\theta}$ includes θ
ACIL	$\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} 3.92 \times SE(\hat{\theta}_i)$

Table 4.5: Model performance parameters: θ is the true parameter estimate, For N_{sim} individual simulations, $\hat{\theta}_i$ is the parameter estimate with associated standard error $SE(\hat{\theta}_i)$. Based on N_{sim} the standard error of $\hat{\theta}$ over all simulations is $SE(\hat{\theta})$. $\bar{\hat{\theta}} = \sum_{i=1}^{N_{sim}} \frac{\hat{\theta}_i}{N_{sim}}$. ACIL: average confidence interval length.

observations at random to be censored. To each dataset the proposed non-proportional model is fit as well as a proportional hazards model for reference. The hazard ratio obtained by fitting a proportional hazards model here is of little general interest but does give some illustration of the effect on parameter estimates of fitting an incorrect model.

The results of the simulation study are given in Table 4.6. Considering initially the parameter estimates from a non-proportional model, a level of bias that may reasonably be expected to occur by chance and the coverage associated with the 95% confidence interval is what may be expected. There is little change in the bias due to the increase but the accuracy and ACIL both decrease as is to be expected.

Comparing the proportional hazards estimates of β to the non-proportional hazards estimates, there is a larger estimate of bias as a consequence of a mis-specified model

and greater uncertainty over the model estimates shown by larger standard error and ACIL estimates. This suggests that mis-specifying a model may result in error in both the point estimate and the confidence interval of the point estimate. In this instance the bias suggests a consistent upwards bias but this may not necessarily be the case for other scenarios. The shift in the point estimate is also demonstrated by the consistent drop in the Coverage for the model that assumes proportional hazards

Sam. Size	Model	Parameters	Mean (SE)	Bias	Acc.	Cov.	ACIL
N = 100	Prop.	$\beta = 0.69$	0.73 (0.25)	0.03	0.06	0.93	0.87
	Non-Prop.	$\beta = 0.69$	0.69 (0.22)	0.00	0.05	0.96	0.80
		$\alpha = -0.28$	-0.29 (0.19)	-0.01	0.04	0.96	0.75
N = 250	Prop.	$\beta = 0.69$	0.73 (0.15)	0.04	0.02	0.91	0.54
	Non-Prop.	$\beta = 0.69$	0.70 (0.12)	0.01	0.02	0.93	0.49
		$\alpha = -0.28$	-0.30 (0.11)	-0.01	0.01	0.96	0.46

Table 4.6: Results of the simulation study for survival data on a hazards scale. Results are given in terms of overall estimates of the mean and standard error, along with estimates of bias, accuracy, coverage and average length of the 95% confidence interval.

The simulation study shows that parameters can be constantly and reliably estimated using standard maximisation techniques. It also demonstrates, that analysing assuming proportional hazards can lead to inflated estimates in terms of both bias and standard error.

4.6.2 Odds models

As with the simulation study for the hazards function, the purpose of the simulation study here is to demonstrate that parameters can be reliably estimated for the proportional odds model. The methods for simulating data are detailed in the Section 4.6.1. Survival times for the control arm are simulated using (4.6) and survival times for the experimental arm are simulated using

$$t_E = \lambda^{-1} \log \left[\exp \left\{ \frac{\log(1 - S_E(t)) - \log(S_E(t)) - \log(\phi)}{\alpha} \right\} + 1 \right].$$

Data are simulated using the same parameters as for the hazards models ($\beta = 0.69$, $\alpha = -0.28$). Figure 4.8 shows the behaviour of the proportional odds function in both a Kaplan Meier and PP-plot. On the PP plot both the proportional and non-proportional odds models are shown for comparison. It may be observed here that the difference between the proportional and non-proportional models is less obvious for an odds relationship than it is for a hazards relationship.

The results from the simulation study are included in Table 4.6 and some notable difference are observed to the use for the hazards models. Here whilst there is again some evidence of a positive bias for the proportional odds model, there is also some evi-

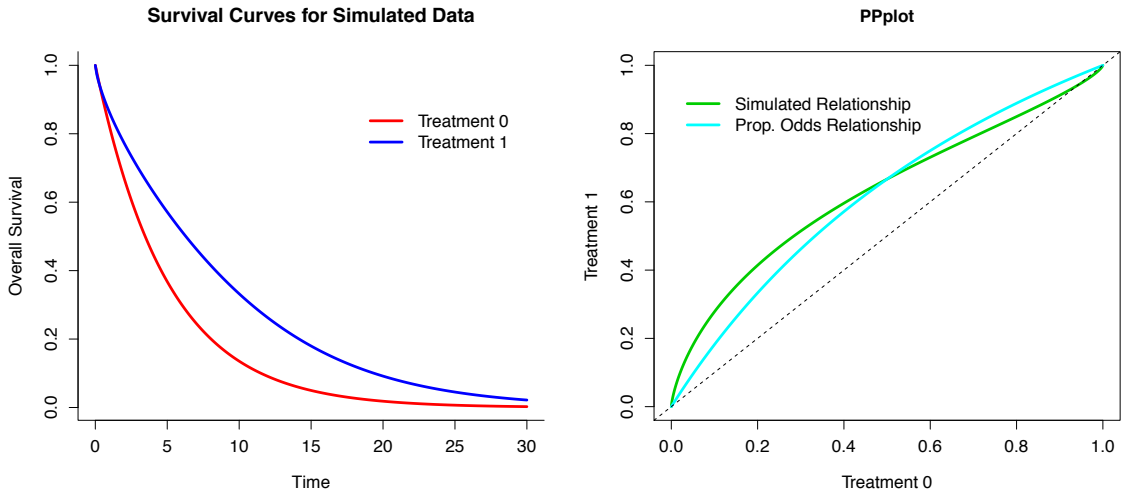

Figure 4.8: Figure to show the behaviour of the proportional odds models with the inclusion of an asymmetry parameter

dence of consistent negative bias for the non-proportional odds model. The asymmetry parameter is consistently measured with a level of bias which is considered acceptable.

Regarding coverage, for the models with a sample size of $N = 100$ then the coverage is what might be expected. For sample sizes of $N = 250$ however, coverage rates greater than what might be expected are observed for β in both the proportional and non-proportional models. The reduction in the ACIL for the non-proportional parameter β compared to the proportional models is evidence of the greater precision for this parameter under the non-proportional model.

Sam. Size	Model	Parameters	Mean (SE)	Bias	Acc.	Cov.	ACIL
N = 100	Prop.	$\beta = 0.69$	0.76 (0.39)	0.07	0.15	0.94	1.42
	Non-Prop.	$\beta = 0.69$	0.64 (0.34)	-0.05	0.12	0.93	1.29
		$\alpha = -0.28$	-0.31 (0.18)	-0.03	0.03	0.94	0.75
N = 250	Prop.	$\beta = 0.69$	0.74 (0.16)	0.05	0.03	1.00	0.89
	Non-Prop.	$\beta = 0.69$	0.63 (0.14)	-0.07	0.03	0.97	0.81
		$\alpha = -0.28$	-0.30 (0.13)	-0.01	0.02	0.93	0.46

Table 4.7: Results of the simulation study on an odds scale. Results are given in terms of overall estimates of the mean and standard error, along with estimates of bias, accuracy, coverage and average length of the 95% confidence interval.

4.7 Application to cancer trials

In this section the asymmetry parameter is applied to both a univariate dataset using data from a clinical trial for gastric cancer and a multivariable dataset using data from the ESPAC-3 trial for pancreatic cancer.

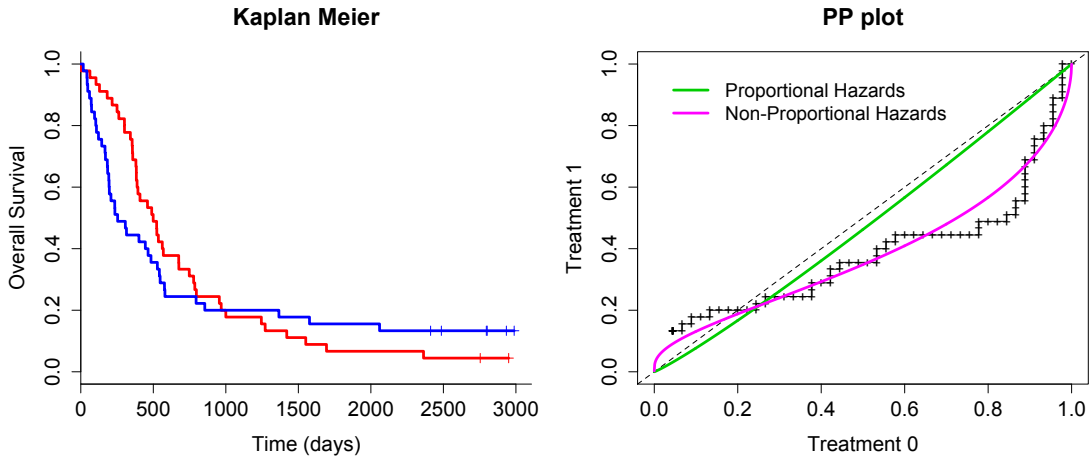

Figure 4.9: Figure to show the fit of a standard Cox model and a model with an included asymmetry parameter

4.7.1 Gastric cancer dataset

The use of the asymmetry parameters is illustrated by applying the method to a dataset of 90 patients taken from the gastric cancer dataset as described in Section 1.4.2. Figure 4.9 illustrates both the survival probabilities as calculated via the method of Kaplan and Meier as well as the PP-plot; both clearly show that the assumption of proportional hazards is not appropriate.

Four models are fitted to the data, both the hazards and odds models with and without the asymmetry parameter. Results are presented in Table 4.8 and illustrate that for both the proportional hazards and the proportional odds models, allowing for the inclusion of an asymmetry parameter improves the fit of the model as is shown by the change in the model likelihoods. The best model in terms of the likelihood is the non-proportional hazards model. The results of the proportional hazards and the non-proportional hazards models are both illustrated as fitted models in the PP-plot in Figure 4.7. Here on the PP-plot, the results of the Cox proportional hazards model (green line) and the model with an additional asymmetry parameter (purple line) are included. The improved fit of the model including the asymmetry parameter is clearly shown. Of interest is the effect on the dispersion parameter $\hat{\beta}$ for the proportional and non-proportional hazard models. This parameter gives a measure of overall dispersion away from the null line of no difference between two levels of a covariate. Here the estimate (standard error) of $\hat{\beta}$ increases from 0.11(0.22) to 0.25(0.24) showing an increase in the divergence favouring chemoradiotherapy over chemotherapy.

Further, note that although the asymmetry parameter acts in the same fashion to the hazards and the odds models, the interpretation of the model parameters does differ. Here for example, for the odds model, there is little difference in the divergence

between the two models with and without the asymmetry parameter with estimates of $\hat{\beta} = 0.75(0.38)$ and $\hat{\beta} = 0.75(0.39)$ respectively. Here then, whilst including the asymmetry parameter does improve the fit of the model, it does not have any substantial effect on the divergence due to $\hat{\beta}$.

Clinical interpretation of the data depends on the model that is chosen. Here, assuming a hazards relationship is preferable due to the improved model fit, the asymmetry parameter is itself significant at the 5% level based on a Wald test giving evidence of non-proportionality. That significance of the asymmetry parameter is sufficient evidence to show that the two survival distributions are different. This can be further verified by a comparison of models via a likelihood ratio test.

That $\hat{\beta}$ is not significant at a 5% level here shows that despite the increased magnitude, there is still not enough evidence to show that the divergence away from the null line is important which may be expected due to the crossing survival curves. Note however, that the importance of the divergence parameter is dependent upon the assumed relationship between treatments (hazards or odds).

Model	Param.	coef	exp(coef)	se(coef)	Z	P-Value	Likelihood
Prop. Hazards	$\hat{\beta}$	0.11	1.11	0.22	0.48	0.63	389.43
Non-Prop.Hazards	$\hat{\beta}$	0.25	1.29	0.24	1.05	0.29	384.52
	$\hat{\alpha}$	-0.60	0.55	0.19	-3.14	<0.001	
Prop. Odds	$\hat{\beta}$	0.75	2.13	0.38	2.00	0.05	388.67
Non-Prop. Odds	$\hat{\beta}$	0.75	2.11	0.39	1.92	0.05	385.78
	$\hat{\alpha}$	-0.47	0.63	0.19	-2.41	0.02	

Table 4.8: Results of applying the asymmetry parameter to the gastric cancer dataset under the special cases of the proportional hazards and proportional odds models

4.7.2 ESPAC-3 data

The asymmetry parameter is further explored with respect to data taken from the ESPAC-3. Of particular interest here are a group of patients who had tumours classified as either 'Ampullary' or 'Other'. Previous analyses of these data have been published [34] and have shown that the addition of chemotherapy offers improved survival over an observation arm. Here a complete case analysis is carried out removing patients with missing data.

Analysis is based on 427 patients. Resection Margin (Negative vs. Positive) and Chemotherapy covariates are 'forced' into the model as a stratification factor in the trial design and the key covariate of interest respectively. Also included are terms for Tumour Type (Ampullary vs. Other), Lymph Nodes (Negative vs. Positive) and Tumour Differentiation (Well vs. Moderate vs. Poor). Taking the standard propor-

tional hazards model, the assumption of proportionality is investigated via inspection of Schoenfeld's residuals [83] and tested via the method of Grambsh and Therneau [85]. The results are given in Table 4.8. Inspection of the residuals show that it may not be appropriate to assume proportional hazards with respect to the Tumour Differentiation covariate. In particular the 'Poor' level of this covariate has significant (P=0.02) evidence of non-proportionality. To account for this, the model is refitted including the asymmetry parameter for this variable only.

Categorical Variable (Level)	Correlation	χ^2_1	P-value
Resection Margins (Positive)	-0.02	0.12	0.73
Treatment Arm (Gem)	0.02	0.06	0.81
Treatment Arm (Obs)	-0.00	0.00	0.98
Tumour Type (Other)	0.00	0.00	0.98
Lymph Nodes (Positive)	0.07	1.39	0.24
Tumour Diff. (Moderate)	-0.07	1.15	0.28
Tumour Diff. (Poor)	-0.15	5.50	0.02

Table 4.9: Results of assessing each variable for proportional hazards via the Schoenfeld residuals

The results of both the standard proportional hazards model and the model that includes the asymmetry parameter are included in Table 4.9. Here it is shown that the inclusion of the asymmetry parameter for the 'Moderate' level is not significant (P=0.356) but that it is for the 'Poor' level (P=0.035). This is encouraging as it agrees with what is observed in the analysis of Schoenfeld' residuals in Table 4.8. It is further observed that for the 'Poor' level of Tumour Differentiation, the effect is to change the estimate (standard error) given for β from 0.745 (0.241) to 0.656 (0.264). The results show that having a poorly differentiated tumour will increase the overall hazard and that this divergence increases as survival probabilities decrease. Again there is little change in the effect of β for the 'Moderate' level of this covariate.

It is important to note that as the asymmetry parameter is included only for the Tumour differentiation variable, the dispersion parameters for all other variables can be interpreted directly as hazard ratios. Lastly, it is valuable to note that by altering the modelling approach for Tumour Differentiation, there is also some effect on other terms due to confounding. Although, alterations are small, there are reductions in the (log) hazard ratios for all the terms apart from Resection Margins for which there is a small increase. There are no detectable differences in the standard errors for parameters that are modelled without the inclusion of an asymmetry parameters.

4.8 Discussion

In this chapter an investigation into the modelling of survival data was carried out when the assumption of proportional hazards is not met. A review was provided on both

			Proportional Hazards			Non-Proportional Hazards		
Factor	Levels	Parm.	coef	se(coef)	P. Val.	coef	se(coef)	P. Val.
Resection Margins	Negative							
	Positive	β_{Rpos}	-0.005	0.174	0.977	0.014	0.174	0.935
Chemotherapy	5FU							
	GEM	β_{GEM}	-0.148	0.161	0.358	-0.137	0.161	0.397
	Obs.	β_{Obs}	0.262	0.157	0.095	0.257	0.157	0.101
Tumour Type	Amp.							
	Other	β_{Oth}	0.675	0.143	<0.001	0.657	0.143	<0.001
Lymph Nodes	Negative							
	Positive	β_{Lpos}	0.932	0.150	<0.001	0.928	0.150	<0.001
Tumour Diff.	Well							
	Mod.	β_{Mod}	0.375	0.223	0.092	0.366	0.228	0.109
		α_{Mod}				-0.182	0.198	0.356
	Poor	β_{Poor}	0.745	0.241	0.002	0.656	0.264	0.013
		α_{Poor}				-0.441	0.209	0.035

Table 4.10: Results of applying the asymmetry parameter to the ESPAC 3 (V2) pancreatic cancer dataset

methods for detecting non-proportional hazards and of methods of modelling data when proportionality can not be assumed. Throughout, the main focus is in assessing the difference between two treatments with the aim of determining which of two treatments is superior.

Data taken from a gastric cancer dataset were used as a case study and three analysis methods in particular were used: a time-dependent covariate approach, a piecewise approach, a restricted mean survival approach. Virtues and drawbacks of each approach were discussed.

Following this, an approach for modelling under non-proportional hazards was introduced whereby an asymmetry parameter is included into the class of semi transformational models of [64] and [65], based on the counting process notation of [40]. In this setting, the asymmetry parameters offer a wide variety of models and allow for the direct of modelling of non-proportional hazards without introducing any explicit dependence on time. Furthermore, it allows for the overall comparison of two treatments via a dispersion parameter which can be interpreted as a hazard ratio when proportionality can be assumed.

The use of an asymmetry parameter does come with some loss of interpretation when compared to the standard proportional models. A particular attraction of the Cox model is that the hazard function for a patient with covariates, z, can be expressed as a multiplicative form of the baseline hazard function. It is argued here however that when the assumption of proportionality is not met, this interpretation is of little practical use and in severe cases, as with the gastric cancer dataset, an average hazard ratio may be misleading in suggesting little difference between two treatment arms.

Throughout, the use of PP-plots as proposed by [123] are incorporated as an improved means of visualising the modelled relationship between two levels of a categorical

variable. It was demonstrated that when including the asymmetry parameter, although the exponent of the parameter that acts proportionally on the baseline hazard function may no longer be referred to as a hazard ratio, it can still be interpreted as a measure that causes overall divergence away from the null line of no-difference and that this measure is independent of time. The asymmetry parameter does not directly cause divergence from the null line but accounts for the lack of symmetry in the model with exponent values > 1 accounting for larger divergence at higher survival probabilities than there are at lower probabilities and values < 1 resulting in larger values at lower survival probabilities.

The approach advocated here offers an alternative to the traditional modelling approach of accounting for departure from proportionality by introducing some dependency on time. Removing time from the model is advantageous as it removes the user from specifying the underlying nature of a models dependency on time. Further, in many situations, time may be a somewhat arbitrary construct on which to introduce a dependency. It has been shown with the use of a pancreatic dataset how model asymmetry can be easily incorporated to account for a non-proportional relationship in a multivariable setting where some variables may be associated with non-proportionality but not all.

Chapter 5

Bayesian Analysis of time-to-event data

5.1 Introduction

In this chapter some exploration of the current uses of Bayesian methods for the analysis of survival data are given. Initially a brief overview is given of Bayesian methods applied in practice. Following this, the data from the ESPAC-3 trial are analysed from a Bayesian perspective using a piecewise exponential model and some of the practical issues discussed. Lastly, an investigation for the use of differing definitions of the time-grid for the piecewise exponential model are investigated and recommendations given for future use.

5.2 The use of Bayesian methods for the analysis of time-to-event data

The analysis of survival data in a Bayesian framework presents a particular challenge due to the sometimes complex censoring patterns that can occur and the often inadequacy of parametric models to correctly model a baseline hazard function. A detailed exploration of Bayesian survival models has been extensively explored by Ibrahim et al.[129] who detail the use of both parametric and non-parametric approaches.

Bayesian non-parametric approaches to the analysis of survival data have been long established with Kalbefleish [57] and Synoms [130] both proposing methods that can be used. The use of a Bayesian piecewise exponential model was given particular attention by Gammerman [49] who developed a dynamic approach to ensure a smoothed hazard function is obtained.

As the capabilities for computation expanded with technology, so did the advances in Bayesian survival analysis with Arjas and Gasberra [131] exploring the uses of the MCMC sampler to fit survival models. The use of Bayesian survival models were further expanded to handle frailty terms [132, 133] piecewise frailty terms [134], and survival

fractions [135, 113]. In the practical situation, Gustafson [136] explored Bayesian survival methods within a multicentre trial.

A direct comparison of frequenstist and Bayesian survival methods was carried out by Gomez et al. [137], whilst a general overview of Bayesian proportional hazards was provided by [138]. The formulation of survival models as a counting process allowed for the exploration of more complex survival models with differing functional forms [65] and censoring patterns [53]. More recently, Bayesian survival models have been expanded into joint modelling with longitudinal data [108], individual patient data meta analysis [56] and semiparametric modelling [139].

The effort to apply Bayesian methods in a practical setting has also increased. Most notably Berry et al. [140] consider Bayesian approaches for analysing non-proportional hazards whilst in a more general sense, He et al. [141] consider Bayesian survival models for direct use in medicine and Omurlu [142] considers a direct comparison between Bayesian models and the Cox model.

5.3 Applied Bayesian analysis of ESPAC-3 data

In this section a Bayesian survival model is fitted to the ESPAC-3 data. In particular the piecewise exponential model (PEM) as detailed in Chapter 2 is utilised. The full dataset is analysed to replicate the final analysis of a clinical trial.

The PEM is chosen as it offers a more flexible approach than standard parametric models whilst methods such as the counting process model [39] are not considered as the large number of parameters required for estimation result in an unfeasible computational burden.

Following Bayes theorem, the posterior distribution is given as

$$Pr(\theta|D) \propto Pr(D|\lambda,\beta)Pr(\lambda)Pr(\beta).$$

Note that this assumes a-priori independence of both λ and β, i.e. $Pr(\beta|\lambda) = Pr(\beta)$. The quantity $Pr(D|\lambda,\beta)$ is the likelihood function for the PEM dependent on the baseline hazard parameters λ and the log hazard ratio β given by

$$Pr(D|\lambda,\beta) = \prod_{i=1}^{N}\prod_{j=1}^{J}\left(\lambda_j \exp(\beta \boldsymbol{z})\right)^{\delta_{i,j}\nu_i} \exp\Big\{ -\delta_{i,j}\Big[\lambda_j(t_i - a_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(a_g - a_{g-1})\Big]\exp(\beta \boldsymbol{z})\Big\}, \quad (5.1)$$

where t_i is the event time for observations $i = 1, 2, .., N$. Throughout this thesis, the hazard rate is presented on the log scale, $\gamma = \log\lambda$. Prior distributions for both $\boldsymbol{\gamma}$

and β are based on normal distributions. That is

$$Pr(\beta_k) \propto \exp\left\{ -\frac{(\beta_k - \mu_{\beta_k})^2}{2\sigma^2_{\beta_k}} \right\}$$

and

$$Pr(\gamma_j) \propto \exp\left\{ -\frac{(\gamma_j - \mu_{\gamma_j})^2}{2\sigma^2_{\gamma_j}} \right\}.$$

For the prior distributions, let μ_{γ_j} and σ_{γ_j} be the mean and variance parameters for the prior distributions for the log hazard rate parameters and μ_{β_k} and σ_{β_k} the prior parameters for the log hazard ratios. Note that for practical purposes set $\mu_{\gamma_j} = \mu_{\beta_k} = 0$ and $\sigma_{\gamma_j} = \sigma_{\beta_k} = 1000, \quad \forall(j,k)$ which are considered vague, uninformative prior distributions for data analysis.

The full posterior distribution for this model is clearly complex and so models are applied in a Bayesian framework making use of the MCMC procedure described in Chapter 2. MCMC routines are produced in R with code available in the Appendix. Here, following initialisation, Normal jumping kernals with mean equal to the current value of each parameter and variance set to obtain an efficient algorithm are used as the jumping distribution for the MCMC procedure (see Gelman et al. [70] for details).

It is noted from initial models that there is a large amount of correlation between the successive log baseline hazard parameters and that this correlation can lead to biased estimates. MCMC routines can be improved by searching for orthogonal transformations of the model parameters. Here however, a batch sampling technique is applied where, as opposed to sampling from each baseline hazard parameter individually, they are sampled collectively as a 'batch'. That is, for each iteration, a sample from the conditional distribution

$$Pr(\boldsymbol{\lambda}|\beta_1, ..., \beta_k, D),$$

is taken. Samples from the conditional distributions for β are obtained in the normal fashion.

The full model is applied following the published analysis by Neoptolemos et al. [32] with the added inclusion of a variable to account for Diabetic status which has since been shown to be of some importance (unpublished). Cancer Antigen 19.9 (CA19.9) is not included in this analysis due to a large number of missing values. For all factor variables of interest, patients with missing values are included as an extra category. For the continuous measurements, tumour size, a dummy covariate is included to account for missing values, the log transformed Tumour Size for each patient is then included as a nested covariate. This approach assumes that missing data are missing completely at random. Lastly, as a piecewise exponential model has been chosen, a time-grid is required. Here a time-grid is somewhat arbitrarily set as $a = (0, 6, 12, 24, 48, 72, 96, t^*)$ where t^* represent the maximum observed survival time in the data.

Initially 10,000 samples are drawn from 3 chains to assess model convergence. Figure 5.1 illustrates the model fit of each chain via a history plot for both the first log hazard rate parameter, γ_1 and the log hazard ratio β_{Arm} for the treatment identifier. Here it is shown that for γ_1, whilst the chains have converged to be drawing from the same density, there is a large amount of auto-correlation. Auto-correlation here refers to the correlation between successive measurements in the same chain for the same parameter within the MCMC routine and is not confused with the correlation between parameters which was the motivation for using a batch sampling routine. The auto-correlation is shown in Figure 5.1 by the chains for λ_1 not directly fitting over one another and is confirmed by the associated autocorrelation plot. The chains for β_{Arm} however show a much smaller degree of correlation and give the desired 'fat caterpillar'.

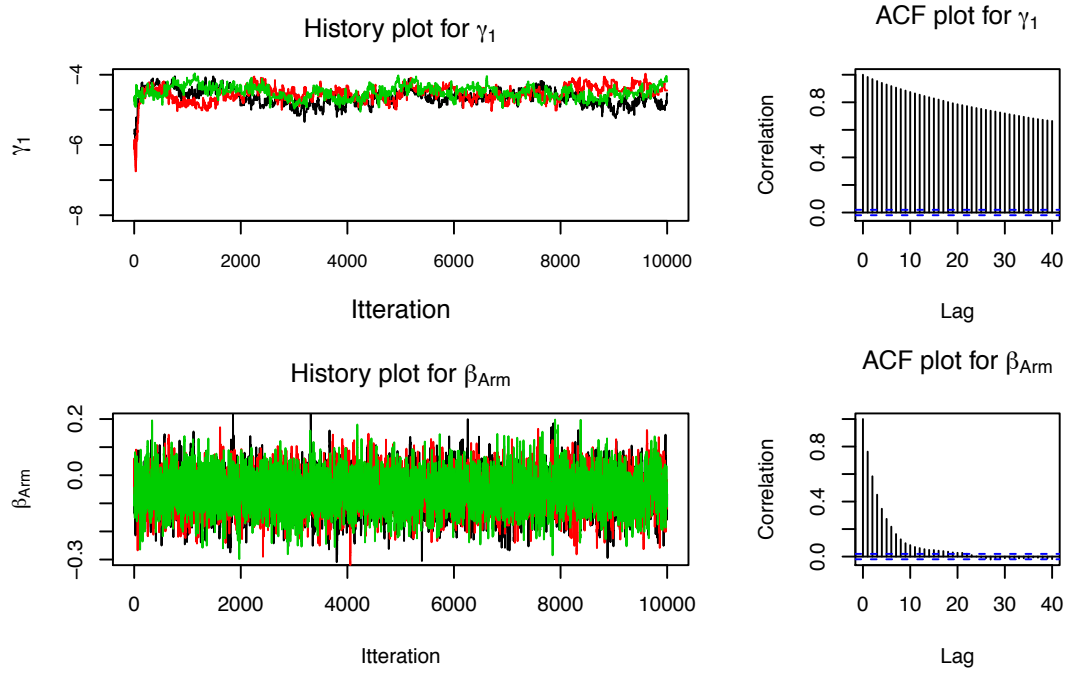

Figure 5.1: History and Autocorrelation plots for γ_1 and β

As the main parameter of interest is the log hazard ratio, model inferences may still be made on what has been drawn. It is somewhat prudent however to account for any autocorrelation as it is observed. In this instance applying a thin of 100, and only recording the 100th sample of each draw from the posterior distribution, can account for the observed auto-correlation. A further 50,000 samples are then drawn from the posterior distribution. The results, again for γ_1 and β_{Arm} are shown in Figure 5.2. Here it is shown that for γ_1 the three chains fit on top of each other . The auto-correlation plots also show that whilst not completely removed, the amount of correlation between successive recorded draws is greatly reduced. Considering β_{Arm}, all evidence of any correlation whatsoever has been completely removed and each successive draw from

the posterior densities can be considered as independent.

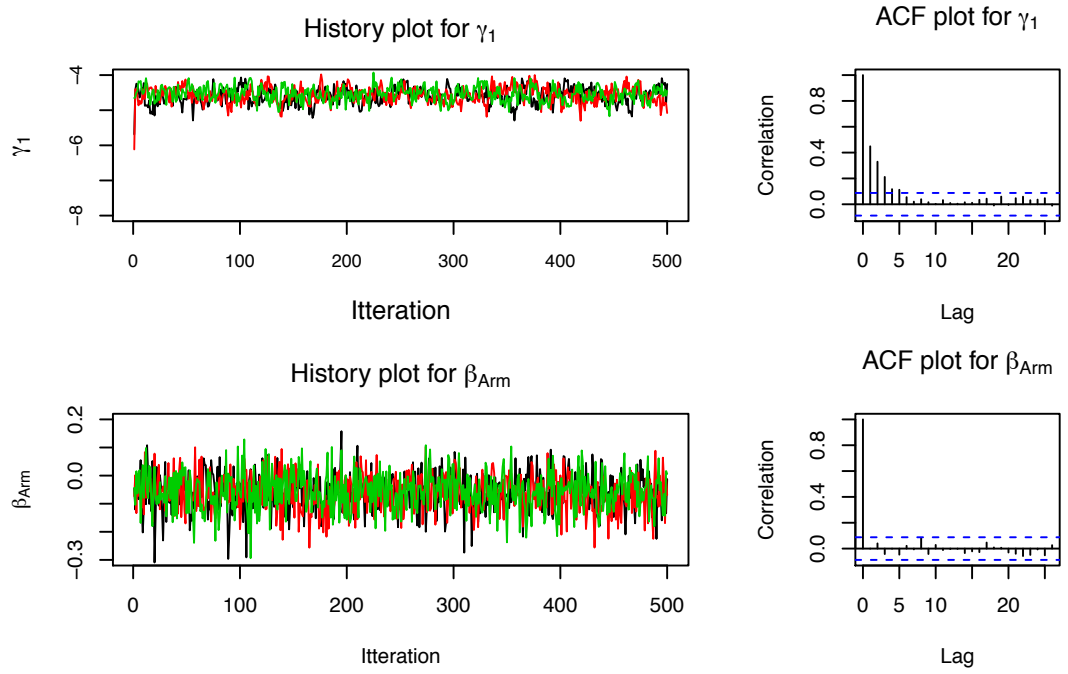

Figure 5.2: History and Autocorrelation plots for γ_1 and β with a thin of 100

Having ensured that each chain is drawing from the posterior distribution and having accounted for the auto-correlation, simulation draws can be used to make inferences about the model. In this instance interest lies in summarising the posterior densities of the log hazard ratios. Model results are presented in terms of parameter means (standard deviations) and associated 95% credibility intervals.

Model results are included in Table 5.1. Here the results of the log baseline hazard parameters are included for reference along with the standard regression parameters. Here a parameter is considered as being important if the 95% credibility doesn't include zero. For example, having positive resection margins increases the hazard compared to negative resection margins (log hazard rate = 0.229, 95% CI = (0.088, 0.366)). There is however no evidence of the Treatment effect being important in explaining overall survival (log hazard rate = -0.059, 95% CI = (-0.192,0.073)).

A further advantage of the Bayesian approach is that the posterior random variables can be transformed to create structures of interest. As a straight forward example, with posterior draws of β_{Arm} (which is denoted $\tilde{\beta}_{Arm}$), the hazard ratio is easily obtained by calculating $\exp\{\tilde{\beta}\}$. Here the point estimate, standard deviation and 95% credibility interval on the exponential scale without any further work necessary. This approach also allows for the calculation of an estimated survival function using the baseline hazard parameters. Given the survival likelihood for the piecewise exponential model defined in Chapter 2, define the survival function as

		Mean (Std. err)	95% Cred. int.
Baseline Hazard			
γ_1		-4.560 (0.214)	(-4.984, -4.146)
γ_2		-3.533 (0.198)	(-3.914, -3.152)
γ_3		-3.286 (0.192)	(-3.659, -2.908)
γ_4		-3.512 (0.196)	(-3.895, -3.125)
γ_5		-4.105 (0.230)	(-4.564, -3.651)
γ_6		-4.745 (0.339)	(-5.453, -4.118)
γ_7		-5.787 (0.815)	(-7.521, -4.427)
Diagnostic Factor	Factor Level		
Resection Margin	Negative		
	Positive	0.229 (0.071)	(0.088, 0.366)
Treatment Arm	5FU		
	Gemcitabine	-0.059 (0.068)	(-0.192, 0.073)
Lymph Nodes	Negative		
	Positive	0.595 (0.083)	(0.435, 0.759)
WHO perf. Status	0		
	1	0.196 (0.075)	(0.047, 0.343)
	2	0.319 (0.117)	(0.091, 0.539)
Tumour Diff.	Well		
	Moderate	0.127 (0.105)	(-0.077, 0.337)
	Poor	0.435 (0.117)	(0.210, 0.665)
Smoking Status	Never		
	Past	0.101 (0.079)	(-0.055, 0.251)
	Present	0.257 (0.101)	(0.055, 0.452)
	Missing	0.210 (0.137)	(-0.058, 0.477)
Diabetic Status	Non Diabetic		
	Diabetic	0.230 (0.08)	(0.072, 0.386)
	Missing	-0.086 (0.232)	(-0.557, 0.353)
Tumour Size	Dummy ind.	-0.326 (0.295)	(-0.915, 0.253)
	log (Tum. Size)	0.182 (0.07)	(0.049, 0.321)

Table 5.1: Summaries of the posterior distributions for all parameters of the piecewise exponential model fitted to the ESPAC-3 data. Summaries are presented in the form of Means (Std. errors) and 95% Credibility Intervals

$$S(t_i, \theta) = \exp\left\{ -\left[\lambda_j(t_i - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(a_g - a_{g-1})\right]\right\}. \tag{5.2}$$

From this survival function each set of draws from the MCMC routine is used to define an estimate of the survival function using the observed survival times t_i from the data. Repeating this process for all draws a posterior estimate of the baseline survival function is obtained. Figure 5.3 illustrates the baseline survival function defined by the posterior densities of γ_j as well as fitted survival functions for patients with negative and positive levels of the Lymph Node variable. Lymph nodes are chosen here as the binary covariate with the greatest divergence and therefore best suited to illustrating

the difference in survival functions.

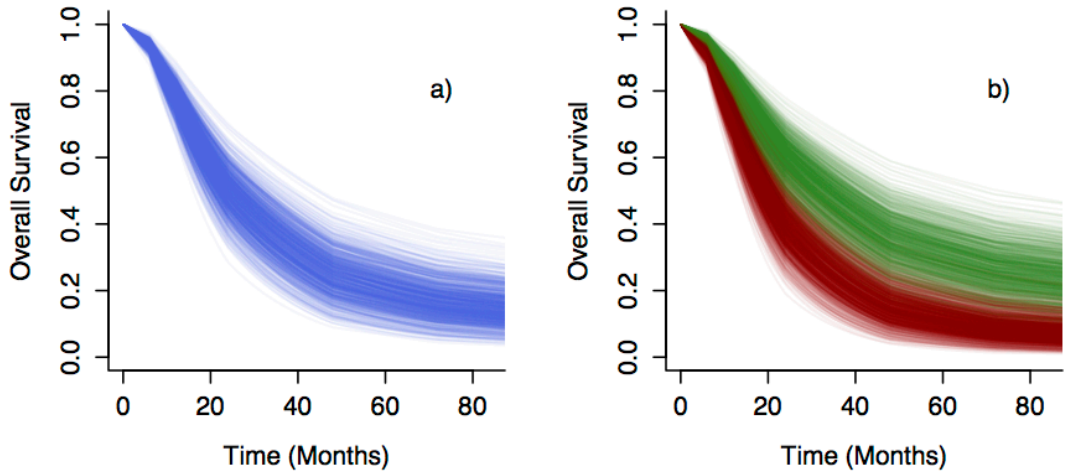

Figure 5.3: Illustration of the survival functions obtained from iteration of the MCMC sample for a) all patients and b) patients with negative (green lines) and positive (red lines) levels of the Lymph Node status variable

This shows how a Bayesian approach can be used to answer more specific clinical questions of interest. Clinicians may, for example, be interested in the probability that a patient will survival beyond 24 months and how this probability changes based on their Lymph Node status. Using (5.1) set $t = 24$ and use the posterior densities to provide a density giving the probability that a patient within the trial is alive at 24 months following randomisation. Figure 5.4 shows these densities for all patients and separated by Lymph Node status. Here it is shown that the overall probability of surviving up to 2 years for all patients is 0.504 (95% CI = (0.370, 0.624)). For patients with negative Lymph Nodes this is 0.601 (95% CI = (0.474, 0.708)) whereas for patients with positive Lymph Nodes it becomes 0.396 (95% CI = (0.265, 0.524)). This shows how patients with positive Lymph Nodes have a poorer prognosis compared to negative Lymph Node patients and allows for model interpretations in a fashion that is acceptable to both clinicians and patients.

Considering a more complicated example, take a non-diabetic patient who was a previous smoker and presented with a zero WHO performance status. Following surgery, a tumour was removed which was shown to be well differentiated and of size 20mm. It was also shown that the patient had positive resection margins and positive lymph nodes. The probabilities that a patient lives up to 6 - 60 months with associated 95% credibility intervals are given in Table 5.2.

Note at this point that all summaries are based on explaining the data that have already been observed. Of further interest may be able to predict the performance of parameters should future data be collected. Summaries here are obtained via the

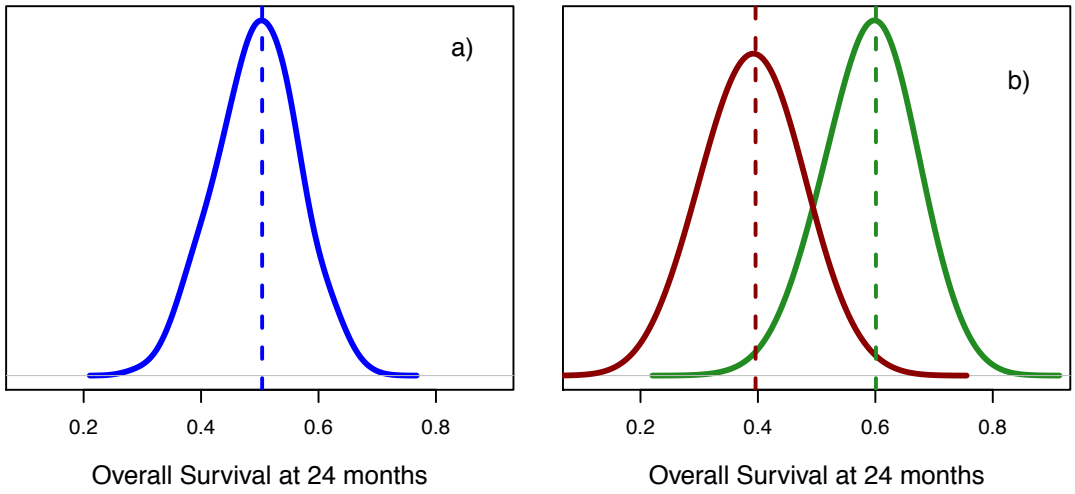

Figure 5.4: Derived Posterior densities showing the probability of patients surviving up to 24 months within the trial for a) all patients and b) patients with negative (green density) and positive (blue density) of the Lymph Node status variable

Months	Estimated Survival (95% CI)
6	0.94 (0.91, 0.95)
12	0.78 (0.72, 0.82)
24	0.49 (0.39, 0.58)
36	0.34 (0.24, 0.43)
48	0.23 (0.15, 0.32)
60	0.19 (0.11, 0.27)

Table 5.2: Table to show the estimated survival probabilities and associated 95% credibility intervals over the course of 60 months.

posterior predictive distribution. The positive predictive distribution of n future data D_n is given by.

$$\int_{\infty} Pr(D_n|\theta)Pr(\theta|D)\delta\theta.$$

Given the complex form of the posterior distribution, $Pr(\theta|D)$, this evaluation is non trivial. Approximations can be formed however by assuming a distributional form for the posterior parameter of interest. As an example, it may be of interest for clinicians to predict the performance of the log hazard ratio for the treatment arm should further patients be randomised into a clinical trial.

Assume that the posterior distribution for β_{Arm} follows a normal distribution,

$$Pr(\beta_{Arm}|D) \sim N(\mu_{Arm}, \sigma^2_{Arm}),$$

where μ_{Arm} and σ^2_{Arm} are the mean and variance of the posterior distribution for β_{Arm}. Using the formulation of [21], a predictive posterior distribution can be given by

$$Pr(D_n|\beta_{Arm}) \sim N(\mu_{Arm}, \sigma^2_{Arm} + \sigma^2_{D_n,Arm}),$$

where $\sigma^2_{D_n,Arm}$ is the variance that is expected to be observed from a future trial with n observations. Previous estimates of $\sigma^2_{D_n,Arm} = 4/m$ have been proposed [143] where m is the number of events. It is prudent to notice that this assumes an equal number of events in each treatment arm; a more accurate formulation is therefore given by $\sigma^2_{D_n,Arm} = 1/E_0 + 1/E_1$ where E_0 and E_1 are the number of events in future treatment arms. Given the parametric form for the survival function given by (5.1), a future study with N patients which is designed with a minimum follow up of t^* will have an estimated number of events in the control arm $E_0 = [NS(t^*, \theta_0)]/2$ and equivalently for E_1. An estimate of the variance of the log hazard ratio from a future study with N observations is then given by

$$\sigma^2_{D_n,Arm} = \frac{4 - 2\{S(t^*, \theta_0) + S(t^*, \theta_1)\}}{NS(t^*, \theta_0)S(t^*, \theta_1)}.$$

Figure 5.5 shows the posterior distribution for β_{Arm} as well as posterior predictive distributions for the same parameter for future trials of size 500 and 750. If it is taken that a clinically important difference is given by a log hazard ratio of $\log(0.8) = -0.22$ (illustrated) then the probability that Gemcitabine will be shown to be at least this much better than Capecitabine is 1% for the observed posterior distribution, and 18% and 14% for future datasets of 500 and 750 patients respectively.

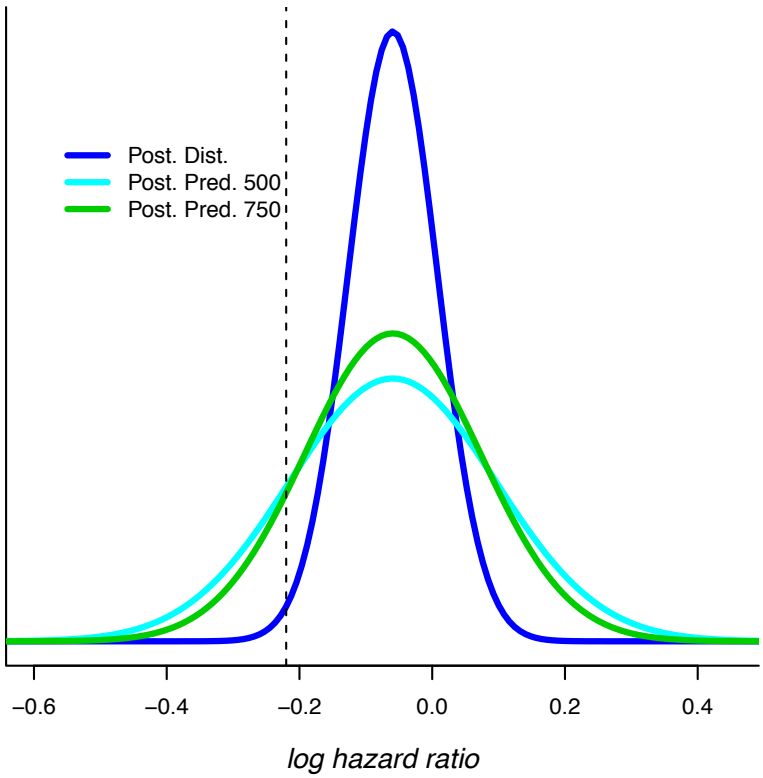

Figure 5.5: Posterior distribution for β_{Arm} and associated predictive posterior distribution for future datasets of size 500 and 750

5.4 Time-grids and the piecewise exponential model

A necessary requirement of the PEM is to set a time-grid, and despite its use in practice, there is no recommended method for how the time-grid should be set. Some discussion on the matter has been provided by Sahu et al. [132] and a review is provided by Han et al. [144].

Specific methods for use in a Bayesian framework have been proposed by Demarquie et al. [145, 146], whilst Goodman et al. [48] propose an approach for selecting change points in a piecewise model based on maximum likelihood approaches. Han et al. take a different approach, starting with a saturated model, where each unique event results in a new partition in the time grid and reducing the piecewise model to only include partitions of import. Each of these approaches may not be appropriate for all situations however and generally concentrate on the accurate estimation of a single cumulative hazard function. Interest in this chapter is primarily on the hazard ratio, with the hazard function being considered a set of nuisance parameters. In this context, little discussion is offered on the choice of the correct time-grid in the literature.

Given that the model parameters and the time-grid can not be considered to be independent of one-another, the need is highlighted for further investigation. In this section, a review of some of the techniques used in practice are presented as well as proposing some further methods. In the section that follows this is applied to a simulation study with the aim of providing recommendations for future use. Some popular methods from the literature are investigated although the method of Han et al. [144] is not considered as some preliminary exploration shows that initial investigations require the comparison of pairwise disjoint single event intervals. Evaluations on such intervals can sometime be unreliable resulting in time-grids that do not always give an appropriate fit to the hazard function. Likewise, the method as proposed by Goodman et al. [48] is altered slightly into the 'split-likelhood' method presented. This still produces time-grids in a forward step-wise fashion based on maximum likelihood methods but is less stringent than the Wald test approach suggested by Goodman et al [48].

It is assumed that the choice of time-grid is to provide as much flexibility as is required to describe the behaviour of baseline hazard function. Simultaneously, it is desirable to avoid wasting information by allowing too many partitions, this is especially true in a Bayesian framework where including extra parameters may greatly increase the computational burden.

5.4.1 Fixed time grid (Kalb.)

Kalbfleisch [57] proposed that a time-grid is set for the analysis before any of the data are observed. Whilst it requires some underlying knowledge of the data, it may be advantageous as it ensures a level of objectivity is imparted into any analysis. Conversely

there is also some risk of having empty partitions or a time-grid which provides an unsatisfactory fit to the data. In practice, a fixed time-grid may be used by observing an appropriate parametric survival or cumulative hazard functions and choosing reasonable points, or by setting time points based on convenient landmarks.

5.4.2 Fixed number of events (n.event)

Under this strategy, the user sets the number of events that are observed in each partition. The time-grid is then fully dependent on the observed data. This has the advantage that 'thinner' partitions will be observed in 'busier' areas of the distribution. However, it may cause problems as fixing a small number of events for a large dataset will result in an unnecessary large number of partitions.

5.4.3 Fixed number of intervals (n.part)

An alternative approach to fixing the number of events is to fix the number of intervals that are required, partitions are then set to occur at regular intervals. Here the model dimensionality and computation effort required can be directly managed but the possibility remains that partitions may be placed at unsuitable points.

5.4.4 Paired event partitions (paired)

The motivation behind this approach lies in the situation where a trial is set up to determine the difference between two treatment regimens. Here it may be considered beneficial to have partitions which contain events from each treatment group. Defining the observed event times for the control arm and experimental arm of a trial as t_c and t_e respectively, the time-grid is set by

1. Set $a_0 = 0$
2. Define $a_1 = max(min(t_c), min(t_e))$
3. Redefine $t_e^{-a_1}$ and $t_c^{-a_1}$ where $t_c^{-a_1}$ are the event times for the control arm with the event included in the first partition removed, and likewise for $t_e^{-a_1}$
4. Repeat steps 2 and 3 to set further partitions
5. Continue until one or both of t_c and t_e has no event times left and set $a_J = max(t_c, t_e)$

Whilst it may be advantageous to have an event from each arm to influence the estimation of the baseline hazard function, it can result in wide partitions if there are large differences between the treatment arms or a large number of partitions if the two arms are similar to one another.

5.4.5 Random time-grid (Demarqui)

The method as defined by Demarqui [145, 146] is tailored for use in a Bayesian framework. Here the-time grid itself is considered to be a random variable to be estimated. Using this approach, each event time is a candidate to define a partition in the time-grid. Demarqui then proposes the use of the posterior predictive distribution to estimate the time-grid. Whilst appealing, a drawback of this approach is it is only set up at present for models without any covariates. Evaluating posterior predictive distributions when covariates are included is more problematic. A two-step approach is taken here whereby the first step estimates the time-grid based on a model with no covariates. This time-gird is then treated as fixed and applied to the model with covariates.

5.4.6 Split likelihood partitions (split.lik)

In a similar fashion to the method as set out for the random time-grid, a time-grid is obtained by searching for a model that provides the best fit. Here, instead of posterior predictive distributions, models are evaluated based on the log likelihood as applied by the following steps

1. Start with a standard exponential model (i.e. a single hazard parameter)
2. Consider each event time as a possible partition in the time-grid and calculate the likelihood for all possible two parameter hazard models
3. Select the partition that gives the best likelihood providing this is an improvement on the single parameter model
4. Repeat this process until no further improvement can be obtained.

This approach, unlike the random time-grid method can easily incorporate covariates. Again a two-step approach whereby the time-grid is compiled in the first step and the model is evaluated in the second step.

5.5 A simulation study to compare the performance of differing time-grids

In the previous section, a series of strategies that may be used to set a time-grid were defined. In this section, each method is applied as part of a simulation study to determine which time-grid strategy may be the most appropriate. It is assumed that in fitting a model with a PEM, the main interest is in the log hazard ratios $\boldsymbol{\theta}$ and the parameters for the baseline hazard function, $\boldsymbol{\lambda}$, are treated as nuisance parameters.

5.5.1 Simulation study design

The simulation study is designed following the recommendations of Burton et al. [82]. The primary outcome of the simulation study are the estimate of the log hazard ratio parameters. These are compared against the 'true' values from which the data are simulated using measure of bias, mean square error, coverage and average confidence interval length (ACIL). Formal definitions of each of these measures are provided in Table 4.5.

To understand the full behaviour of the effects of time-grids on hazard ratios, parameter estimation is considered under a series of different scenarios as outlined by:

- censoring: 5%, 10%, 25%, 50%
- Sample Size: N = $100, 250, 500$

Data are assumed to follow a relationship similar to that of patients with advanced pancreatic cancer as displayed in Figure 5.5. There are a total of $4 \times 3 = 12$ scenarios. For each scenario, 500 datasets are simulated.

5.5.2 Simulation of data

Given that PEM models are applied under differing time-grids, simulating data from a PEM itself may produce some bias towards whatever time-grid was used in the simulation process. Furthermore, it is desirable to avoid using any standard distributions as firstly, they often do not represent the hazard functions that are observed in practice and secondly, if the data are known to follow some parametric distribution, it removes the need for a PEM altogether.

In light of this, an approach is used similar to that described by Bender [128] to simulate survival time data. Here for the 'control' group, describe a cumulative hazard function that is dependent on time but does not follow any specific distribution such that

$$S_0 = g(t).$$

Here $g(t)$ is a function of time 't' alone. The approach by Bender is to simulate survival probabilities from a uniform distribution and then derive observed survival times using the function $g(t)$. Whilst Bender uses this approach to illustrate the simulation of data from standard parametric distributions, this method can be easily adapted to handle more complex survival functions as shown by Crowther et al. [147].

Given the definition of $g(t)$, survival data are simulated on a patient level adjusting for covariates such that.

$$S_i = S_0^{\exp\{\beta^T(z_i)\}}.$$

For the purposes of this simulation study, data are simulated for patients randomised to one of two treatment arms with a log hazard ratio of $\beta_{Arm} = 0.6$. Data are simulated to ensure that half of the patients are randomised to each treatment arm. A second covariate is simulated which follows a standard normal distribution $z_{cov} \sim N(0, 1)$ and is connected to the baseline survival function via a log hazard ratio of $\beta_{cov} = 0.1$.

Given survival estimates at defined time-points, a survival function, $g(t)$ is obtained assuming a spline function and adjusted due to the appropriate covariates. Survival times for each patient are obtained using a five step process defined below. No administrative censoring is applied as part of the simulation study.

1. Simulate treatment arm and covariate information for each patient
2. For each patient, calculate survival function based on baseline priors and covariate values
3. Simulate survival probabilities from a uniform distribution
4. Derive survival times

To aid interpretation, this process is illustrated in Figure 5.6.

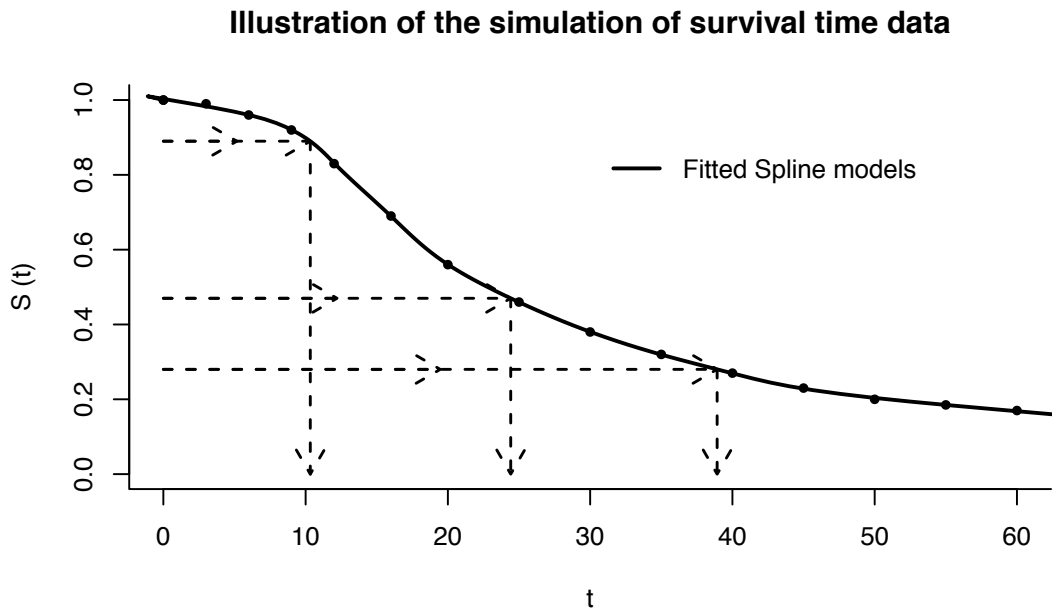

Figure 5.6: Illustration of the process of simulating survival time data using cubic splines to estimate the baseline survival function

5.5.3 Analysis of results

The results of the study are presented here in terms of model bias as a means of evaluating the point estimate and the ACIL as a means of evaluating the measure of spread of the point estimates. Further results on model accuracy and coverage are not included here for brevity.

Table 5.3 shows the results from the simulation study. For the time-grid defined by the number of events, partitions are defined by having five, ten and twenty events in each partition. In a similar fashion, time-grids defined by the number of partitions, five ten and twenty partitions are chosen.

The results in Table 5.3 show that all approaches demonstrate good levels of bias at all sample sizes. As is to be expected, the bias reduces as sample sizes increase. In terms of ACIL, again the length of the confidence interval is as to be expected. Relative inspections show that larger ACIL are observed for the paired approaches. Choosing this approach may therefore result in credibility intervals which are artificially large.

Due to the large amount of information, graphical methods are applied. Here, each set of results from each of the twelve scenarios are standardised so as to be represented on the same scale. Figure 5.7 shows the standardised absolute bias against the standardised absolute ACIL. Both the bias and ACIL are standardised by subtracting the overall mean for each summary and scaling by the observed standard deviation. Here, points that are represented towards the bottom left corner represent the best performing time-grid. This plot is useful for showing firstly the best performing time-grids and secondly consistency of different strategies across the different scenarios.

Take for example, the Demarqui time-grids (orange), these are not only the best performing, but they are consistent across different scenarios. Also good performers are the fixed time-grid approach (Kalb. - light blue) and the scenario defined by having 5 partitions (n.part (5) - dark green). Also highlighted in Figure 5.7 are the results from the Cox model (black) for reference.

In terms of consistency of performance, the approaches based on the number of events per partition and those based on at least 10 or 20 partitions (shades of red and green respectively) demonstrate the greatest spread indicating that these approaches are somewhat dependent on either the number of events or the censoring mechanism employed. The paired events (shades of blue) show clear evidence of consistently larger ACIL estimates. Lastly note that the results of including an exponential model results in consistently the largest bias although the ACIL are among the smallest observed. This illustrates how assuming the wrong relationship can result in systematically biased estimates.

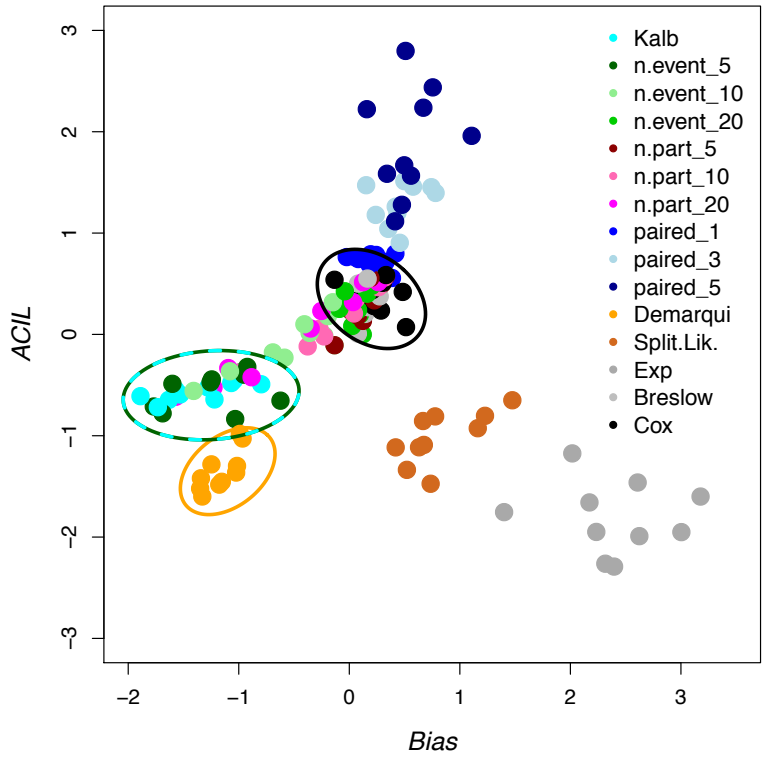

Figure 5.7: A visualisation of the simulation study results via standardised bias and ACIL estimates.

5.6 Discussion

This section investigated the use of Bayesian techniques for the analysis of time-to-event data. In a pure Bayesian framework, direct fitting of the standard Cox model is troublesome due to the difficulty of using MCMC techniques on the partial likelihood formulation in statistical packages such as WinBUGS. Furthermore, there are practical advantages in having a parametric form for the baseline hazed function that can aid in practical interpretation of any survival model.

The approach chosen as an alternative to the Cox model is the piecewise exponential model as detailed in Chapter 2. This has the advantage of being flexible enough to fit a large number of hazard functions and provides a fully parametric form for the survival function. Furthermore, the work by Laird and Oliver [59] allows that the models easily fit in a Bayesian framework.

This approach was applied to the analysis of the data from the ESPAC-3 trials. All models are applied using manual MCMC routines coded using the statistical package R and some practical issues discussed. It was shown how models such as these can provide an improved insight into patient prognosis from a clinical perspective.

A necessary requirement of the PEM is the setting of a time-grid, denoting the points

in the time domain within which the hazard rates are constant. A number of different strategies were investigated and it was shown that a time-grid that is considered itself as a random quantity to be estimated, as introduced by Demarquie et al. [145], to be the most consistent and reliable approach.

Generally however, a wide number of approaches gave consistently reliable approaches and the overall levels of bias associated with any approach were small. As a result any reasonable approach may be chosen and attention should be on limiting the computational burden - especially for large datasets. For example, choosing a fixed time-grid as proposed by Kalbfleish can result in reliable estimates and can greatly reduce the computational burden. For simplicity, this is the approach that is taken throughout the reminder of this thesis.

Sample Size	Time-Grid	Censoring 5%	10%	25%	50%
100	Kalb.	-0.01 (0.84)	0.02 (0.87)	0.01 (0.95)	0.02 (1.17)
	n.event (5)	-0.02 (0.84)	0.03 (0.88)	0.01 (0.96)	0.03 (1.18)
	n.event (10)	0.01 (0.85)	0.05 (0.89)	0.03 (0.96)	0.05 (1.20)
	n.event (20)	0.04 (0.87)	0.07 (0.89)	0.05 (0.98)	0.06 (1.20)
	n.part (5)	0.04 (0.87)	0.07 (0.89)	0.05 (0.98)	0.05 (1.20)
	n.part (10)	0.03 (0.86)	0.05 (0.89)	0.03 (0.97)	0.03 (1.18)
	n.part (20)	-0.01 (0.85)	0.01 (0.87)	0 (0.95)	-0.01 (1.16)
	paired (1)	0.04 (0.88)	0.07 (0.91)	0.06 (1.00)	0.07 (1.24)
	paired (3)	0.05 (0.9)	0.08 (0.94)	0.07 (1.03)	0.08 (1.3)
	paired (5)	0.05 (0.93)	0.09 (0.96)	0.08 (1.07)	0.08 (1.39)
	Demarqui	0.00 (0.82)	0.02 (0.84)	0.00 (0.92)	0.03 (1.13)
	Split Lik.	0.07 (0.83)	0.11 (0.86)	0.08 (0.94)	0.11 (1.16)
	Exponential	-0.16 (0.81)	-0.14 (0.83)	-0.15 (0.91)	-0.13 (1.12)
	Breslow	0.04 (0.87)	0.07 (0.90)	0.05 (0.98)	0.06 (1.20)
	Cox	0.04 (0.87)	0.07 (0.90)	0.06 (0.98)	0.08 (1.21)
250	Kalb.	0.01 (0.53)	-0.01 (0.54)	0.00 (0.60)	0.03 (0.73)
	n.event (5)	0.01 (0.53)	-0.01 (0.54)	0.02 (0.60)	0.03 (0.73)
	n.event (10)	0.03 (0.53)	0.04 (0.55)	0.05 (0.61)	0.07 (0.75)
	n.event (20)	0.07 (0.54)	0.05 (0.56)	0.05 (0.61)	0.08 (0.75)
	n.part (5)	0.07 (0.54)	0.05 (0.56)	0.06 (0.61)	0.09 (0.75)
	n.part (10)	0.07 (0.54)	0.05 (0.56)	0.06 (0.61)	0.08 (0.75)
	n.part (20)	0.06 (0.54)	0.04 (0.56)	0.04 (0.61)	0.05 (0.74)
	paired (1)	0.07 (0.55)	0.05 (0.56)	0.06 (0.62)	0.09 (0.76)
	paired (3)	0.08 (0.55)	0.06 (0.57)	0.07 (0.63)	0.10 (0.78)
	paired (5)	0.08 (0.56)	0.06 (0.58)	0.07 (0.64)	0.11 (0.79)
	Demarqui	0.03 (0.52)	0.00 (0.53)	0.01 (0.58)	0.04 (0.71)
	Split Lik.	0.08 (0.52)	0.06 (0.54)	0.08 (0.59)	0.11 (0.72)
	Exponential	-0.13 (0.51)	-0.15 (0.52)	-0.14 (0.57)	-0.12 (0.71)
	Breslow	0.07 (0.54)	0.05 (0.56)	0.06 (0.61)	0.09 (0.75)
	Cox	0.07 (0.54)	0.05 (0.56)	0.06 (0.61)	0.09 (0.76)
500	Kalb.	0.02 (0.37)	0.01 (0.38)	0.02 (0.42)	0.01 (0.52)
	n.event (5)	0.03 (0.38)	0.02 (0.39)	0.03 (0.42)	0.00 (0.51)
	n.event (10)	0.07 (0.38)	0.02 (0.38)	0.07 (0.43)	0.02 (0.52)
	n.event (20)	0.07 (0.38)	0.06 (0.39)	0.07 (0.43)	0.06 (0.53)
	n.part (5)	0.07 (0.38)	0.06 (0.39)	0.08 (0.43)	0.06 (0.53)
	n.part (10)	0.07 (0.38)	0.06 (0.39)	0.08 (0.43)	0.06 (0.53)
	n.part (20)	0.07 (0.38)	0.06 (0.39)	0.07 (0.43)	0.06 (0.53)
	paired (1)	0.07 (0.39)	0.06 (0.4)	0.07 (0.44)	0.06 (0.53)
	paired (3)	0.08 (0.39)	0.06 (0.4)	0.08 (0.44)	0.07 (0.54)
	paired (5)	0.08 (0.39)	0.07 (0.4)	0.08 (0.44)	0.07 (0.54)
	Demarqui	0.04 (0.37)	0.02 (0.38)	0.03 (0.41)	0.02 (0.5)
	Split Lik.	0.09 (0.37)	0.07 (0.38)	0.08 (0.41)	0.08 (0.51)
	Exponential	-0.13 (0.36)	-0.14 (0.37)	-0.13 (0.41)	-0.14 (0.5)
	Breslow	0.07 (0.38)	0.06 (0.39)	0.08 (0.43)	0.06 (0.53)
	Cox	0.07 (0.38)	0.07 (0.39)	0.08 (0.43)	0.07 (0.53)

Table 5.3: Table to show the results of the simulation study for a number of differing time-grids in terms of bias (ACIL)

Chapter 6

Bayesian Design of Clinical Trials with Time-to-Event Endpoints

6.1 Introduction

In this chapter, Bayesian methodology for the design of clinical trials is reviewed, paying particular attention to Bayesian sample size calculations. Initially an overview of the literature of Bayesian methods with respect to clinical trials design is given, following this Bayesian sample size methodology is reviewed and the key differences to frequentist approaches highlighted. The application of Bayesian methods to survival data is emphasised along with the importance of specifying a baseline hazard function. Finally, the ViP trial, investigating two chemotherapy treatments for patients with advanced pancreatic cancer, is re-designed from a Bayesian perspective. Note that throughout this chapter, whilst trial design is considered from a Bayesian perspective, no consideration is given to the uses of informative priors to inform trial design and analysis, this will be explored in Chapter 7.

6.2 Bayesian clinical trials

In practice, frequentist clinical trial design methods are dominant. Many funding bodies in particular will request frequentist quantities such as power and alpha levels to be set. There remains however, a large quantity of literature investigating Bayesian design methods for clinical trials. As far back as 1988 Sylvester [148] explored a Bayesian method for the design of phase II clinical trials, and these methods have begun to be used in practice.

Most notable is the MD Anderson centre, which as of 2005 had 20% of all trials designed from a Bayesian perspective [149]. More recently, two further reviews of Bayesian methodology, one with regard to medicine [18] and another more specific to the pharmaceutical industry [19] have evaluated the development of Bayesian methods over the previous 25 years.

It is clear that the decision over whether to use a frequentist or Bayesian framework (or some mix of the two) in clinical trials is a topic of some debate and given the increasing literature with respect to Bayesian trial design, Statistics in Medicine presented two publications by Berry [150] and Whitehead [1] who argue the case for each framework respectively. Here in particular, Berry argues against the 'rigidity' of the frequentist framework and notes that interpretations of trial results under a frequentist design are less direct, do not make use of all available information and are valid only under the pre-specified design conditions. The point is further made that frequentist designs are typically dependent upon two quantities, 'power' and 'clinically relevant difference' which are both chosen arbitrarily and often manipulated to '...give a sample size acceptable to investigators and sponsors'. A point reinforced by Amri and Kordestani [2] who note that the observed true magnitude of a difference is '...nearly always less than what was predicted at the time the trial was designed'. It is further noted that as a Bayesian trial can be analysed independent of any design conditions, no formal design parameters are required to begin a trial. Whitehead by contrast argues that for phase III trials in particular, a Bayesian approach should be discouraged due the possibility for subjectivity to enter the interpretation of results and that when a specific question is required to be answered, the frequentist approach can provide '...one of the most powerful tools of clinical research.'

With the growing accessibility of Bayesian techniques the Statistics in Medicine journal presented four further papers prefaced by Herson [13] with the specific aim of investigating the Bayesian analysis of cancer clinical trials. Specifically Wieand and Cha [17] introduce a trial designed under a frequentist framework to compare five treatments against a control arm for the treatment of patients with colorectal cancer. To the trial data, Dixon and Simon [14] consider using a Bayesian subset analysis to analyse the trial, whereas Freedman and Spiegelhalter [15] use Bayesian sequential stopping rules with informative priors on the hazard ratios. Lastly, Greenhouse [16] provides a review of the methods applied and shows that the overall conclusions between the two frameworks are in agreement.

Despite the dominance of frequentist methodology, interest in the Bayesian approach was clearly growing and to this end, Spiegelhalter et al. [151] provided a set of practical approaches for applying Bayesian techniques to clinical trials with Hughes [152] suggesting a set of guidelines for reporting of trials in a Bayesian framework. Following this, Spiegelhalter et al. published a book discussing the use of Bayesian methodology in medical research [21].

Still, practical uptake has been slow and consequently publications by Howard et al. [11], Gonen [10] and Moye [27] all encouraging the use of Bayesian methods. Furthermore, Berry [153, 3] argues for the use of Bayesian techniques in cancer research, citing the increased efficiency and making ethical arguments. Lastly both Perneger [5]

and Berry [4] argue that Bayesian methods are more readily interpretable by medical professionals.

The majority of the literature concerning Bayesian methodology for trial design focuses around Phase I and Phase II trials. Phase I trials concerned with dose estimation are explored by Gatsonis and Greenhouse [154] and Whitehead [155]. A trial that makes uses of an adaptive Bayesian approach to assess both efficacy and safety is considered by Berry et al. [156], whereas Fan et al. [157] consider Bayesian Phase I trials designed from a decision theoretic perspective. Chevret [158] considers Phase I trials using a continual reassessment method, a unique feature of Bayesian designs which allows for the trial to be assessed after each individual patient is evaluated.

For Bayesian techniques following on from phase I trials, Thall and Estey [159] consider an approach for screening treatments to determine which is the most appropriate to carry through to a Phase II trial. For Phase II designs, both Tan and Machin [160] and Lewis and Berry [161] consider group sequential designs. Both Bandyopadhyay et al. [162] and Zhao et al. [163] consider two stage designs with a survival endpoint. Johnson and Cook consider the continual reassessment method in a Phase II setting whereas Parmar et al. [164] consider the monitoring of clinical trials using 'enthusiastic' and 'sceptical' priors with applications. Resnic [165] focuses on Bayesian methodology to monitor trial safety.

Adaptive designs, whereby the trial design is altered based on accumulated results have been explored by Berry [166, 167] and Yin et al. [168]. The latter of these use a predictive probability approach whereby the predicted results of the trial are estimated and used to inform the randomisation procedure during the trials progress. These are further utilised by Inoue et al. [169] who propose a method for seamlessly expanding from Phase II to Phase III, by Lee and Liu [170] who show that a predictive probability approach can be more efficient than frequentist approaches and by Hong and Shi [171] who use the approach to aid the decision to progress to a Phase III trial.

6.3 Bayesian sample size calculation

In this section, the theory behind Bayesian sample size calculations is considered. Previous reviews of which, and comparisons to a frequentist framework, have been given by Addock [172] and more recently Inoue et al. [173]. Please note that throughout this chapter, the traditional notation for Type I and Type II error rates as α and β respectively are used. These are not to be confused with the intensity function and log hazard rate parameters used in previous and forthcoming chapters.

Sample size calculations for specific data types have been previously explored, see M'Lan et al. [174] for binomial responses and Joseph and Belisle [175] for normal responses. Both Gould [176] and Gubbiotti and De Santi [177] consider the design of equivalence studies. Typically however, due to the added complexity in the Bayesian

framework, there are few analytical solutions to the Bayesian sample size problem. As a result simulation techniques are generally required, an approach for which is well defined by Wang and Gelfand [178] and Rubin and Stern [179]. These make use of utility functions on the posterior distribution of interest such as those described by Lindley [180] and Pham-Gia [181].

Under the typical design of a Phase II/III randomised clinical trial, the aim in a frequentist framework is to control the Type I and Type II error rates. These design parameters are defined conditional on some fixed minimum clinically relevant difference, denoted δ. A difference that must be observed for a trial to be determined a success.

Using the notation of Chow et al. [182] denote the trial outcome as positive (negative) using $C = +$ ($C = -$) and the 'true' outcome as $T = +$ ($T = -$). In a frequentist framework the definitions for α and β are given by

$$\begin{aligned}\alpha =& Pr(C = +|T = -)\\ \beta =& Pr(C = -|T = +).\end{aligned}$$

A frequentist trial will then control designed parameters conditional on the 'true' trial outcome which can never be known. Further criticisms of the frequentist approach to be noted are:

- Typically frequentist sample size calculations will be based on some estimate of the standard deviation for the key parameter of interest. This parameter is often treated as known when it rarely is.
- Often prior information regarding the treatment effect or behaviour about a particular treatment arm may be available at the design stage but must be disregarded in a frequentist framework.
- Setting a minimum clinically relevant difference, δ, can be difficult in practice.

Further difficulties have been noted as to the effect on trials that a minimum clinically relevant difference can have. In many areas, such as oncology, any improvement, however small, may be considered clinically relevant. Designing trials on this basis results in unfeasibility large trials however. This can lead, as has been previously noted, to values of δ being set to satisfy a sample size calculation and therefore undermines the rigorous philosophy of a frequentist design. Furthermore, it allows for trials to observe a statistically significant difference but a point estimate that does not pass the criterion of being clinically relevant.

Bayesian sample size estimations are advantageous as they not only account for any parameter uncertainty, in key or nuisance parameters, but also can build in prior information at the design stage. Furthermore, as Bayesian designs typically rely on

simulation approaches, any layer of complexity/variability required can be built into the modelling approach and provide more informed sample size calculations than may otherwise be available.

To describe Bayesian sample size calculations, first consider that a trial is to be conducted which will collect data x which it is planned to model based on some parameters θ. Assume the final analysis will be based on a model which will provide a log likelihood of the form $l(\theta|x)$. Given $Pr(x|\theta)$ is proportional to $l(\theta|D)$ and the prior distribution for the model parameters, $Pr(\theta)$, the marginal distribution for the data x is

$$Pr(x) = \int_{\Theta} Pr(x|\theta)Pr(\theta).$$

The full posterior distribution in a Bayesian analysis is dependent on the marginal distribution. The Bayesian approach to sample size calculations are not based on controlling Type I and Type II error, rather they concentrate on controlling aspects of the posterior distribution. Denote the statistic on which the posterior distribution is evaluated as $T(x)$. Many forms of $T(x)$ have been previously proposed. Three popular approaches given by Joseph and Belisle [175] are the Average Coverage Criterion (ACC), the Average Length Criterion (ALC) and the Worst Outcome Criterion (WOC) with Wang and Gelfand suggesting two further approaches, the Average Posterior Variance Criterion (APVC) and the probability of detecting a treatment effect of size δ^*. These criteria are described in further detail in Sections 6.3.1 to 6.3.5.

Typically, analytical solutions are not available and so simulation approaches are required. The approach taken is set out by Wang and Gelfand [178, 183]. The general formulation is

1. Sample a value $\tilde{\theta}$ from the prior distribution $Pr(\theta)$
2. Sample data $\tilde{x}$ from the marginal distribution, dependent on $\tilde{\theta}$
3. Calculate $T(\tilde{x})$
4. Repeat the process for a total of N simulations

From N simulations we can directly calculate $\mathbb{E}[T(x)]$ as the arithmetic mean of $T(\tilde{x})$ over all simulations and $\mathbb{P}[T(x)] \in \mathbb{A}$ as being the proportion of the N simulations for which $T(\tilde{x})$ belongs to $\mathbb{A}$.

6.3.1 Average coverage criterion

Here some fixed length of the posterior distribution is set and the aim is to estimate the coverage of the posterior distribution provided by a given sample size. Define the point estimate (mean or median) of a symmetric posterior distribution for ψ and set

some fixed length l such that an interval $\mathbb{A}(y^{(n)}) = (\psi - l/2, \psi + l/2)$ can be formed. Given some $\alpha \geq 0$, define the ACC as

$$\mathbb{E}[Pr(\psi \in \mathbb{A}(y^{(n)})|y^{(n)})] \geq 1 - \kappa.$$

For non symmetric distributions, the length l can be amended to represent some highest posterior density (HPD) such that $\mathbb{A}(y^{(n)}) = \{\psi : Pr(\theta|y^{(n)}) \geq c_n(l)\}$. Here $c_n(l)$ is chosen such as the Lebesque measure of $\mathbb{A}(y^{(n)}) = l$. Typical values of κ are 0.05 or 0.1 which are equivalent to 95% and 90% credibility intervals.

6.3.2 Average length criterion

Similarly to the ACC, here the coverage of the posterior density is fixed and a length of some desirable credibility interval is calculated for a given sample size. As with the ACC we can define the length l either assuming a symmetric distribution or via an HPD for non-symmetric distributions. Firstly define the interval $\mathbb{A}(y^{(n)} = (F^{-1}_{\psi|y^{(n)}}(\kappa/2), F^{-1}_{\psi|y^{(n)}}(1-\kappa/2))$ where $F^{-1}_{\psi|y^{(n)}}(\kappa/2)$ defines the $\kappa/2$ quantile of the posterior distribution. The ALC is obtained by a given n which satisfies

$$\mathbb{E}[F^{-1}_{\psi|y^{(n)}}(1-\kappa/2) - F^{-1}_{\psi|y^{(n)}}(\kappa/2)] \leq l.$$

6.3.3 Worst outcome criterion

For each of the ACC and the ALC there is a 50% chance that the the coverage/length will be greater than what is desired as it only controls the required quantities on average. When this is of concern, the worst outcome criterion (WOC) is a viable alternative. Here, of instead of using the expectation some appropriate subset of the sample space $\mathbb{S}_0$ is defined such that

$$\inf_{y^n \in \mathbb{S}_0} [Pr(\psi \in \mathbb{A}(y^{(n)})|y^{(n)})]$$

6.3.4 Average posterior variance criterion

The average posterior variance approach aims to control the variance of the posterior distribution $Var(\psi|y^{(n)})$ and seeks an n for some $\epsilon \geq 0$ such that

$$\mathbb{E}[Var(\psi|y^{(n)})] \leq \epsilon.$$

6.3.5 Effect size criterion

In the clinical trials setting, it is often the case that interest lies in whether or not a parameter is greater than (or less than) some pre-specified effect size. This may, or may not be analogous to a minimum clinically relevant difference under a frequentist

framework. Setting some threshold value as ψ^*, a sample size n is obtained which satisfies

$$\mathbb{E}[Pr(\psi > \psi^* | y^n)] \geq \varrho.$$

Here ϱ is some given quantile of the posterior distribution. In many situations it is simply set that $\psi^* = 0$ which would represent evidence of some positive effect.

6.3.6 Successful trial criterion

The effect size criterion is extended by considering that at the outset of a trial, there are conditions under which the trial would be considered a success. The Successful Trial Criterion (STC) is then set up to estimate the probability that the posterior distribution of the key parameter of interest meets some pre-defined criterion.

Taking for example the situation where a trial would be defined a success if a sufficient proportion of the posterior density ψ is greater than (or less than) some threshold value ψ^*, we estimate n to satisfy the criterion

$$\mathbb{P}\{Pr(\psi^{(\vartheta)} \geq \psi^* | y^{(n)})\}$$

where $\psi^{(\vartheta)}$ is the ϑ level of the posterior distribution. Here for example, setting $\kappa = 0.1$ and $\psi^* = 0$ would consider the trial a success only if the 0.1 quantile of the posterior distribution is less than zero. Note that this criterion can be used to provide Bayesian equivalents of the frequentist Type I and Type II error rates although these are some what dependent on the prior distributions that are set for the parameters in the model. Treating all design parameters as fixed and setting prior distributions for the parameter of interest to match the null and alternative hypotheses with zero variance will give results analogous to frequentist Type I and Type II errors.

6.4 Bayesian sample size for survival data

Here it is shown how Bayesian design criteria are applied to time-to-event data. An illustration is provided with respect to the ViP trial carried out at the Liverpool Cancer Trials Unit to compare a treatment arm of combination chemotherapy, consisting of gemcitabine and vandetanib, against an arm of gemcitabine alone in patients with advanced pancreatic cancer.

Sample size calculations for clinical trials such as this are traditionally carried out in two steps. Firstly, the number of events are calculated, typically using Schoenfeld's formula [184] or Freedman's approximation [185]. Having determined the number of events required, some information is required regarding the survival distribution for the patient population under consideration to calculate the number of patients required. Often this step is carried out assuming the survival distribution to be fixed without allowing for any parameter variability.

An approach to Bayesian sample size calculations for survival data is proposed based on the assumption of proportional hazards and using a Piecewise Exponential Model (PEM), the likelihood for which is obtained by equation (2.6) in Chapter 2.

The PEM is dependent upon a parametric baseline hazard function, defined by the parameters $\boldsymbol{\lambda}$ and the log hazard ratio β. Prior distributions for both of these are required. From an analysis perspective, it may be desirable to leave both of these vague and uninformative. This can be damaging regarding trial design however as declaring that all survival functions are equally likely is unrealistic. In reality there is a good basis of information on the survival function. Indeed, it is typically treated as fixed in a frequentist framework. Methods for deriving baseline hazard parameters for the design and analysis of survival data are the main topic of Chapter 7 and are not discussed any further here.

The approach of Wang and Gelfand is followed in setting informative 'design' priors to estimate a sample size without specifying that these are required for the analysis of future trial data.

Given data D, define the marginal distribution for the piecewise exponential model as

$$Pr(x) = \int_{\Theta} Pr(D|\lambda, \beta) Pr(\lambda) Pr(\beta).$$

Simulating times from the sampling distribution is not straight forward due to the complex nature of the likelihood in use. Survival times are then simulated from the marginal distribution using the method of Bender [128] and following the steps of Section 5.5. Assuming a derived survival function $S(t|\lambda)$ given by (5.2), and derived from parameters taken from the design priors for $\boldsymbol{\lambda}$ and β, a single patients survival is obtained by simulating from a uniform distribution with limits $(0, 1)$. A survival time, $\tilde{t}$, is then derived based from

$$\tilde{t} = [-\log(S(t) - \sum_{g}^{j-1}(a_g - a_{g-1})]/\lambda_j + a_j \tag{6.1}$$

This formulation requires that it is known which partition each patient belongs. For a large number of observations, using (6.1) can become cumbersome. On this basis, a set of survival times may be taken from

$$t^* = [\log\{S(t)\} - \Psi]\phi^{-1}, \tag{6.2}$$

where $\Psi = \sum_{i=1}^{J-1} \sum_{g=1}^{i} \phi_{g-1} r_g$ and r is the distance between successive measurements of the time-grid such that $r_g = a_g - a_{g-1}$. In this form, t^* is a vector of length $J - 1$. Further adjust $t_j^* = \min(0, t_j^*)$ and $t_j^* = \max(s_j, t_j^*)$ for each j. Following this, set $\tilde{t} = \sum_{j=1}^{J} t^*$ to obtain simulated survival times $\tilde{t}$.

Denote $S_0(t|\lambda)$ as the survival function for the control arm and $S_1(t|\lambda,\beta) = S_0(t|\lambda)^{\exp\{\beta\}}$ as the survival function for the experimental arm. The approach taken is to first simulate survival probabilities $S_0(t|\lambda,\beta)$ and $S_1(t|\lambda,\beta)$ from a standard uniform distribution and use the results above to generate survival times. A censoring mechanism is generated assuming that the censoring mechanism is independent of the data generating mechanism for the survival times. In this case, censored patients are established by firstly applying 'administrative' censoring to any survival time larger than the maximum follow-up time of the study. Secondly, random censoring is applied using the same approach as Section 3.4 whereby a small number of patients are randomly selected as censored to control the overall censoring rate to match that of previous trial data.

6.4.1 Bayesian design of ViP

Here application of Bayesian sample size theory to the ViP trial for advanced pancreatic cancer is carried out. Note however that although a Bayesian methodology is being applied, no advocation for the use of any informative priors is included here. Investigations into informative design priors that may be used for the Bayesian design of a clinical trial are the main focus of Chapter 7. Please note that throughout this section, the notation of β to represent the log hazard ratio is retained and is not to be confused with the power attributed to a study design.

The initial trial was designed from a frequentist perspective with Type I and Type II error rates both set to 0.1 and a minimum clinically important difference given by a hazard ratio of 0.6. The initial sample size formula concluded that 100 events were required for the trial.

Information regarding the survival distribution for the intended patient population is derived from 3 previous trials [186, 187, 188], all of which included a group of patients with a similar prognosis who received gemcitabine alone. Figure 6.1. shows the survival curves from these trials. Using this information the initial sample size calculation estimated that 120 patients with a minimum of 12 months follow-up would be required to obtain 100 events.

To estimate the required sample size, definitions are required for the criteria under which the trial is designed. That is, for the ACC a length 'l' is required, for the ALC a coverage level is required, for the effect size criterion a threshold value ψ^* is required and for the successful trial criterion an ϑ level of the posterior distribution and a threshold value are both required.

Re-designing ViP from a Bayesian perspective requires re-evaluating the initial aims of the trial as noted by Simon [189]. As has been previously stated, both the Type I and Type II error rates are quantities that are calculated by conditioning upon some fixed value of the key parameter of interest (typically $\beta = 0$ and $\beta = \delta$). As the Bayesian perspective assumes β to itself be random variable, fixing it at some arbitrary values is

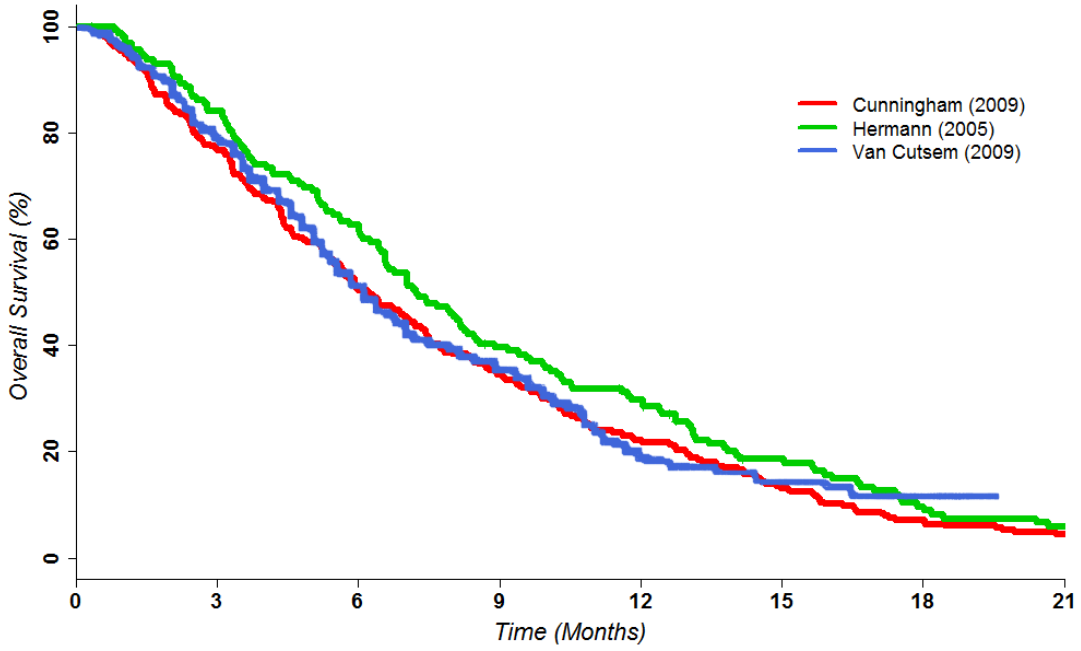

Figure 6.1: Kaplan Meier plot of the trials including a Gemcitabine arm in patients with advanced pancreatic cancer in preparation for the design of the ViP trial

not appropriate.

The ViP trial will be a success in a frequentist framework if a one-sided P-value less than 0.1 is observed. From a Bayesian perspective, the trial is considered successful if less than 10% of the posterior distribution is greater than zero. Based on this, setting a desired coverage for the ALC of (0.1, 0.9) seems reasonable. A minimum clinically relevant hazard ratio of 0.6 (a log hazard ratio of $\delta = -0.51$) was set and with this is mind, $l/2 \leq \delta$ is required and a value of $l = 0.8$ is chosen as a conservative measure to ensure that if a clinically relevant difference is observed, then a posterior credibility interval should not contain zero.

For completeness, an effect size of interest is set as being $\psi^* = 0$. Another option may be to set $\psi^* = \delta$ however here it is considered that any value of $\beta > 0$ is of clinical interest. For a successful trial criterion set $\vartheta = 0.1$ and retain $\psi^* = 0$. Here, a successful trial is defined if <10% of the posterior distribution is greater than zero. This allows evaluation of success directly under conditions that replicate the initial trial design.

A summary of the design conditions are given in Table 6. For completeness when designing the trial the variance criterion (VAC) is also assessed by setting a standard deviation of $\sigma = 0.24$. This is chosen to give an approximate 95% interval length to match the conditions of the ACC. Lastly, the worst outcome criteria (WOC) is set as being the maximum length obtained to give a coverage of (0.1, 0.9).

Having defined the design criteria for the trial, the estimation of these criteria are

Design Criterion	Critical Value
ALC	coverage $= (0.1, 0.9)$
ACC	length $= 0.6$
VAC	$\sigma = 0.241$
ESC	$\psi^* = 0$
STC	$\psi^* = 0$, $\vartheta = 0.1$
WOC	coverage $= (0.1, 0.9)$

Table 6.1: Table illustrating the design criteria for the Bayesian design of the ViP trial.

illustrated through the means of a single sampled dataset. Following the design priors approach of Wang and Gelfand [178], prior distributions for the baseline hazard function and the log hazard ratio are required from which the data will be sampled. The prior distributions for the baseline hazard function in turn depend upon the specified time-grid. Here a fixed time grid of $a = (0, 3, 6, 12, 18, 30)$ is set. The design priors for the log baseline hazard parameters $\boldsymbol{\lambda}$ and the log hazard ratio β are set as

$$\boldsymbol{\lambda} \sim MVN(\boldsymbol{\gamma}, \boldsymbol{\Sigma}_\gamma) \tag{6.3}$$

$$\beta \sim N(\log(0.6), 1/3). \tag{6.4}$$

With values of $\boldsymbol{\gamma} = (-2.68, -2.12, -2.02, -2.15, -2.73)$, the covariance matrix,$\Sigma_\gamma$ is a diagonal matrix with elements given by $(0.33, 0.41, 0.50, 0.58, 0.66)$. Further discussion on the derivation of prior distributions for the baseline hazard function are the topic of Chapter 7. The prior for β is set to have a mean equal to the minimum clinically relevant difference of the original trial design, and a variance of $1/3$. This keeps the trial design consistent with the initial trial design, the value for the variance is chosen so that 99% of the prior distribution is of approximate length two.

To describe the sampling procedure of a Bayesian design, initially a single dataset is investigated. Sampling a single set of parameters from the prior distributions, values are obtained such that $\tilde{\boldsymbol{\lambda}} = (-3.62, -2.63, -2.31, -1.66, -1.74)$ and $\tilde{\beta} = 0.34$. A single dataset of 100 patients is sampled from these parameters using the approach defined in Section 5.5. The Kaplan Meier estimates from these data are shown in Figure 6.2. Also shown are a set of 1,000 draws from the baseline hazard function from which the data are sampled.

To the sample data a Bayesian piecewise exponential model is fit with non-informative prior distributions. For simplicity, in this case the time-grid is kept as fixed. In practice however, redefining the time-grid in light of the sampled data may be carried out. The model parameters from this model are given in Table 6.2. Note that the model parameters do not match exactly the sampled parameter estimates from the design priors, due to the random variability that is introduced in the sampling of survival times.

For the single dataset, the posterior distribution for β is used to estimate the design

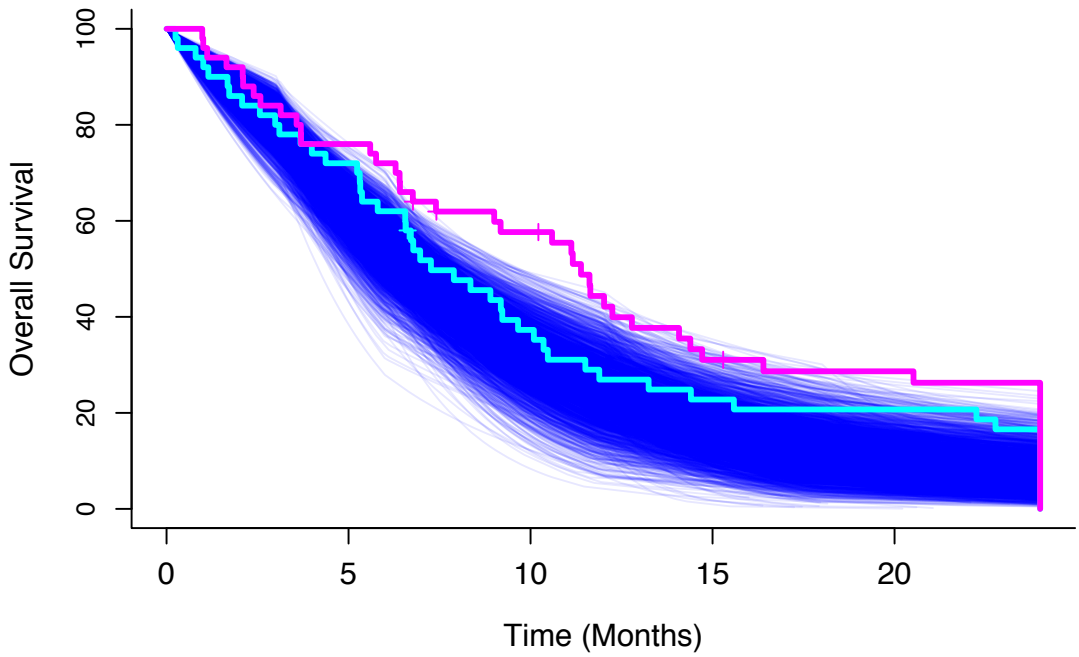

Figure 6.2: Illustration of the behaviour of the survival function under the sampling priors for $\boldsymbol{\lambda}$. Also plotted are the data from a single simulated dataset, the sampled parameters here are $\tilde{\boldsymbol{\lambda}} = (-2.96, -2.18, 2.04, -2.11, -2.82)$ and $\tilde{\beta} = -0.34$

	Mean (Std. Dev.)	(95% Cred. Int)
λ_1	-2.638 (0.257)	(-3.166, -2.159)
λ_2	-2.619 (0.279)	(-3.191, -2.105)
λ_3	-2.168 (0.212)	(-2.597, -1.763)
λ_4	-2.666 (0.335)	(-3.366, -2.044)
λ_5	-1.645 (0.237)	(-2.135, -1.204)
β	-0.245 (0.205)	(-0.658, 0.153)

Table 6.2: Table showing the parameter summaries from the analysis of a single sampled dataset from the sampling distribution with design priors.

criterion defined in Section 6.3 with the criteria set in Table 6.1. Figure 6.3 shows the posterior probability density from the single sampled dataset. Also shown are an illustration of how each criterion is evaluated. Note that the worst outcome criteria is not included here as this is only evaluated over all simulated datasets.

It is observed that for this sample dataset, the length of the posterior distribution that covers the 80% credibility interval is 0.85. Conversely, for a fixed length of $l = 0.6$, we would obtain a coverage of 0.75. The effect size criterion here shows that 0.01% of the posterior distribution is greater than $\psi^* = 0$. Lastly the successful trial criterion here gives a value of 1 as the 10% quantile of the posterior distribution is greater than the threshold value indicating that we would conclude from this trial that the treatment is effective and worthy of further investigation.

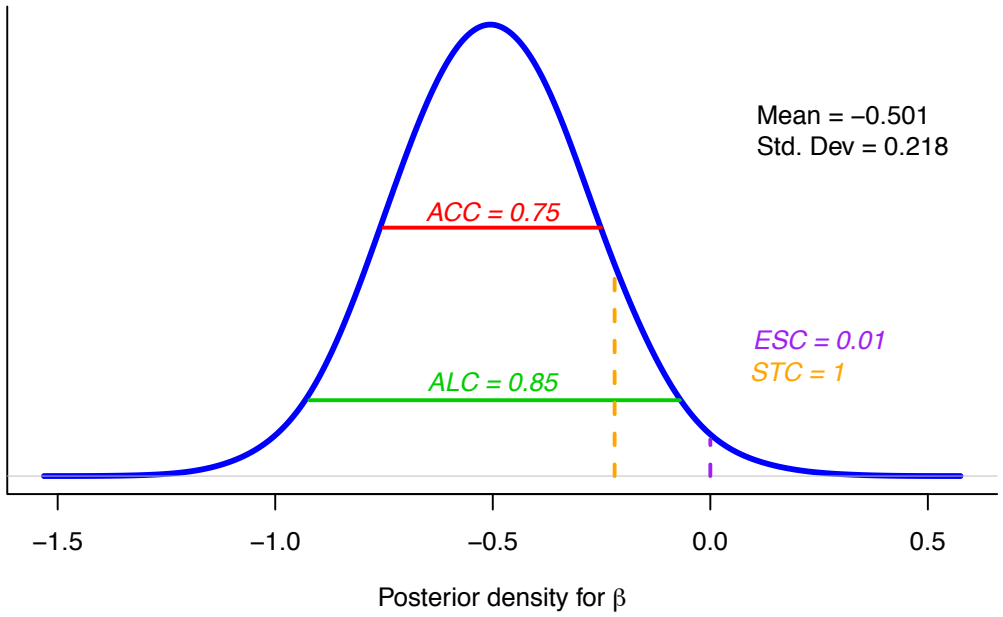

Figure 6.3: The posterior distribution distribution for $\tilde{\beta}$ from a single sampled dataset with Bayesian sample size criterion

The full Bayesian design criteria are obtained by simulating further datasets and evaluating the criteria over all simulated datasets. Expectations of ACC, ALC, VAC and EFC are obtained by taking the arithmetic mean of the criteria and the probability of a successful trial is calculated as the proportion of occasions that a successful trial is observed. The WOC is obtained as the maximum length giving a coverage of $(0.1, 0.9)$ over all datasets.

Data are simulated under total sample sizes of $N = (60, ..., 150)$ by intervals of ten. For each sampled dataset 5% of the observations are randomly selected to be censored and further administrative censoring is applied on any survival time $t \geq 24$ where time is measured in months.

Five hundred datasets are sampled in total. The results are shown in Figure 6.4. Here the effect that an increased sample size has on each criterion is shown. Taking initially the ALC, to obtain a coverage of $(0.1, 0.9)$ the length of the posterior distribution required decreases with increased sample size. For example, given the fixed coverage level, 100 patients are required to ensure an interval length of 0.6 is obtained (illustrated).

Considering the ACC, it is shown that with a fixed length of 0.6, the coverage of the posterior distribution increases with the sample size. Here a sample size of 100 ensures a coverage of 0.8 which is not surprising considering the similarities to the ALC. The VAC illustrates how the variance (standard deviation) decreases with the increased

sample size. Here again 90 patients are required to obtain a standard deviation of 0.241.

The ESC and STC are both included here but these criterion are dependent upon the prior distribution for β. Taking for example the ESC, it is shown how there is a general decrease from approximately 0.15 to less than 0.1 as sample sizes increase. This decreases as sample sizes increase as the posterior distribution for β is here more dependent upon the prior distribution $Pr(\beta)$ and less dependent upon the prior distribution for the baseline hazard parameters. The STC decreases in a similar fashion. Note that as here, the $Pr(\beta > 0)$ is being measured and a decrease in *beta* is desirable then $1 - STC$ may be of greater interest. Here for example, as the sample size increases to 150 patients there is approximately an 80% chance of a successful trial being observed. This is again dependent upon the prior distributions set however and should be treated with some caution.

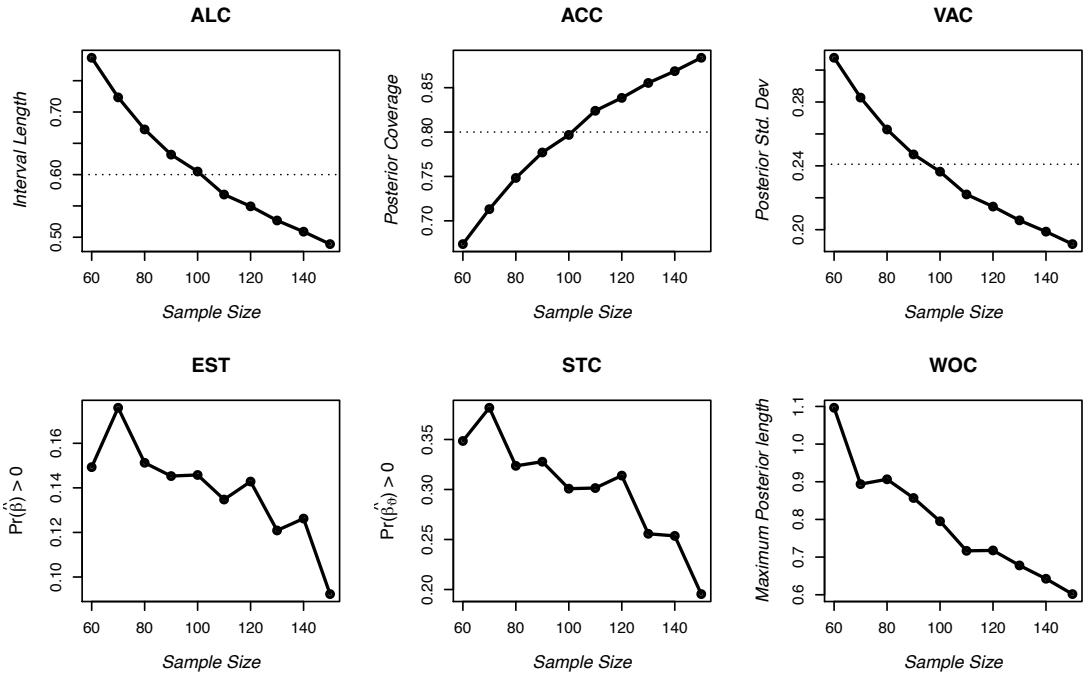

Figure 6.4: Bayesian sample size criteria for the ViP trial

Finally the WOC is considered and it is observed that whilst this decreased with an increased sample size, as expected due to its relationship to the ALC, there is much more variability in this criterion. The reason for this is illustrated in Figure 6.5. Here an illustration is given to the calculation of both the ALC and the WOC. Presented are the estimated lengths that give the required coverage from each individual sampled dataset. Both the ALC which is the arithmetic mean of all estimated lengths and the WOC which is the maximum observed length are shown. In this case the WOC is some way off the rest of the estimated lengths and may be considered an outlier. This helps

to explain the more erratic nature of the WOC.

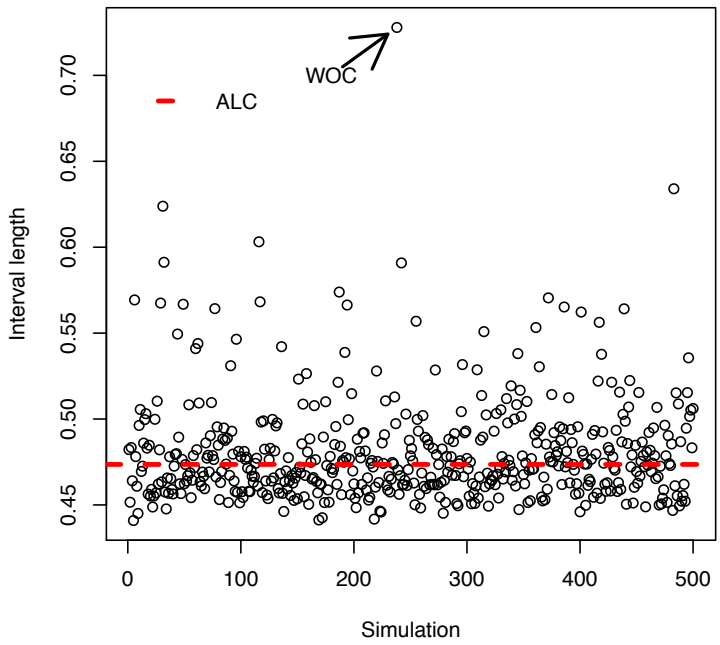

Figure 6.5: An illustration of the calculation of the ALC and WOC criteria

Given the six criteria considered in this section, the ALC and the ACC are considered the most appropriate criteria to use in this setting. The VAC provides a reliable estimate but it is considered that setting a desired variance (or standard deviation) may not be a straight forward undertaking prior to a trial. Both the ESC and the STC are quantities that are overly dependent on the prior distribution for β with the $1 - STC$ in particular being interpreted as the probability of observing a statistically important difference. In this regard, similarities to frequentist Type I and Type II error rates can be obtained but care should be taken here as traditional Type I and Type II error rates are dependent upon fixed known differences between two treatment arms whereas the STC takes this difference (through the prior for β) as a random variable. There is reduction in the variability for both the STC and ETC for larger sample sizes as there is less effect from the variability about $\boldsymbol{\lambda}$ and the criteria become more dependent on the prior distribution for β.

When there is a reasonable amount of prior information the ESC and STC quantities may be appropriate, however the extra variability in these results suggests that more simulations may be required which may be computationally expensive. The WOC may be an appropriate measure conceptually but has practical difficulties which render it unsuitable in this instance. A utility based on the 90^{th} quantile of the distribution of

interval lengths may be one solution here as a more conservative measure than the ALC without the erratic nature of the WOC.

The ALC and ACC both have the attractive properties of controlling desirable aspects of the posterior distribution and providing stable estimates upon which sample sizes can be estimated. They may be criticised as they only control the required quantities on average [182]. That is to say, prior to a trial beggining, there is only a 50% chance that the trial will control the desired parameter at a given quantity, however compensations can easily be made to account for this. The choice on whether to control the length of the interval or the coverage is somewhat arbitrary, and the best choice may be to use whichever criterion is most accessible to clinicians. Fixing the coverage in this case may be the more straight forward approach as many clinicians will understand the concept of 95% confidence intervals, and may even interpret them as credibility intervals from the onset [5].

6.5 Discussion

In this chapter the main ideas behind the use of Bayesian methodology for the design of clinical trials were introduced. The main concepts are compared to the frequentist methodology typically used in practice and some arguments made for the Bayesian approach. Bayesian sample size methods typically require some sort of simulation techniques and it is illustrated how these techniques are applied to a survival context. Lastly some application to the design of the ViP trial was given.

The purpose of any trial is to gain new information. In a clinical trial setting this will often, if not always require comparing some new experimental treatment to a placebo or active control. Intuitively, the concepts of Type I, Type II error and a minimum clinically relevant difference are not as accessible as direct Bayesian probabilities. These quantities all depend upon estimating some fixed but unknown quantity, declaring that it needs to pass some threshold value to be of clinical import and then calculating sample sizes based on these assumptions. Whilst it is claimed that the frequentist approach has the advantage of objectivity [1], this can be misleading as

- The idea of a clinically relevant difference (CRD) is inherently subjective
- Sample size calculations will often be based on assumptions such as the baseline hazard rate. Typically Bayesians will allow these parameters to be random variables whereas frequentist methodology will treat them as fixed.
- Parameters such as the CRD, Type I and Type II rates will often be set to arrive at a suitable sample size [150].

Bayesian designs have the advantage that they can answer the clinical question directly. For example, ‘what is the probability that an experimental treatment is better

than the control?'. Sample size calculations are based on the principle of controlling aspects of the posterior distributions as opposed to controlling against the chance of being wrong. Furthermore, Bayesian designs have the added advantage of flexibility, allowing for changes in the trial without harming the final interpretation whereas frequentist P-values, being a probability statement based upon the theoretical reproducibility of the data, can be difficult to define following alteration to the trial design once the trial has begun.

Lastly the point is made that a Bayesian design typically implies a Bayesian analysis although this need not be the case. Furthermore, Bayesian designs do not imply that some sort of informative priors are going to be incorporated into the analysis of any trial and can be utilised more for the added freedom they give as opposed to attempting to alter any analyses with prior opinions.

The flexibility of Bayesian designs can result in additional complexity. Whilst a Bayesian simulation approach to sample size calculation allows for variability to be built into multiple aspects of a trial design, it does inhibit the development of a 'one size fits all' design package. Typically the user is required to write some simulation routine tailored to the study they are designing. Here, simulation routines were written in the statistical package 'R', using the code that is provided in Appendix A4. The difficulty in producing Bayesian designs compared to ease of the standard approach goes some way to explaining the dominance of the frequentist approach. There is ground being made however and Bayesian designs are becoming more common which is best demonstrated by the increasing number of trials designed by a Bayesian approach at the MD Anderson Cancer Center [149] and there uses in the design of studies with rare disease [**?**].

In the next chapter, some exploration into informative priors are made. In particular, informative prior distributions on the baseline hazard function are set with the aim to show how such priors can be used to design clinical trials which make a more efficient use of costly data.

Chapter 7

Bayesian Design and Analysis of a Cancer Clinical Trial with a time-to-event endpoint

7.1 Introduction

In this chapter, a Bayesian design of a cancer clinical trial with a time-to-event endpoint using informative priors is considered. Again the ViP trial coordinated by the Liverpool Cancer Trials Unit (LCTU) is used as an example and a design proposed which makes use of the historical information available for a single arm of the trial.

This chapter is structured as follows, firstly a review of the uses of historical information in the design and analysis of clinical trials is provided. Following this, a method by which informative prior distributions can be derived from summary information for a single arm of a trial with a time-to-event endpoint is proposed. Also used are local step and trapezium priors which are uninformative only when observed data agree with prior information and are designed to encourage likely solutions when only sparse data are available.

Lastly, it is demonstrated that the proposed methodology can be incorporated into the design of the ViP trial, resulting in a trial that can be carried out with fewer patients.

7.2 Historical controls in clinical trials

All clinical trials with a time-to-event endpoint are designed with reference to previous trials and/or data. Typically however, whilst historical information may guide issues such as sample size calculations or estimated event rates, this information is discarded when analysing the final results of a trial.

The exploration of more formal inclusion of historical information has gained attention. As far back as 1976, Pocock [190] was making a case for the use of historical controls in clinical trials. Importantly here, Pocock introduces six key criteria which

should be met in order to compare experimental treatments against historical controls. It has been noted by Speigelhalter et al. [21] that these conditions may be overly stringent and some relaxation may be reasonable, especially where some allocation is given to a control arm.

A further review by Sacks et al. [191] in 1982 compared the uses of randomised controlled trials to historical control trials and noted that using historical controls alone can lead to bias in the interpretation of trial results. Results such as this have lead to the domination of randomised control trials.

Attention has been given to the use of historical information to inform elements of a trial. Initially, the methodology was confined to carcinogenicity studies, Tarone [192] in particular introducing a method to incorporate historical information in a test for trends which has been used and adapted in practice [193, 194, 195]. With respect to trial design, Thall and Simon [196] consider the use of historical information for the formal use in the design of Phase II studies.

As an extension of the analyses for carcinogenicity trials, methodology for the incorporation of historical information to the analysis of Poisson means has been developed [197, 198] with application of historical controls to the analysis of bioassay data explored by Chen et al. [199]. Whilst methodology has being developed, its application outside of carcinogenicity trials remained sparse, despite reviews of Bayesian methods by Racine et al. [200] commenting on the '...tremendous scope for improved design and analysis using historical information'.

More recently, interest has grown with Ibrahim et al. [201] considering using historical information to adjust covariates in logistic regression models and French et al. considering their use in three-state models[202]. Chen et al. [203] consider the inclusion of historical controls into the the Bayesian design of non-inferiority trials with a binary endpoint.

One approach that has gained popularity is the use of power priors, see for example De Santis [204]. Here, it is assumed that some previous clinical trial data are available in their entirety. Historical and current data can then be analysed simultaneously with the results of the historical data informing the current data weighted by some power parameter. This approach was applied to survival data by Ibrahim et al. [58] and further by De Santi [183] who considers this approach for the use of Bayesian sample size calculations with a simple example set to the parametric exponential survival model.

Further to this, Neuenschwander et al. [205] introduce a meta-analytical approach of previous trial data for summarising the control information (MAP). Here a single parameter ς^* is assumed to contain all information about a control arm in a trial. The meta analysis approach allows for both within and between study variability to be summarised and the authors suggest using the predictive distribution of ς^* as a prior distribution for inclusion in the design and analysis of a future trial. Gsteiger et

al. [206] summarise Neuenschwander's approach for the case of over dispersed Poisson data. Importantly here, aggregate data as well as complete data are used in estimating ς^*.

The idea of a set of robust priors termed 'commensurate' has also been explored by Cook, [207] Fuquene et al. [208] and Hobbs et al.[209]. Here prior distributions influence the likelihood along with a 'commensurability' parameter which measures the degree of agreement between the information in the prior distributions and that collected in the data. Where the data agree with the prior information the priors are relatively informative; where there is poor agreement however, the posterior distributions are more dependent upon the observed data and the effect of the priors is lessened.

A recent review of the differing methods of incorporating historical information in the evaluation of a treatment effect is given by Viele et al. [210] comparing the [205] MAP method to the power prior approach, Pocock's method and more rudimentary methods of pooling data. Though no recommendations are given, the authors do promote the methods as a means for obtaining smaller more efficient trials and for considering trials with allocation ratios other than the standard 1:1.

Whilst there is some well established methodology for the formal incorporation of historical priors into the analysis of clinical trials data, some key questions still remain unanswered. To date, much of of the methodology depends upon historical information being known in its entirety. This is often impractical. Furthermore, developing priors only on available data may introduce a level of selection bias into any analysis.

Whilst Gsteiger et al. [206] do explore the inclusion of aggregate data, the application of such an approach to survival data presents a number of challenges. Under a standard two arm trial, the control arm can be defined by the set of baseline hazard parameters and in the vast majority of survival models, sufficient information will not be reported for an aggregate meta analytical approach to be a viable option. Furthermore, it is unlikely that a single parameter will be sufficient to contain all information regarding the control arm of a trial in this context.

In the remainder of this chapter, a method is proposed for estimating baseline hazard parameters where only summary information is available. Following the estimation of a historical baseline hazard function, some discussion is given to the definition of prior distributions and the effect these priors have on the design and analysis of a trial with a time-to-event endpoint.

7.3 Derivation of priors for baseline hazard parameters

Here a method is proposed for the estimation of baseline hazard parameters where only summary information is available.

Assume initially that prior information, $D_p = (\vartheta_p, t_p)$, is available which takes the form of survival probabilities ϑ_p at associated time points t_p. This information could

be taken from obtained datasets, published material such as Kaplan Meier curves or derived from expert opinion.

It is further assumed at the design stage that the data will be modelled using one of the parametric family of models, i.e. a model which includes some parametric description of the baseline hazard function as well a hazard ratio upon which the trial will be assessed. Methodology is presented here specific to the piecewise exponential model as this provides a flexible modelling approach with practical applications. Adaption to other parametric forms are not presented here but can be easily obtained.

Assume a PEM and that previous data, D_p, are available in the form of a set of survival probabilities $\{\Phi_p\}$ along with associated time points $\{t_p\}$. This information could be taken from obtained datasets, published material or derived from expert opinion. Given D_p, the objective is to convert the prior survival probabilities $\{\Phi_p\}$ into survival probabilities, $\{\phi_j\}$, at time points corresponding to those in the time grid used in the analysis of the PEM $\{a_j\}$. Estimates of $\{\phi_j\}$ could be obtained via simple linear interpolation or by fitting a spline model with knots κ such that $\Phi_p = f(t_p, \kappa)$; estimates of ϕ_j can be taken as the fitted values of the spline function at the partitions of the time grid. It is the structure $\{\Phi_j\}$ and the time-grid $\{a_j\}$ which are used to estimate $\{\gamma_j\}$, the hyper parameters which define the point estimates of the prior distributions $Pr(\boldsymbol{\lambda})$.

From the definition of the PEM given by 2.2, a survival function is defined as

$$S(t) = \exp\Big\{ - \Big[\lambda_j(t - a_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(a_g - a_{g-1})\Big]\Big\}.$$

Replacing λ_j with γ_j and evaluating the survival function at partitions in the time grid,

$$\phi_j = \exp\{-\gamma_j(a_j - a_{j-1})\}\phi_{j-1} \tag{7.1}$$

is obtained. Consequently point estimates for prior distributions on a baseline hazard function are derived via

$$\gamma_j = -\frac{\log(\frac{\phi_j}{\phi_j - 1})}{a_j - a_{j-1}}.$$

A graphical illustration of the process of obtaining point estimates for the prior densities is given in Figure 7.1.

The full form of the prior distributions depend on whether hazard ratios are measured on the standard or log scale. Including baseline hazard parameters on a nominal scale, priors are set using a Gamma distribution for each γ_j individually such as $\Gamma \sim (\eta_j, \zeta_j)$ which is constrained by $\gamma_j = \eta_j/\zeta_j$. For practical and computational convenience, throughout this thesis all baseline hazard parameters are defined on the log scale and $\varpi = \log(\gamma)$ defined. Prior distributions are defined such that

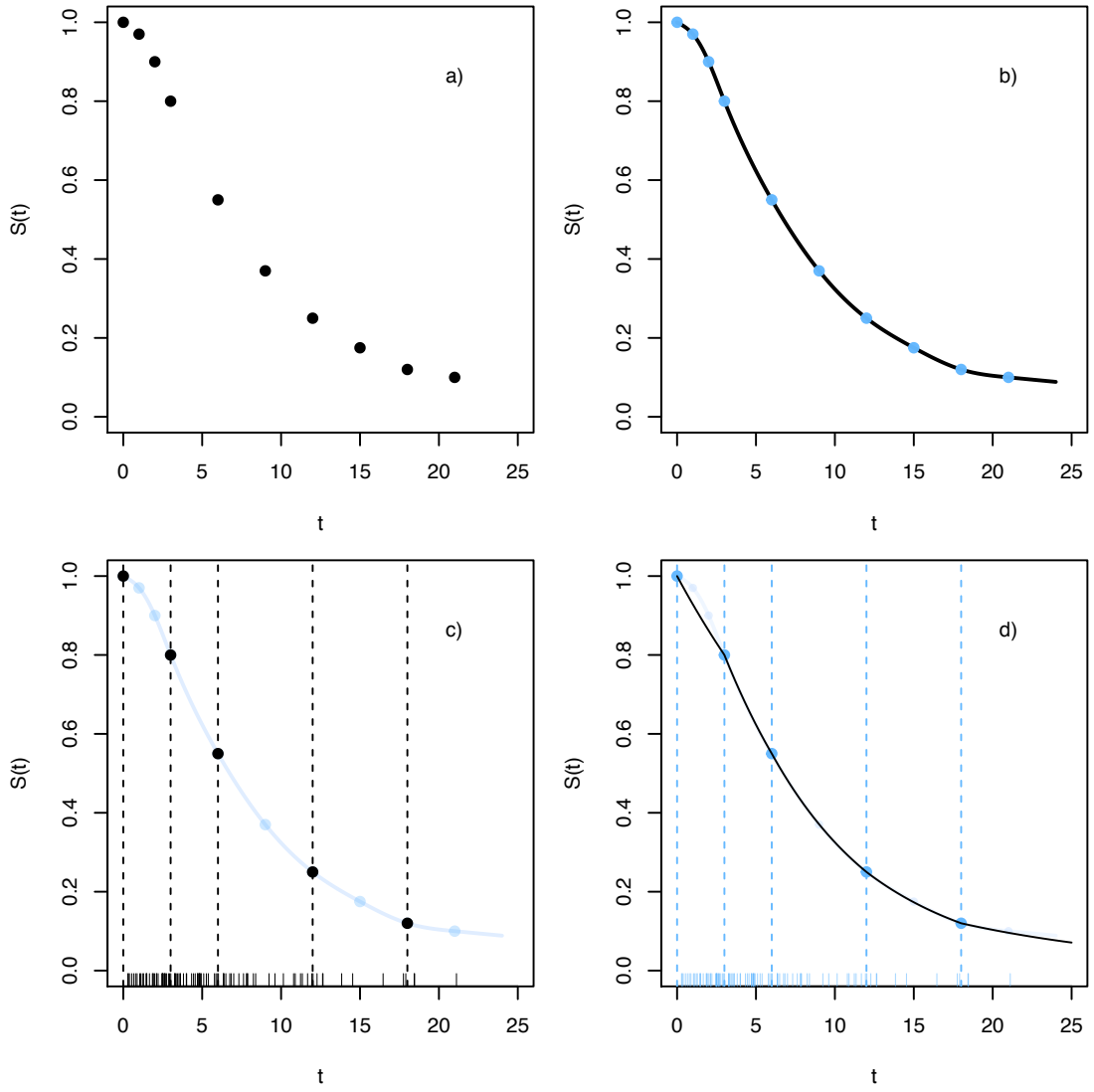

Figure 7.1: Figure to illustrate the process of deriving parameters for informative prior distributions on a baseline hazard function. Figure a): the prior estimates of survival probabilities and associated times are obtained. Figure b): a spline function fitted to the prior estimates. Figure c): data are observed (rug plot) and the time grid is set. Figure d): prior parameter estimates γ are obtained and the resulting piecewise model estimate is given.

$$Pr(log(\boldsymbol{\lambda})) \sim MVN(\boldsymbol{\varpi}, \boldsymbol{\Sigma}). \tag{3}$$

Here $\boldsymbol{\Sigma}$ is a $j \times j$ covariance matrix with elements $\rho_{i,j}$ which quantifies the degree of confidence in the prior parameters. It is assumed a-priori that all baseline hazard parameters are independent and set $\rho_{i,j} = 0$ if $i \neq j$. This is not to assume that the baseline hazard parameters themselves are independent, only that there is no prior knowledge of any correlation between prior parameters. Here this assumption is made for convenience but may be relaxed by assuming structured correlation structures. Full definitions of the diagonal elements are non trivial and discussed further in Section 7.3.2

7.3.1 Prior precision for the baseline hazard function

A particular challenge in this approach is to set the precision of the prior distributions. Fully data dependent methods such as those proposed by Neuenschwander [205] are based on predictive distributions which allow for both between and within study variability. Deriving prior probabilities from summary information however makes it difficult to obtain reliable estimates of both within and between study variability.

In deriving prior distributions from summary information, clinicians and statisticians need to determine to what degree a future trial can be considered relative to historical information. Whilst this may be no easy task, this does offer an advantage over fully data dependent approaches as prior distributions can be amended to reflect the degree to which future data are believed to relate to historical information, or to what extent any scepticism over the validity of historical information exists. For example if the historical information is taken from data collected a number of years previously, clinicians may wish to inflate prior variability to account for the fact that medical standards may have progressed.

Two approaches are explored here, a graphical approach and an approach based on the effective number of events. Both approaches are adapted to weight more prior information on the earlier partitions as these prior survival estimates are expected to be more reliable.

For the graphical approach, each of the j diagonal elements of the matrix Σ is defined by $\rho_{jj} = c\varsigma_j$ where c is an overall level of variability and ς is a variance inflation function. In practice, appropriate values of c and ς will be dependent on the amount of data being analysed and the time-grid that is set and it is for this reason that a graphical approach is proposed. Examples of prior survival functions are given in Figure 7.3. These figures are obtained by taking samples from the design prior distributions based on informative priors and converting into survival functions using equation (5.2). Figures such as these can also be useful in explaining to medical professional how the prior information is likely to impact on future data analysis.

As an example, useful functions of ς may be to increases from one to two by equal steps over the J categories. Other approaches are applicable, for example letting $\varsigma = 1/\sqrt{a_{j-1}}$, which has the property of allowing prior precision to decrease at a rate proportional to the assumed survival function.

A more practical approach may be to determine the prior information in the number of effective events. For example, given a prior baseline function, a clinician may state that they want this function to have the equivalent effect of 20 events in an upcoming trial.

Following this approach, consider that for each individual interval in the PEM, the hazard rate parameter can be considered to follow a gamma distribution with parameters η_j and ζ_j, where η_j can be taken as the number of events observed within an interval and ζ_j is the patient time at risk. This distribution has mean given by

$$\gamma_j = \eta_j \zeta_j^{-1}$$

which is estimated using (7.1). As prior distributions are defined on the log scale, variability about $log(\gamma_j)$ can be obtained via the delta method such that

$$\begin{aligned} Var(\log \gamma) =& Var(\gamma)[(\log \gamma)']^2 \\ =& \eta \zeta^{-2} (\eta \zeta)^2 \\ =& \eta^{-1} \end{aligned} \tag{7.2}$$

where $'$ represents the first derivative. Denoting hazard rates on the log scale therefore shows that variability about a baseline hazard rate can be determined using only the number of events within each interval. Given that a hazard rate and an effective number of events (E_p) has been pre-specified therefore, the prior variability for each partition η_j can be defined as

$$\eta_j = \frac{E_p(\phi_{j+1} + \phi_j)}{2 \sum_{i=1}^{J} \phi_i}.$$

.

This again weights the available events so that more prior events are attributed to earlier time partitions as these are the partitions upon which estimates are considered to have greater reliability. For the remainder of this paper, prior distributions are formed based on the effective number of events approach.

It is lastly noted that it may be reasonable in practice to set more than one set of prior distributions. Ultimately however, a single scenario may need to be defined on which a future trial will be assessed.

7.3.2 Definition of the time grid

A brief note is included here to highlight that the time-grid, which is generally assumed fixed for the PEM will play an important role in both the derivation of prior parameters and the analysis of the data. It is possible to have one time-grid that is responsible for deriving prior distributions, perhaps in a design setting, and a second time-grid which is data dependent and is used for analysing trial data. One complication in this approach however is that prior distributions based on the 'design' time-grid would have to be amended in light of the 'analysis' time-grid.

Whilst this approach is both feasible and in some respect desirable as the 'design' time-grid may be sub-optimal for analysis of trial data, it does introduce an extra layer of complexity. For this reason, a fixed time-grid as proposed by Kalbfleisch is used throughout the remainder of this thesis for both the design and analysis of trial data.

7.4 The analysis of time-to-event data with informative priors on a baseline hazard function

Here, an exploration of the effects of an informative prior on the baseline hazard function is carried out with specific application to the GemCap trial.

Data to be analysed are taken from the Cunningham trial (GemCap) [186], a trial to investigate the use of Gemcitabine and Capecitabine for the treatment of patients with advanced pancreatic cancer. The trial recruited a total of 534 patients. The final analysis, although showing some survival benefit for the Gemcitabine and Capecitabine arm (P-value = 0.08).

Aside from the GemCap trial, two further trials are available: Herman et al.[187] and Van Cutsem et al. [188], with total sample sizes of 319 and 301 patients respectively, both of which contain a Gemcitabine arm and are used to estimate baseline hazard priors. Table 1 provides summaries of the prior information in terms of survival estimated at given time-points taken from the Herman and Van Cutsem trials along with arithmetic means of the two, $\{\phi_p\}$ which shall be use to derive prior point estimates.

Time (Months)	1	3	6	9	12	15	18	21
Hermann (2005)	0.98	0.84	0.63	0.40	0.30	0.19	0.10	0.06
Van Cutsem (2009)	0.96	0.78	0.50	0.36	0.21	0.18	0.15	0.15
ϕ_p	0.97	0.80	0.55	0.37	0.25	0.18	0.12	0.10

Table 7.1: Derivation of prior survival estimates at given time points.

Using a fixed time grid given by $a = (0, 3, 6, 12, 18, t^*)$ where t^* the maximum observed time, baseline hazard point estimates, $\boldsymbol{\varpi}$, are obtained using the methods outlined in Section 7.2. Here estimates of $\boldsymbol{\varpi} = (-2.60, -2.08, -2.03, -2.10, -2.59)$ are

obtained.

Figure 7.2 shows the Kaplan Meier survival estimates of each treatment arm of the GemCap trial along with the fixed time-grid and the survival function derived from the prior point estimates. Of main interest here is to compare the prior survival function against the estimates obtained from the Gemcitabine arm of the GemCap trial. Figure 7.2 shows a general good level of agreement although the survival estimates in the Gemcitabine arm are slightly below that of the estimates obtained from the prior data.

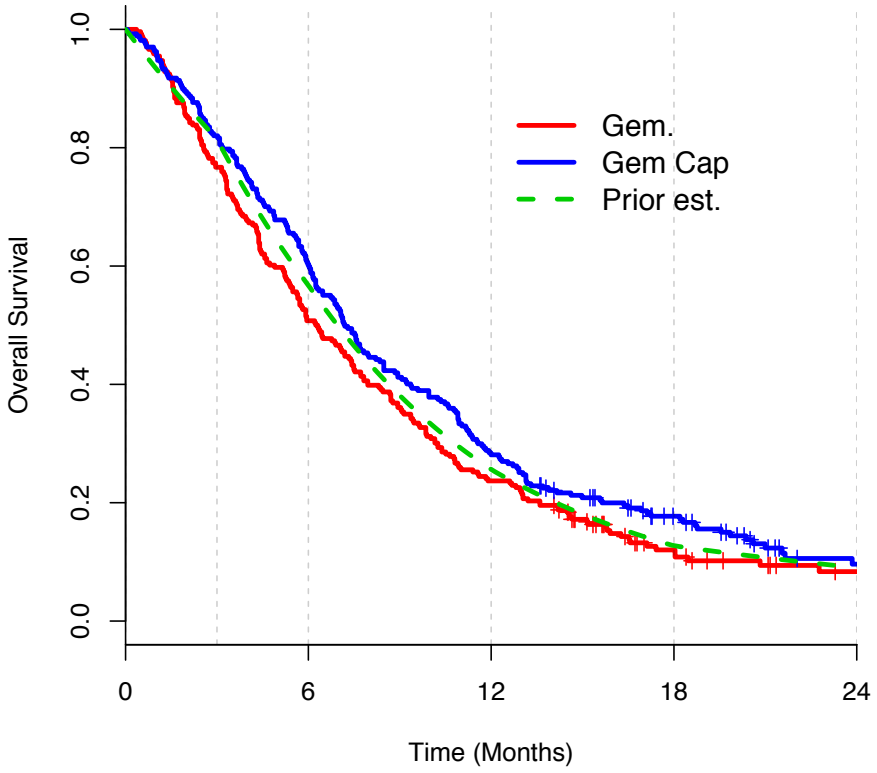

Figure 7.2: Kaplan Meier estimates from the GemCap data along with an estimate of the survival function obtained from the point estimates of the prior distributions

Some consideration is now given to the level of variability that is attributed to the prior point estimates. Here the effective events approach as defined in Section 7.3.2 is followed. In practice, multiple priors which give varying degrees of belief, reflected in differing total number of effective events may be set to reflect a reference, sceptical and optimistic approach suggested by Speigelhalter et al. [21].

It is noted that an attraction of the MAP approach is to account for uncertainty both within and between studies wheres the pooled baseline estimates effectively ignore any between study variation. With reference to the GemCap study therefore, whilst the

prior information is reasonably consistent, there may still be a desire to use relatively vague priors to ensure that the data remain dominant in the interpretation of the trial. Despite the prior data being derived from approximately 300 events therefore, effective event sizes of 10, 25, 50 and 100 events are investigated. Figure 7.3 provides a graphical illustration of the survival functions that are obtained from these effective numbers of events.

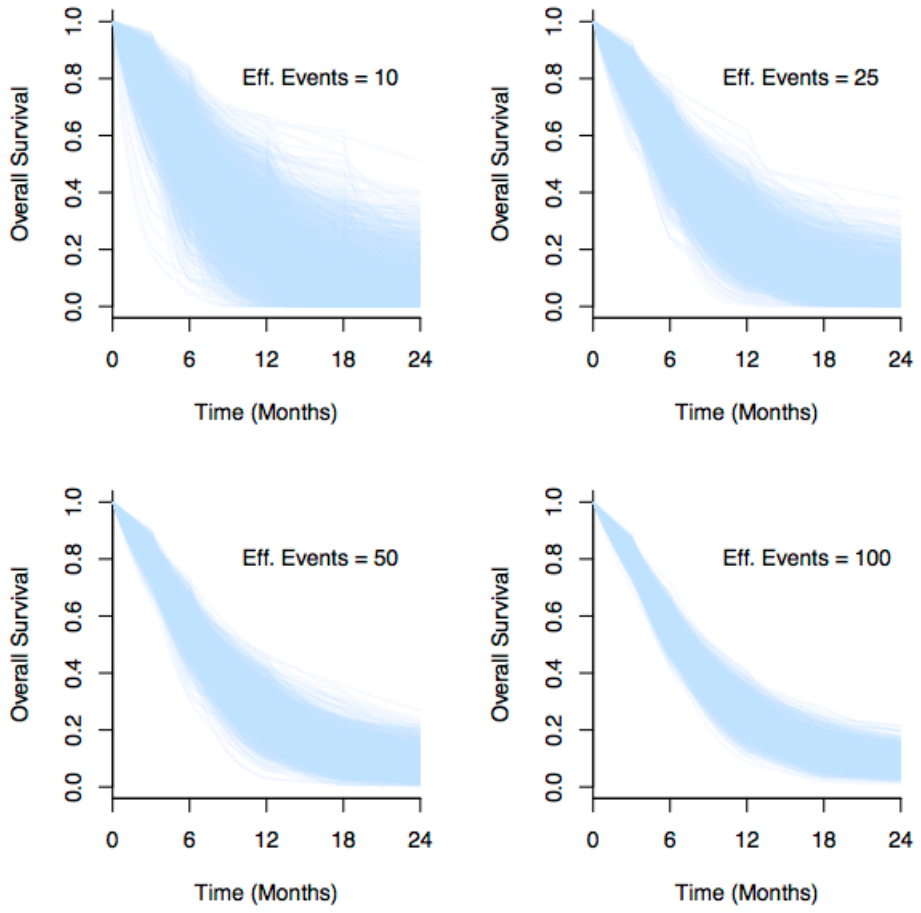

Figure 7.3: Illustration of the survival functions obtained from informative baseline hazard priors

For the purposes of this analysis, $\zeta = 50$ effective events are chosen to define the prior distributions, being a scenario which will well accentuate the effect of the informative priors without being overly influential. The data from the GemCap trial are re-analysed using a PEM. Two models are applied, a 'reference' model with vague uninformative priors and a model incorporating an informative baseline hazard prior. Figure 7.4 shows the prior densities for all parameters along with the posterior densities for the reference model and the model with informative priors.

The effects of the informative priors is shown to reduce the posterior estimate of each

log baseline hazard parameter which is as expected as the prior estimate of the survival function is consistently higher than the Kaplan Meier estimate for the Gemcitabine arm as shown in Figure 7.2. Each baseline hazard parameter is also estimated with a greater degree of precision. The log hazard ratio also has a point estimate which is closer to zero, again as expected as Figure 7.2 illustrates that the prior information will 'drag' the survival curve for the control arm closer to the experimental arm. More importantly, the precision of the estimate of the log baseline hazard has increased, without any prior information being included on this parameter.

The full set of parameter estimates from each model are given in Table 7.2. Here there are some results which may be slightly non-intuitive. As an example, the posterior distribution for ϖ_4 with an informative prior has mean -2.25 which may be unintuitive given the reasonably good agreement between the prior distribution and the data respectively (-2.10 and -2.12). Upon further inspection, it is clear that there is some correlation between the baseline hazard parameters and that all baseline hazard parameters must be considered collectively. It may also be noted that an informative baseline hazard function has the effect of smoothing the baseline hazard function. This can be shown graphically.

Lastly, the effect on the log hazard rate is again noted with some shrinkage towards zero, in agreement with Figure 7.4. Also associated with the shrinkage is an increase in the parameter precision as the standard deviation decrease from 0.09 for the reference model to 0.06 for the informative model.

Parameter	Time-grid	$\boldsymbol{\varpi}$	'Reference'	Informative ($\zeta = 50$)
$log(\lambda_1)$	3	-2.60	-2.36 (0.17)	-2.55 (0.10)
$log(\lambda_2)$	6	-2.08	-1.91 (0.17)	-2.07 (0.11)
$log(\lambda_3)$	12	-2.03	-1.84 (0.16)	-1.99 (0.11)
$log(\lambda_4)$	18	-2.10	-2.12 (0.20)	-2.25 (0.15)
$log(\lambda_5)$	42	-2.59	-2.33 (0.26)	-2.52 (0.21)
β			-0.15 (0.09)	-0.06 (0.06)

Table 7.2: Results of applying informative baseline hazard priors to the analysis of GemCap

Aside from the results of parameters in Table 7.2, Figure 7.5 shows the resulting survival functions that are obtained from the reference and informative models. In each case, the Kaplan Meier survival estimates for the Gemcitabine arm are included for reference. These show explicitly how the survival function is increased and is estimated with a larger precision compared to the reference model. Note that the Kaplan Meier estimates do not agree entirely with the survival estimates obtained in the reference model, this is as the model estimates are also influenced by the experimental arm as well as the control arm and that some evidence of non-proportionality can distort this estimate.

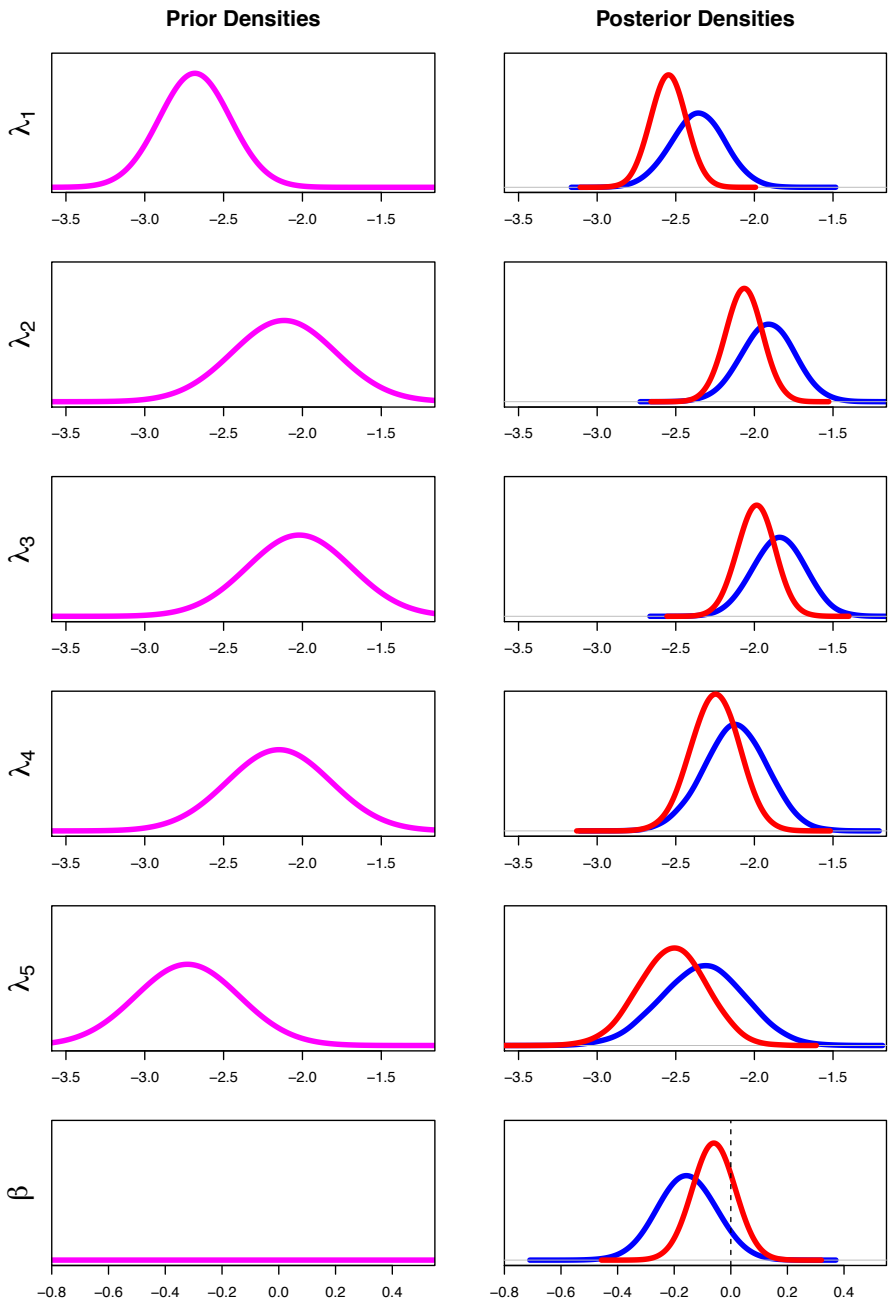

Figure 7.4: Illustration of prior and posterior densities for a selection of parameters for the analysis of GemCap data

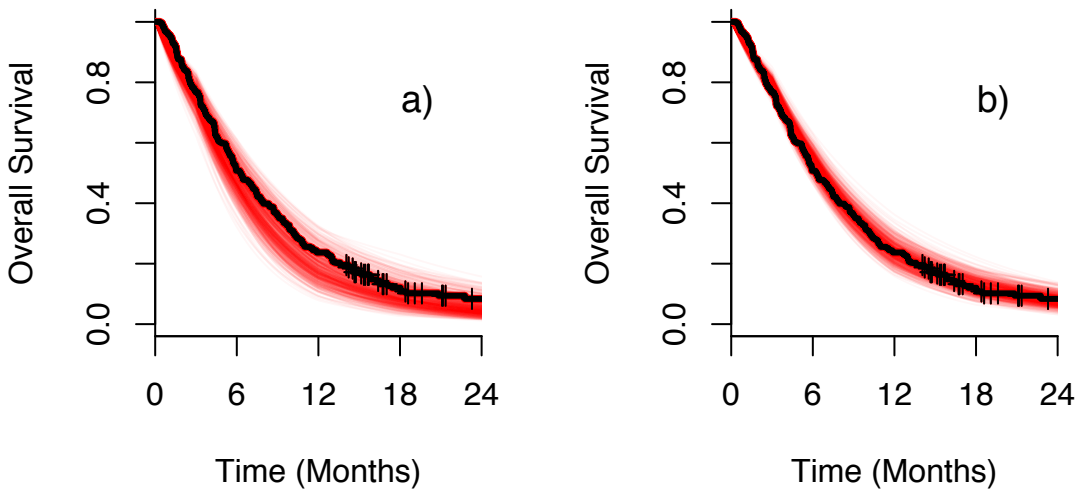

Figure 7.5: Illustration of fitted survival function for a) vague and b) informative prior distributions

7.5 Local step and trapezium priors

Here local step and trapezium distributions are introduced for use in clinical trial design and analysis. Previous efforts for the use of non-standard prior distributions in trial design have been explored by Cook et al. [207], Fuquene et al. [208] and Hobbs et al. [209]. Here these priors are termed as being 'robust' in that they have less influence when there is disagreement between the priors and the observed data. Hobbs et al. in particular present commensurate priors where the influence of the priors is discounted when there is disagreement with the data. Locally flat priors by contrast are only uninformative when there is broad agreement between prior information and the observed data. The aim of locally flat priors is then to discourage unlikely solutions. This approach may be particularly useful when data are sparse or expensive to obtain and the user wishes to encourage a set of likely solutions.

The local step and trapezium priors are characterised by being flat uninformative only within some given bounds. The motivation for these priors lies in the fact that prior to a trial taking place, previous information and/or expert opinion may reliably show a parameter to fall within an interval but that a point estimate may be more difficult to obtain. These are both applied to priors on the baseline hazard function.

'Step' prior distribution

The step distribution is characterised by solutions within some inner bounds having a greater density than those outside of the inner bounds. Outer bounds define the upper and lower limits outside of which the probability density is zero. The step distribution

requires five parameters to be set (a, b, c, d, p). The extreme points of the distribution, a and d give the outer bounds, the inner limits are given by b and c. The result is a step type distribution where a predefined proportion 'p' of the distribution lies between b and c. Formally the distribution is defined as:

$$\pi_{step}(x) = \begin{cases} 0 & \text{if x} < \text{a} \\ \frac{(1-p)(b-a)}{[(b-a)+(d-c)]^2} & \text{if a} \leq \text{x} < \text{b} \\ \frac{p}{c-b} & \text{if b} \leq \text{x} \leq \text{c} \\ \frac{(1-p)(d-c)}{[(b-a)+(d-c)]^2} & \text{if c} < \text{x} \leq \text{b} \\ 0 & \text{if x} > \text{d} \end{cases}.$$

Note that for a symmetrical distribution, the density between the intervals $[a, b)$ and $(c, d]$ simplifies to $\frac{1-p}{2(b-a)}$. Figure 7.6 shows the behaviour of the step distribution with the given limits. Here the limits are set to produce a symmetrical distribution but this need not be the case.

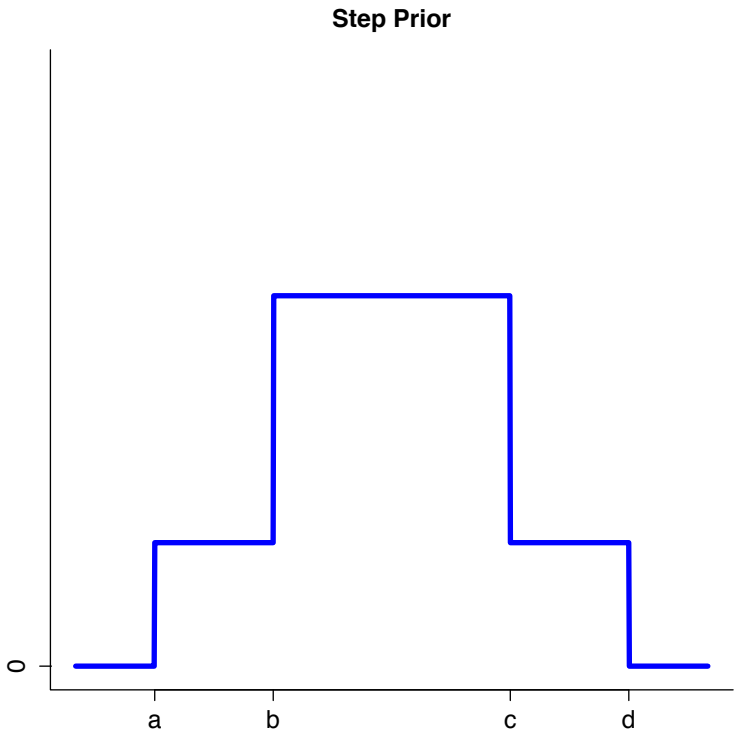

Figure 7.6: Illustration of the behaviour of the Step distribution

'Trapezium' prior distribution

A second approach is to define a trapezium distribution. Here the same limits (a, b, c, d) are defined although p is no longer required. The distribution is similar in its properties to the 'Step' distribution above but differs in that the densities between the points (a,

b) and (b, c) reduce at a linear rate as opposed to a stepwise fashion. Define the density for x under this distribution as

$$\pi_{trap}(x) = \begin{cases} 0 & \text{if x} < \text{a} \\ \frac{(x-a)}{\delta(b-a)} & \text{if a} \leq \text{x} < \text{b} \\ \frac{1}{\delta} & \text{if b} \leq \text{x} \leq \text{c} \\ \frac{(d-x)}{\delta(d-c)} & \text{if c} < \text{x} \leq \text{d} \\ 0 & \text{if x} > \text{d} \end{cases}$$

Figure 7.7 show the behaviour of the trapezium prior with the same limits set as for the step-prior.

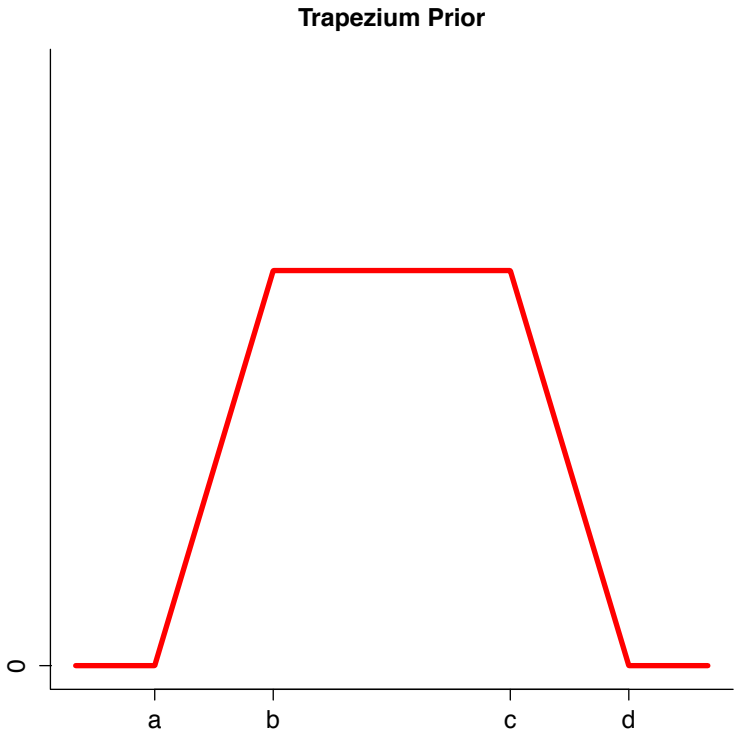

Figure 7.7: Illustration of the behaviour of the Trapezium distribution

7.5.1 Survival analysis with various prior distributions

Here, uninformative (reference), normal, local step and trapezium priors are applied to the analysis of the GemCap data presented in Section 7.3. Parameters are set for the locally flat priors with the inner and outer bounds, (a, b, c, d), derived from the $0.005, 0.1, 0.9$ and 0.995 quantiles respectively of the normal distributions used to define the prior distributions $Pr(\boldsymbol{\lambda})$ in the analysis of the GemCap data in Section 7.3.

The results are presented in Table 7.3 and show that similar, although less accentuated results are obtained for local step and trapezium priors than for priors based on normal distributions. This is shown by both the reduced effect on the point estimate

and the smaller increase in precision for all baseline hazard parameters as well as the log hazard ratio. Comparing the step and trapezium priors, the step distribution model has the greater influence on model parameters, the standard error of the hazard ratio for the step model and trapezium model are 0.06 and 0.07 respectively. Even so, the diminished effect of the trapezium prior distribution still offer a notable increase in the precision of the posterior distributions of the estimates obtained from the reference model.

Parameter	Time-grid	ϖ	'Reference'	Normal	Step	Trapezium
$log(\lambda_1)$	3	-2.60	-2.36 (0.17)	-2.55 (0.10)	-2.53 (0.11)	-2.50 (0.11)
$log(\lambda_2)$	6	-2.08	-1.91 (0.17)	-2.07 (0.11)	-2.05 (0.11)	-2.03 (0.13)
$log(\lambda_3)$	12	-2.03	-1.84 (0.16)	-1.99 (0.11)	-1.98 (0.12)	-1.95 (0.13)
$log(\lambda_4)$	18	-2.10	-2.12 (0.20)	-2.25 (0.15)	-2.25 (0.16)	-2.23 (0.17)
$log(\lambda_5)$	42	-2.59	-2.333 (0.27)	-2.52 (0.21)	-2.53 (0.21)	-2.47 (0.24)
β			-0.15 (0.09)	-0.06 (0.06)	-0.07 (0.06)	-0.08 (0.07)

Table 7.3: Results of applying informative baseline hazard priors to the analysis of GemCap with locally flat priors

The results presented are as to be expected as the priors used are only informative for solutions that do not agree with the prior information. In general, less information will be included into a model when locally flat priors are used as the priors are only influential when they disagree with the data that have been observed. It should be noted here that these priors differ from the robust priors proposed by Cook et al. [207]. The local step and trapezium priors proposed here are uninformative when the data agree with the priors but discourage the data when they do not agree with prior estimates.

The use of local step and trapezium priors may be particularly attractive to clinicians as they encourage solutions which are in keeping with current medical thinking. The aim of these priors therefore is not to estimate one solution on which to base prior distributions but to define a range of likely solutions. Taking for example the design of a study with a time-to-event endpoint, clinicians may be nervous in accurately estimating a survival function for the control arm before the trial has commenced. More confidence may be given to statements such that it is not expected that all patients will die within the first month following the trial opening or that the median survival for a certain group will be within a given set of bounds. Locally flat priors aim to discount solutions by stating that parameter estimates outside the (a, d) limits are not possible. Further, any solutions that lie outside of the (b, c) bound, but within (a, d), are justified by arguing that any solution outside of these bounds would be met with scepticism by the general medical community.

7.6 Bayesian design of the ViP study

Here a Bayesian design of the ViP trial with the inclusion of informative priors on the baseline hazard function is considered.

The target population for the trial is the same as for the three trials that are illustrated in Figure 6.1, with the GemCap trial in particular being administered in the same trials unit. ViP, in a similar fashion to the previous trials, also includes a control arm which is Gemcitabine alone. Furthermore, the data shown in Figure 6.1 were also used in the trial design, informing the sample size calculation.

To illustrate the effect of the informative baseline hazards, two approaches are taken. Firstly, a fully Bayesian sample size technique is followed based on the sampling methodology of Wang and Gelfand [178] and De Santis, [211]. Secondly an approach is taken whereby the main efficacy parameter of interest is assumed fixed at pre-determined values. The purpose of this second approach is to obtain quantities similar to the frequentist Type I and Type II error rates for comparison with the initial trial design.

7.6.1 Bayesian sample size for ViP

The Average Length Criterion (ALC) is chosen as the utility function on which to base sample size calculations and it is define a-priori that a posterior length of 0.6 is of interest to obtain a coverage of 90%. Prior point estimates are obtained using the estimates that are given in Table 7.1 along with the survival estimates that are obtained from the GemCap trial.

Data are sampled using the marginal distribution of the Bayesian PEM as shown in Section 6.2. Prior variability is defined using the effective number of events approach and effective number of prior events are set as $\zeta = 10, 20, 30$ and 50. Design priors are set from Normal distributions with the most informative log baseline hazard priors, ($\zeta = 50$) along with a prior distribution for the log hazard ratio of $N \sim (\log(0.6), 0.5)$. This is chosen to replicate the initial design parameters of the ViP trial. For each sampled dataset, administrative censoring is applied to any survival time greater than 24 months. Patterns of censoring are obtained using the same methods as Section 3.4 and Section 6.4.

Data are simulated for total sample sizes of 60 to 150 by increments of ten. The resulting ALCs from each set of simulations, for each model are shown in Figure 7.8. These show the resulting ALC estimates obtained from varying sample sizes for uninformative priors and informative priors based on the four effective event scenarios described above. This shows how the behaviour of the ALC criterion alters depending on the prior distributions set. Including prior information improves the behaviour in all cases. Improvement is negligible for the less informative of the locally flat priors

however. Normal priors consistently out perform the locally flat priors in passing the 0.6 threshold at smaller sample sizes.

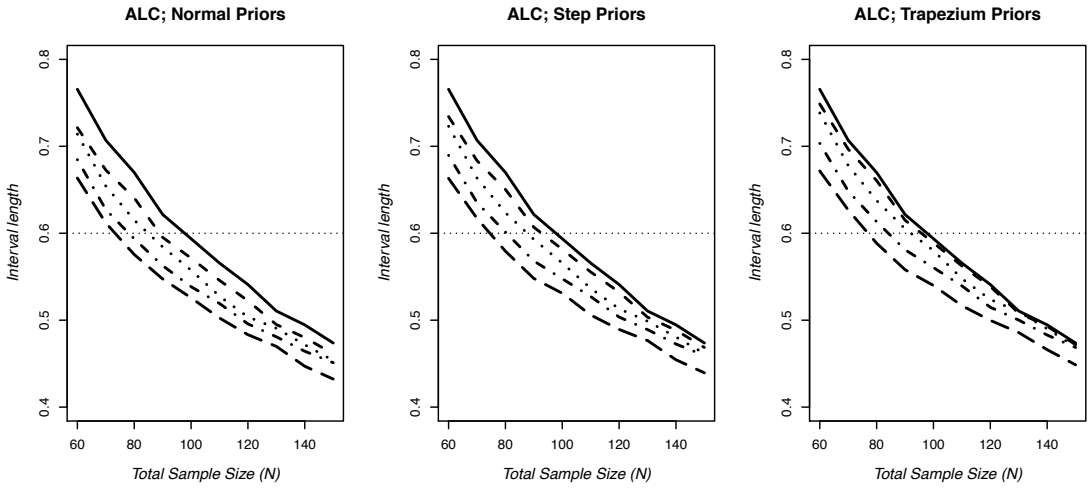

Figure 7.8: Figure to show the performance of the ALC for normal, step and trapezium prior distributions

The results are given in Table 7.4. As a reference model, where the priors remain uninformative, 98 patients are required on average to ensure that a length of 0.6 will contain 90% of the posterior distribution. This is smaller than the 120 patient required for the frequentist design which is in part due to the differing approaches of the two methodologies as a Bayesian approach attempts only to control some aspect of the posterior distribution whereas frequentist approaches by contrast attempt to control against two types of error. Some disparity is also expected based on the parameters chosen on which to base the Bayesian design.

Sample size estimates show that as more information enters the model through the priors, smaller numbers of patients are required to control the width of the posterior distribution. In the most extreme case, 74 patients are required to obtain an ALC of 0.6. Again, this effect is accentuated for normal priors compared to the locally flat alternatives. Considering the locally flat priors, the Step prior has a larger effect than the Trapezium prior.

	Effective Sample Size			
	$\zeta = 10$	$\zeta = 20$	$\zeta = 30$	$\zeta = 50$
Normal	92	86	78	74
Step	94	88	79	74
Trapezium	97	92	84	76

Table 7.4: Sample size estimates for the ViP trial under various differing priors on the baseline hazard function under the ALC.

Both Figure 7.8 and Table 7.4 show that smaller sample sizes are obtained for the

Normal priors in comparison to the locally flat priors in terms of the ALC. The locally flat priors may still be preferred in practice as they may be easier to derive and can inform a trial design and analysis without a-priori setting a point estimate for the most likely solution.

7.6.2 Bayesian type I and type II error rates

To evaluate Bayesian Type I and Type II error rates, a Successful Trial Criterion (STC) is utilised. In the context of the ViP trial, according to the initial design parameters, the trial is a success only if $\geq 90\%$ of the posterior distribution is less than zero. To evaluate Bayesian Type I and Type II error rates, two special conditions of the STC are considered where the design priors for β are set to fixed values of 0 and δ respectively. Specifically setting $\beta = 0$ and calculating the STC will give the Type I error rate and equivalently $\beta = \delta$ for a Type II error rate. Whilst from a Bayesian perspective, sampling from a distribution where the key parameter of interest is considered fixed is inappropriate, this method allows estimation of quantities analogous to the frequentist Type I and Type II error rates.

To ensure that reliable estimates of Type I and Type II errors are obtained, 2000 datasets are simulated following the same procedure as in Section 5.1 but with a fixed sample size of 120 patients to replicate the initial ViP design. The aim here is to show the 'error rates' can be improved over the proposed design as opposed to searching for a sample size based on controlling Type I or Type II error rates.

The results are given in Table 7.5 and show that for the reference model, design parameters similar to that for the ViP trial are obtained. As with sample size calculations based on the ALC, as more information enters into the design through informative priors, the Type I and Type II error rates improve. Considering prior distributions based on normal distributions, for the most informative priors Type I and Type II error rates of 0.07 and 0.08 respectively are obtained. Again, the effect is lessened for locally flat priors with only the most informative priors having any noticeable effect on the Type II error rates.

It is also of interest to note that there is a plateau in the effect that increasingly informative priors have. This is due to the reasoning that as more information enters the prior distributions through effective events, the further information that obtained from the events in the control arm during the course of a trial is reduced. Design parameters here become more dependent on what is observed in the experimental arm as the data form the control arm contribute less towards the estimate of the log hazard ratio.

There is little effect on the Type I error rates, showing when data are simulated with the efficacy parameter fixed at zero the effect of falsely concluding that a new therapy is superior does not change.

Priors	Error	Effective Prior Events 10	20	30	50
Reference	Type I	0.12	0.11	0.12	0.11
	Type II	0.10	0.10	0.11	0.12
Normal	Type I	0.11	0.09	0.10	0.09
	Type II	0.08	0.07	0.07	0.07
Step	Type I	0.11	0.10	0.10	0.10
	Type II	0.09	0.07	0.07	0.08
Trapezium	Type I	0.11	0.10	0.10	0.10
	Type II	0.10	0.08	0.08	0.08

Table 7.5: The effect of different priors and effective prior events on Bayesian Type I and Type II error rates

7.7 Discussion

In this chapter, a method by which summary information on the survival rates can be taken from previous trials or expert information and incorporated into the design and analysis of clinical trial data with a time-to-event endpoint has been introduced. As an example, the GemCap trial carried out at the Liverpool Cancer Trials Unit is used. It was shown that increased precision in the log hazard ratio can be achieved. Though in this instance this also results in some shrinkage towards the 'null point' of no difference.

It is argued here that the main interest in any trial with at time-to-event endpoint is the (log) hazard ratio, however this quantity may only be deemed as clinically important if the survival rates in the control arm agree with the current medical thinking. Take for example the situation of a positive result showing the experimental arm to be an improvement over a control arm, but where survival probabilities in the improved experimental arm do not show any improvement over previously published data. In situations such as these, it is not immediately clear whether the within trial comparison should take precedence and an important difference declared or whether the results on the new therapy should be compared against other available information on patient performance.

In some way at least, the results of a single trial are always going to be compared against other trials or respective data available to the medical community. The Bayesian methods here allow for that information to be formally incorporated into the design and analysis and can therefore further inform the clinical decision making process.

Further introduced were the local step and trapezium priors, which penalise solutions which do not agree with prior information but have no effect when the data agree with previous evidence. It is argued that these priors may be particularly attractive as they do not require a single most likely solution be defined a-priori, rather that a set

of solutions within given bounds all deemed to be equally likely are defined. They may be of particular use in situations where data are sparse or expensive to collect and can therefore be used to encourage likely solutions without being overly influential.

The step and trapezium priors and the normal priors are applied to the design of the ViP trial and show that by incorporating prior information on the baseline hazard function, smaller sample sizes based on the average length of the posterior distribution of the log hard ratio are obtained. Considering trial design on the basis of Type I and Type II error rates, small improvements are also observed when trial designs incorporate informative prior distributions.

There have been recent methodological advances in the incorporation of historical information into the design of clinical trials, most notably the use of commensurate priors and power priors. Despite this however, there is still yet to be many examples of these approached being used in practice. This chapter introduced some of the practical steps that must be considered in the design of a clinical trial with a time-to-event outcome which incorporates historical information. In the next chapter the methodology presented here is extended into trial design to investigate the possibility of deviation away form the standard 1:1 allocation ratios that is common in randomised controlled trials.

Chapter 8

Unequal Allocation Ratios in a Bayesian and Frequentist Framework

8.1 Introduction

This chapter is concerned with the optimal allocation of patients to treatment arms in a two-arm clinical trial using reliable prior information. Initially a review is given on the allocation ratios that are used in practice. Following this, some theoretical results are obtained for binary and continuous outcomes where prior information is available. Analytical results for the optimal allocation ratios are obtained and applied to simple examples.

The use of informative priors for survival outcomes are investigated for a standard exponential and a PEM. An analytical form for the optimal allocation ratio is derived based on the assumption of all patients having equal follow-up.

8.2 The use of unequal allocation ratios in practice

Whilst the use of unequal allocation ratios is not new, they are relatively rare. Much is made over the statistical preference for equal allocation ratios with Schultz and Grimes [212] going as far as to claim that an equal allocation ratio itself can be treated as an endpoint in a trial.

Whilst there are some examples of unequal allocation ratios used in practice, literature devoted to this topic is relatively rare. A review by Dumville et al. [213] carried out in 2006 identified 65 trials that have been carried out using unequal allocation ratios. Reasons for unequal allocation were identified as cost, further arguments for which are given by Vozdolska et al. [214] and the need for a learning curve with newer experimental therapies/treatments. Ethical issues were highlighted as a possible reason for unequal allocation, a point further made by Alvin [215] who argues that 1:1 allocation should only be used when there is true equipoise between two treatments.

Lastly Dumville et al. consider unequal allocation ratios for 'other' conditions such as multi-arm trials or trials which wish to compensate for an inflated drop-out rate for one particular treatment.

What is highlighted is the lack of any statistical justification for the use of allocation ratios other than 1:1. Some research has been carried out on the use of unequal allocation ratios for binary endpoints, most notably Raudenbush [216], Raudenbush and Li [217] and Raudenbush et al. [218] who consider optimal designs for binary endpoints for cluster randomised trials, multisite trials where the treatment effect is expected to vary across sites and under an adaptive randomisation method respectively. Under a Bayesian perspective, Brooks [219] investigates optimal allocation about an odds ratio for trials with a binary endpoint.

Throughout the remainder of this chapter, it is assumed that prior information may be available for each arm of a two-arm trial individually and not on the main efficacy parameter of interest. For instance, the prior interpretation might be on the mean in each arm of a trial but not on the difference in means.

Note that throughout this chapter, the notation shall be used whereby N is the total number of patients randomised to a trial with n_1 and n_2 being the number of patients randomised to each treatment arm. The proportion of patients randomised to arm one is given by γ, i.e. $n_1 = \gamma N$, $n_2 = (1-\gamma)N$.

8.3 Optimal allocation ratios under Bayesian analysis

In this section some investigation into optimal allocations for a variety of different outcomes is explored. The overall aim for each outcome is to optimise the amount of information that can be gained on the key efficacy parameter of interest, δ. Calculations are then presented based on minimising the associated standard error. For each outcome, the methods are then extended into a Bayesian framework allowing for prior distributions to be placed on each arm individually.

8.3.1 Normal outcomes

Assuming that data from each arm of a two arm trial are summarised by the associated means, define the key efficacy parameter of interest to be the difference between means such that

$$\delta = \mu_2 - \mu_1$$

with an estimated standard error given by

$$se(\delta) = \sqrt{\sigma^2/n_1 + \sigma^2/n_2}.$$

For simplicity it is assumed that there is equal variation in each treatment arm. The estimated standard error from a future trial with N total patients is then given by

$$se(\delta) = \sqrt{\sigma^2/N\gamma + \sigma^2/N(1-\gamma)}.$$

It is straight forward to show by squaring both sides and differentiating with respect to γ that the smallest estimate of $se(\delta)$ will be obtained by a value of $\gamma = 0.5$.

Extending into a Bayesian framework, it is now assumed that informative priors are available for each treatment arm individually. Let $\{x_{ij}, i = 1, 2; j = 1, ..., n_i\}$ be the observations that will be obtained from the two arms of the trial. Note that the first treatment arm will be designated as the control arm when appropriate. For the convenience of conjugate Bayesian analysis, it is assumed that the prior information for the mean of the control arm is given by

$$\pi(\mu_1) \sim N(m_1, \tau_1)$$

and likewise for the prior distribution associated with μ_2. Taking the variance of μ_i as fixed, the posterior distribution for μ_i is a normal distribution when the data and the prior information both follow a normal distribution. A form for the posterior distribution, given the data, D, is given as

$$\pi(\mu_i|D) \sim N\Big(\frac{\mu_0/\tau_i^2 + \sum_{j=1}^{n} x_{ij}/\sigma^2}{(1/\tau_i^2 + n/\sigma^2)}, (1/\tau_i^2 + n_i/\sigma^2)^{-1}\Big),$$

see for example [21] . The difference between the means of the two posterior distributions, δ_π, has a standard error given by

$$se(\delta_\pi) = \sqrt{(1/\tau_1^2 + N\gamma/\sigma^2)^{-1} + (1/\tau_2^2 + N(1-\gamma)/\sigma^2)^{-1}}.$$

Squaring both sides and differentiating with respects to γ, an expression for the allocation ratio that minimises $se(\delta_\pi)$ is given by

$$\gamma = \frac{\sigma^2(\tau_1^2 - \tau_2^2) + N\tau_1^2\tau_2^2}{2N\tau_1^2\tau_2^2}.$$

Note here that if vague uninformative priors are placed on each treatment arm such that $\tau_1^2 = \tau_2^2 \to \infty$ then a solution of $\gamma = 0.5$ is obtained as would be expected. Taking the example of informative priors for a single arm only, allowing $\tau_2^2 \to \infty$ gives

$$\gamma = \frac{\sigma^2 + N\tau_1^2}{2N\tau_1^2}. \tag{8.1}$$

It is of interest to note that the optimal precision of δ_π here is not dependent on the prior point estimates, m_1 or m_2. As may be expected it is dependent upon the total number of patients that are to be randomised.

Table 8.1 gives the allocation ratio for the number of patients to be allocated to the control arm under differing degrees of prior information. These results are based on standardised data, i.e. $\sigma^2 = 1$. Table 8.1 also illustrates the condition that as the total

number of patients increases, the optimal allocation ratios tend towards $\gamma = 0.5$. This is to be expected as when data sets increase in size, posterior distributions are more dependent upon the observed data than they are on the prior distributions.

τ_1^2	τ_2^2	N = 20	N = 50	N = 100	N = 250
0.05	0.1	0.25	0.40	0.45	0.48
	0.25	0.10	0.34	0.42	0.47
	0.5	0.05	0.32	0.41	0.46
	∞	0.00	0.30	0.40	0.46
0.1	0.25	0.35	0.44	0.47	0.49
	0.5	0.30	0.42	0.46	0.48
	∞	0.25	0.40	0.45	0.48
0.25	0.5	0.45	0.48	0.49	0.50
	∞	0.40	0.46	0.48	0.49

Table 8.1: Optimal allocation ratios for standardised data under differing prior distributions and varying total sample sizes.

Figure 8.2 gives a contour plot to display the behaviour of the optimal allocation function for Normal outcomes under differing total sample sizes and the variability about the prior distribution for μ_1. This shows how for smaller sample sizes, the effect of informative prior distributions is more pronounced. Beyond sample sizes of 200 say, only prior distributions with a high degree of precision will have any meaningful effect on deviation away from the 1:1 allocation ratio.

Example

As a simple example, consider a two arm trial comparing a new therapy against current clinical practice. The primary outcome of the study follows a normal distribution, and previous data for the clinical practice arm have shown that an estimated outcome, μ_1, follows a normal distribution such that $\mu_1 \sim N(100, \sqrt{10})$. A previous estimate for the standard deviation for the population of interest is $\sigma = 20$. For simplicity, this parameter shall be treated as fixed though there is no requirement for this assumption. It is determined that a sample size of 100 shall be used. Using (8.1), an optimal allocation ratio of 0.35 is obtained.

As a method of verification, Bayesian sample size estimates were carried out under a variety of different allocation ratios using the average length criterion (ALC) as proposed by Joseph and Belisle [175] as the utility function of choice. This is set up to measure some fixed quantity of the posterior distribution, for example the 95% credibility interval. The results of the design simulations are given in Figure 8.2 and show general agreement with the lowest ALC given for a proportion of 0.35.

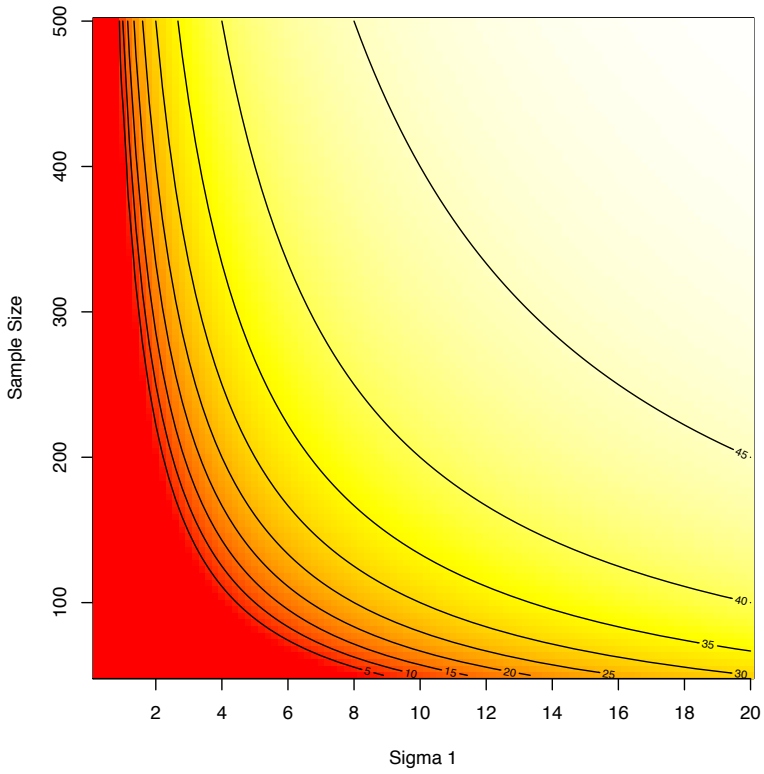

Figure 8.1: Contour plot to show optimal allocation ratios for differing total sample sizes and estimates of prior variability for the control arm τ_1

8.3.2 Binary endpoint

Sample size calculations for a binary endpoint have been considered in detail by Sambucini [220] and M'Lan et al. [174] although neither consider unequal allocation ratios. Under a binary endpoint, it is assumed that a trial of two treatments is carried out to assess the number of successes and failures in each arm. Given the probability of a success in each treatment arm as p_1 and p_2 respectively, the distribution of data is given in Table 8.2.

Treatment	Success	Failure	Total
Trt 1	$k_{11} = N\gamma p_1$	$k_{12} = N\gamma(1-p_1)$	$N\gamma$
Trt 2	$k_{12} = N(1-\gamma)p_2$	$k_{22} = N(1-\gamma)(1-p_2)$	$N(1-\gamma)$
Total	Np	$N(1-p)$	N

Table 8.2: Table to display the distribution of data for a two-arm trial with a binary endpoint

Let k_{j1} be the number of successes in treatment arm j and k_{j2} be the number of failures. The key efficacy parameter is then given by the odds ratio defined by

$$OR = \frac{k_{11}k_{22}}{k_{12}k_{21}}.$$

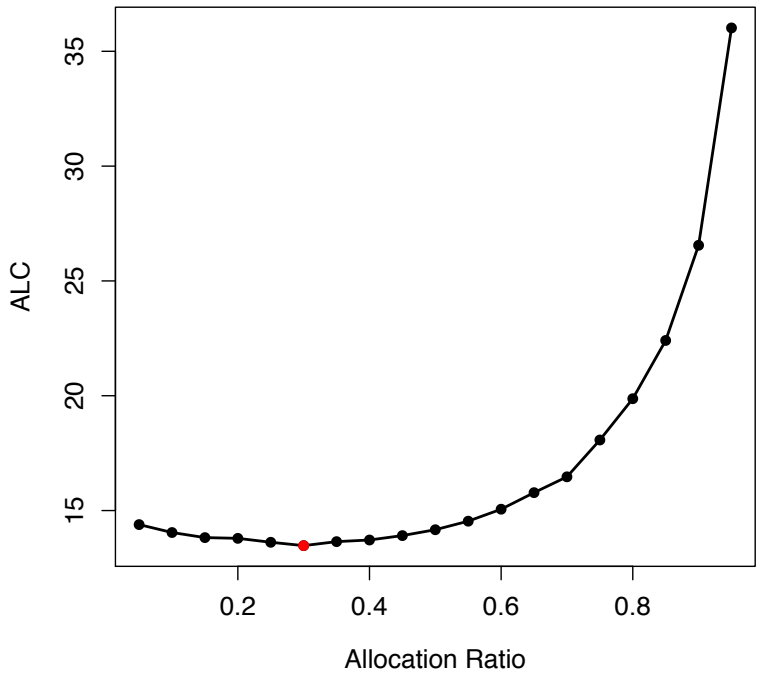

Figure 8.2: Figure to demonstrate the behaviour of the ALC design criterion under differing allocation ratios for a fixed sample size of 100 patients.

Other measures of efficacy such as the risk difference or the relative risk are not considered here. Measuring the odds ratio on the log scale, an estimated associated standard error for a future trial is given by

$$se(log(OR)) = \left\{ \frac{1}{N\gamma p_1} + \frac{1}{N\gamma(1-p_1)} + \frac{1}{N(1-\gamma)p_2} + \frac{1}{N(1-\gamma)(1-p_2)} \right\}^{1/2}.$$

Squaring both sides and differentiating with respect to γ gives an optimal allocation ratio of

$$\gamma = (1+A)^{-1}, \tag{8.2}$$

where

$$A = \left\{ \frac{p_1(1-p_1)}{p_2(1-p_2)} \right\}^{1/2}.$$

The behaviour of the allocation ratio based on varying expected response rates in each of the two treatment arms is illustrated in Figure 8.2. Under standard trial design for a superiority trial, it is customary to assume the null hypothesis that $p_1 = p_2$ thereby giving the optimal allocation ratio of 0.5. It is argued here however that it may be of greater benefit to design a trial under the alternative hypothesis. This is as under a positive or borderline positive trial, being one that is close to showing a clinically

important difference, a standard error which is sub-optimal will be obtained if the trial is designed with an equal allocation ratio. Under an equal allocation ratio, the standard error of the log odds ratio will only be optimal when there is no difference between the treatment arms, a scenario where the precision of the log odds ratio is of the least interest.

It is shown by (8.2) that under the alternative hypothesis, an equal allocation ratio will be obtained when the mean success rate, given by $\bar{p} = (p_1 + p_2)/2$, is equal to 0.5. When an increase in the success rate is considered desirable, for $\bar{p}$<0.5 a optimal standard error is obtained when a greater proportion is given the the experimental arm. For $\bar{p}$>0.5 an optimal standard error is obtained when a greater proportion is given to the control arm.

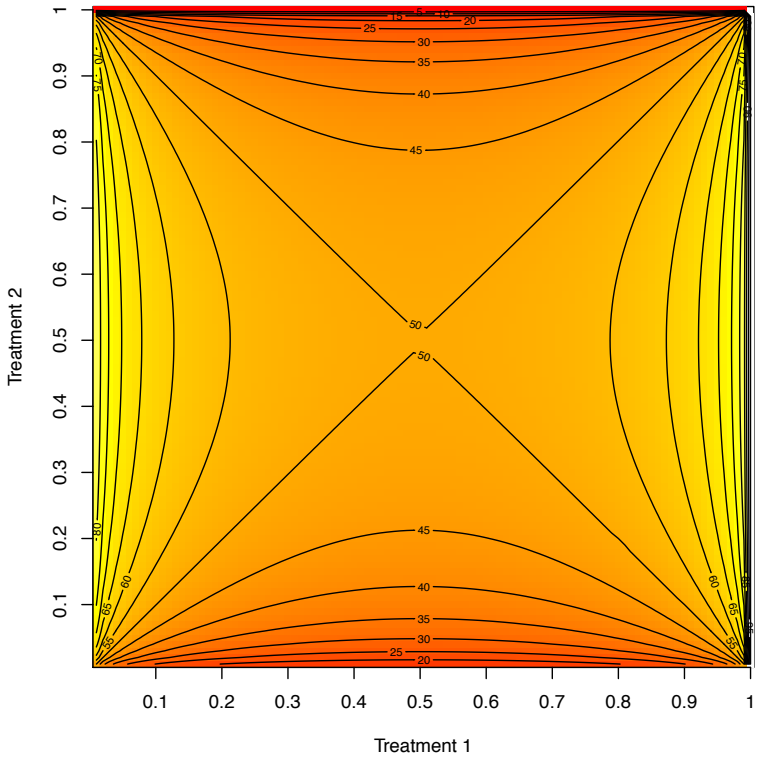

Figure 8.3: Contour plot to show the optimal allocation ratio for a trial with estimated response rates in each arm.

Extending into a Bayesian design and assuming that the likelihood for the number of successes in each arm follows a binomial distribution, a conjugate analysis for a binary endpoint can be obtained by setting a prior distribution using a Beta distribution with parameters α_j, β_j. It is straight forward to show that a posterior distribution for each arm is itself given by a Beta distribution defined by

$$Beta(\alpha_j + k_{j1}, \beta_j + k_{j2}).$$

A log odds ratio under informative priors, denoted $\log\{OR_\pi(\theta|D)\}$ is

$$\log(OR_\pi(\theta|D)) = \frac{(\alpha_1 + k_{11})(\beta_2 + k_{22})}{(\alpha_2 + k_{21})(\beta_1 + k_{12})}.$$

For a future trial with unobserved data, an estimated standard error is obtained via

$$se\{log(OR_\pi(\theta|D))\} = \Big\{\frac{1}{p_1(\alpha_1 + \beta_1 + N\gamma)} + \frac{1}{(1-p_1)(\alpha_1 + \beta_1 + N\gamma)} + \frac{1}{(1-p_2)(\alpha_2 + \beta_2 + N(1-\gamma))} + \frac{1}{(1-p_2)(\alpha_2 + \beta_2 + N(1-\gamma))}\Big\}^{1/2}.$$

Here is is worth noting that a-priori $\alpha_1 \approx p_1(\alpha_1 + \beta_1)$. Squaring both sides and differentiating with respect to γ, an optimal allocation ratio for an analysis with informative priors is obtained when

$$\gamma = \frac{N - A(\alpha_1 + \beta_1) - (\alpha_2 + \beta_2)}{N\Big\{1 + A\Big\}}. \tag{8.3}$$

An illustration of the optimal allocation ratio for binary outcomes with informative priors on the control arm is given in Figure 8.4. Here it is shown that as the sample size increases, allocation ratios tend towards the allocation ratio that are obtained where no prior information is included. Binary outcomes do differ to normal outcomes here as even at reasonable large sample sizes, optimal allocate ratios other than 1:1 will still be obtained.

Note that unlike the optimal allocation for the normal outcome, the point estimate of the response rate is important in determining the optimal allocation ratio. It can be shown that when both $\alpha = \beta = 0$ for both arms of the trial, the optimal allocation ratio formula become (8.2).

As the optimal allocation ratio here depends upon the mean outcome rate itself, there may be some discussion on what value of p_2, being the rate of success in the experimental group, should be set at in terms of trial design. As has been noted, setting $p_2 = p_1$ ensures an optimal standard error for the log odds ratio only when there is no difference between the treatment arms. Conversely, setting $p_2 = \delta$, some minimum clinically relevant difference, would ensure an optimal allocation ratio only when this difference is observed. It may be argued that the most information is required about the efficacy parameter for borderline trials and in this case setting p_2 to some value between p_1 and the clinically relevant difference may be advised.

Example

As an example, a two arm trial with a binary endpoint is considered. As with the previous section, the sample size is fixed at 100. A prior distribution for the mean

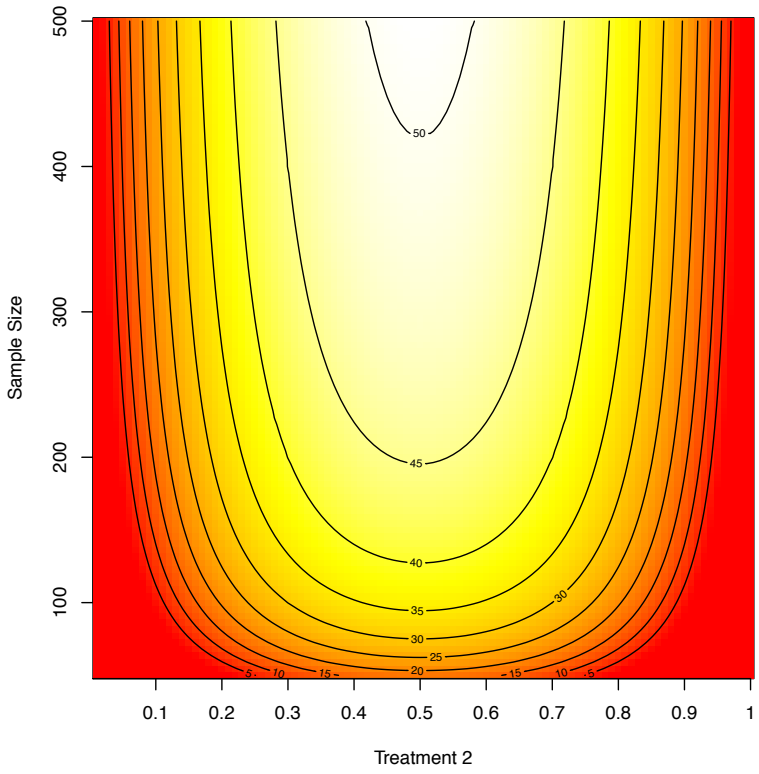

Figure 8.4: Figure to show the behaviour of the optimal allocation ratio for a binary endpoint with different total sample sizes and estimates for the performance of the control arm

response rate in the control arm, p_0 is set as $p_0 \sim Beta(12, 28)$. This is equivalent to assuming a mean (standard deviation) for the response as 0.3 (0.087). A clinically relevant difference is defined a-priori as being an odds ratio of 0.42 which is equivalent of an absolute increase from a response rate of 30% in the control arm to 50% in the experimental arm. Note that as the mean response rate is <50% it is expected that more patients should be randomised to the experimental arm.

From (8.2) an optimal allocation ratio of 0.36 is obtained resulting in 36 patients being randomised to the control arm and 64 patients being randomised to the experimental arm. Bayesian sample size calculation simulations were carried out with data simulated from two binomial distributions with mean proportions equal to 0.3 and 0.5 respectively. The average length criterion of the log odds ratio was used as the utility function of interest. Figure 8.5 shows the agreement here with an allocation ratio of 0.35 providing the smallest ALC on average.

8.3.3 Survival outcomes

Studies with a time-to-event outcome are typically designed with either a log-rank test or some proportional hazards modelling. As the log-rank test is most powerful

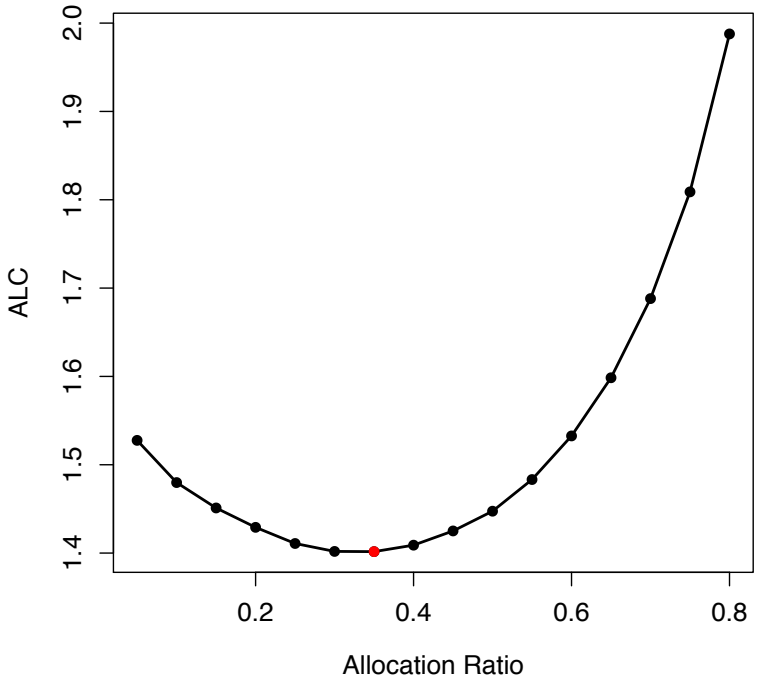

Figure 8.5: Results of the Average length Criterion (ALC) for an example study with a binary endpoint and informative priors on the control arm

under the proportional hazard assumption, the key efficacy parameter of interest is the hazard ratio, ϕ, typically measured on the log scale. An estimate for the standard error of $\log(\phi)$ is given by

$$se(log(\phi)) = \sqrt{(1/E_1) + (1/E_2)},$$

where E_1 and E_2 are the number of events in the control and experimental arm respectively (see for example Parmar et al. [143]). In a trial setting with a fixed sample size, the smallest standard error is obtained when there is an equal number of events in each arm. Under traditional designs with a 1:1 allocation ratio, the smallest standard error will occur with a hazard ratio $\phi = 1$. As with a binomial endpoint, designing a trial assuming the null hypothesis in a frequentist approach ensures an optimal standard error is obtained only when there is no difference between the treatments.

In this context the aim prior to a study beginning should be to distribute the anticipated number of events evenly across the treatment arms. For any single patient on a trial, the probability of observing an event before time t is given by $1 - S(t)$. Let the survival function for the control group be $S_1(t)$ and for the treatment group, $S_2(t)$ and hence

$$S_2(t) = S_1(t)^{\phi}.$$

Without any prior information, the smallest standard error is obtained when

$$\gamma = \left\{1 + \sqrt{\frac{1 - S_1(t)}{1 - S_1(t)^{\phi}}}\right\}^{-1}. \tag{8.4}$$

Figure 8.6 gives a contour plot to illustrates the behaviour of differing allocation ratios based on varying hazard ratios and baseline survival estimates at the point of analysis. This shows that under the null hypothesis $\log \phi = 0$ the optimal allocation is always 0.5. The optimal allocation for values of $\log \phi \neq 0$ differ from 0.5 but it should be noted that even under the most extreme estimates there is relatively little deviation away from 0.5.

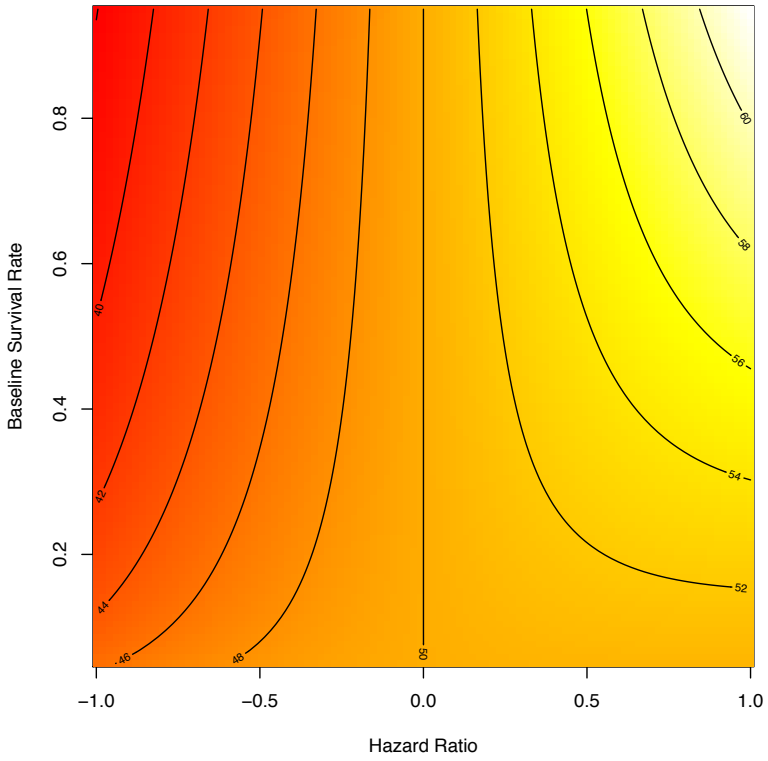

Figure 8.6: Contour plot to show optimal allocation ratios for a trial with a time-to-event endpoint based on baseline survival rates and an assumed hazard ratio

Whilst this is a general result that may be used in any trial with a time-to-event outcome, extending into a Bayesian framework may result in some difficulties since there is no generic conjugate analysis and will only be possible for the simplest of models. For illustration, the Exponential survival model is considered making use of a gamma prior distribution. Extensions to the more flexible Piecewise Exponential Model are included in later in this section.

Exponential model

Here, the survival distribution for the control arm is $S_1(t) = \exp\{-\lambda_1 t\}$. The likelihood is given by $\prod_{i=1}^{N\gamma} \lambda_1^{\nu_i} \exp\{-\lambda_1 t_i\}$, where ν_i is an indicator variable (1 = event, 0 = non-event). The maximum likelihood estimate of λ_1 is $\hat{\lambda_1} = \sum \nu_i / \sum t_i = \eta_i / \varsigma_i$.

Extending to a Bayesian analysis, a conjugate prior distribution is a Gamma(ξ, ϵ) where ξ and ϵ are the scale and shape parameters. In practice, ξ may be selected as the effective prior number of events and ϵ derived to satisfy the a-priori estimate of λ.

The posterior distribution of λ_1 has a probability density which is proportional to $\lambda_1^{\eta_1+\xi_1-1} \exp\{-\lambda_1(\varsigma+\epsilon)\}$. The posterior mean and variance of λ_1 are $(\eta_1+\xi_1)/(\varsigma_1+\epsilon_1)$ and $(\eta_1+\xi_1)/(\varsigma_1+\epsilon_1)^2$ respectively. Thus the prior distribution essential adds ξ_1 events with extra patient risk time of ϵ_1.

Using the delta method, the approximate mean and variance of $\log(\lambda_1)$ are $\log\{(\eta_1+\xi_1)/(\varsigma_1+\epsilon_1)\}$ and $1/(\eta_1+\xi_1)$ respectively. Similar arguments apply for the treatment arm. The difference in $\log(x_2)$ and $\log(x_1)$ leads to the estimate of ϕ with

$$se\{\log(\phi)\} = \{1/(\eta_1+\xi_1) + 1/(\eta_2+\xi_2)\}^{1/2}.$$

Now $\eta_i = N\gamma\{1-\exp(\lambda_i t)\} = N\gamma k_1$, say, under the prior distribution. After some algebra the optimum value of γ is

$$\gamma = \frac{N\sqrt{k_1}\sqrt{k_2} + \frac{\sqrt{k_1}}{\sqrt{k_2}}\xi_1 + \xi_2}{N[k_1 + \sqrt{k_1}\sqrt{k_2}]}. \tag{8.5}$$

It can be seen that setting $\xi_1 = \xi_2 \to 0$ and dividing (8.5) through by $\sqrt{k_1}$ will give equation (8.4).

Example

As an example, again take a trial which is to have a total sample size of one hundred patients. The endpoint for the patient group is taken to be $t^* = 24$, measured in months and it is estimated that patients undergoing the standard therapy will have a survival estimate $S(t^*) = 0.5$. For the new therapy being investigated it is envisaged that a hazard ratio of $\phi = 0.75$ is obtained which is equivalent to an improvement in the survival rate at t^* to 0.59.

Before the trial begins, there is already a source of reliable information on which the survival estimates in the control arm are made and it is considered that this information should represent ten events within the analysis of the trial, ($\xi_1 = 10$). There is no information for the experimental arm ($\xi_2 = 0$). Using (8.5) an optimal allocation ratio of 0.36 is obtained with a greater proportion of patients allocated to the experimental arm.

As the optimal allocation ratio for survival endpoints are more complex, a contour plot is given in Figure 8.7. This shows how optimal allocation ratios are dependent

both on the baseline survival rate and the hazard ratio. Also shown here is the optimal allocation ratio for the scenario described above.

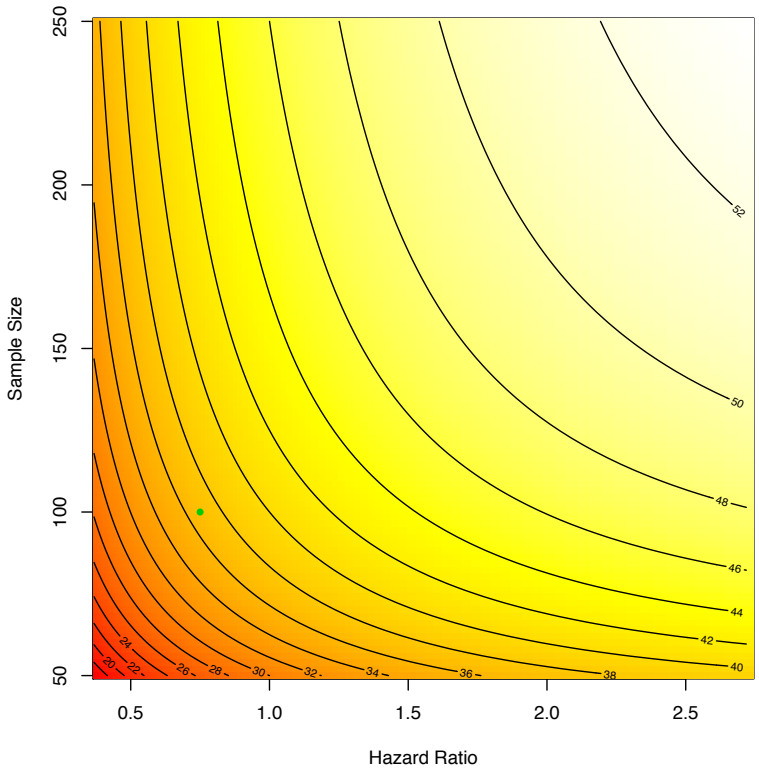

Figure 8.7: Heat map to show optimal allocate ratios dependent on total sample size and trial hazard ratio. Included (green dot) is the optimal allocation ratio for the scenario described above.

As a means of verification, Bayesian sample size calculations were carried out via simulation. The results are displayed in Figure 8.8. Here with a fixed sample size of 100 it is shown that the smallest ALC for the hazard ratio is observed at approximately 0.36 in agreement with what is obtained from (8.6).

Piecewise exponential model

Here the basis behind optimal allocation for an exponential model is expanded into the analysis under a piecewise exponential model (PEM).

Assuming a time grid with J intervals at $a_1, ..., a_{j+1}$, the hazard rate in each interval is defined as $\hat{\lambda^{(j)}} = \eta^{(j)}/\varsigma^{(j)}$ where $\eta^{(j)}$ are the number of events in the observed interval and $\varsigma^{(j)}$ is the total patient time at risk for the j^{th} interval. Note that for the j^{th} interval patient time at risk is limited by $a_{j+1} - a_j$.

For each individual interval, the expected number of events for the control arm is given by

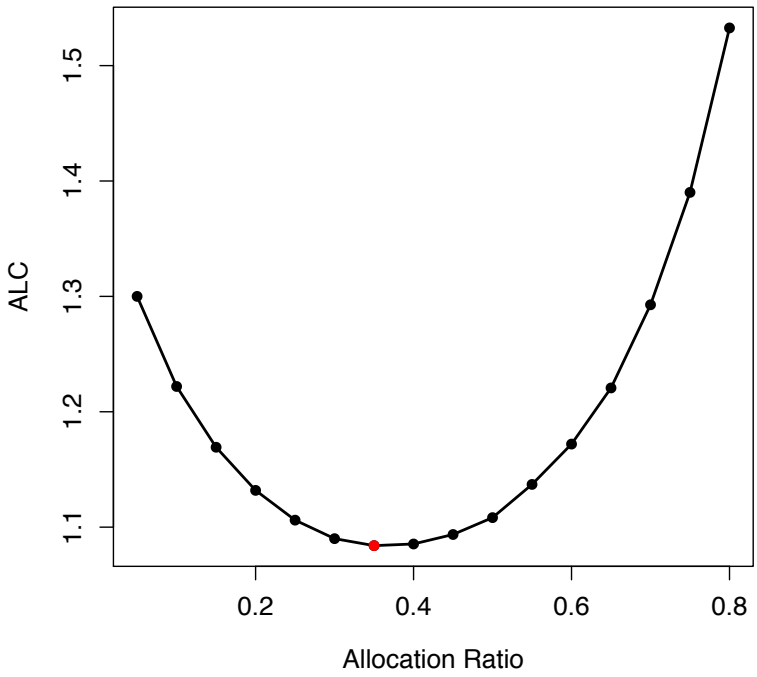

Figure 8.8: ALC estimates obtained from bayesian design simulations. Included (red dot) is the optimal allocation ratio for the given scenario.

$$N\gamma[1-\exp\{-\lambda_j(a_j-a_{j-1})\}].$$

Over all intervals, the total expected number of events obtained is then

$$N\gamma\sum_{j=1}^{J}[1-\exp\{-\eta_j/\varsigma_j(a_j-a_{j-1})\}].$$

Expressing as $N\gamma k_1$ and equivalently $N(1-\gamma)k_2$ then allocation ratios may be obtained as for the exponential distribution.

Extending into a Bayesian framework and allowing each individual partition to have an informative prior based on a Gamma distribution with prior parameters ξ_j and ϵ_j, a posterior distribution for each interval is given as

$$Pr(\lambda_j|D)\sim\Gamma(\eta_j+\xi_j,\varsigma_j+\epsilon_j).$$

The effective number of observed events in the j^{th} interval is then

$$\eta_j+\alpha_j=N\gamma\left[1-\exp\left\{\frac{\eta_j+\xi_j}{\varsigma_j+\epsilon_j}(a_j-a_{j-1})\right\}\right]$$

.

Following similar arguments as for the exponential model and noting that a-priori, each hazard rate λ_j can be considered independent, an estimate of the standard error for the log hazard ratio is

$$se\{\log(\phi)\} = \{1/(\sum_{j=1}^{J}[\eta_{1j} + \xi_{1j}]) + 1/(\sum_{j=1}^{J}[\eta_{2j} + \xi_{2j}])\}$$

where the subscript $1j$ identifies the j^{th} interval of the control arm and $2j$ identifies the j^{th} interval from the experimental arm. Now $\sum_{j=1}^{J}\eta_{1j} = \sum_{j=1}^{J} N\gamma[1-\exp\{\lambda_1 j(a_j - a_{j-1})\}] = N\gamma k_1$ and similarly for the experimental arm. Following the same steps as for the exponential model therefore, an optimum value for the standard error is obtained by

$$\gamma = \frac{N\sqrt{k_1}\sqrt{k_2} + \frac{\sqrt{k_1}}{\sqrt{k_2}}\sum_{j=1}^{J}\xi_1 + \sum_{j=1}^{J}\xi_2}{N[k_1 + \sqrt{k_1}\sqrt{k_2}]}. \tag{8.6}$$

In terms of an optimum allocation ratio therefore, it is only required to know the total number of effective events that are used to derive prior distributions and not what intervals the events fall in.

8.3.4 Accounting for recruitment

It should be noted that the estimation of optimum allocation ratios in survival studies, being inherently dependent on the number of events that are observed are also dependent upon the recruitment rate in a trial, which is denoted as $\alpha(t)$ for the purposes of this discussion.

Following [44], if the recruitment rate is constant over an accrual period, define $\alpha(t) = \alpha$ and let t_α denote the time taken to recruit all the patients into a trial. The trial then continues until $t_{\alpha+f}$, the point at which the last patient recruited has had some pre-determined fixed follow-up time. Then the number of events in the control and treatment arm are $N\gamma\alpha^{-1}\int_{t_f}^{t_{\alpha+f}}\{1-S_1(u)du\}$ and $N(1-\gamma)\alpha^{-1}\int_{t_f}^{t_{\alpha+f}}\{1-S_1(u)^\phi du\}$ respectively.

For the simple exponential model, it is straight forward to show that the expected number of events in the control arm, η_1 is

$$N\gamma\alpha^{-1}\Big[t_\alpha + \lambda_1^{-1}\{\exp(-\lambda_1 t_{\alpha+f}) - \exp(-\lambda_1 t_f)\}\Big],$$

with a similar expression for the experimental arm.

Estimation of the number of events in the piecewise exponential model is less straightt forward as it requires the integration of a complex survival function.

Given a survival function with a given time-grid a;

$$S(t) = \exp\{-[\lambda_j(t - a_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(a_g - a_g - 1)\}.$$

It is assumed that integration is desired between two limits $(t_\alpha, t_{\alpha+f})$. Integration is carried out here by defining the time grid as $a^* = (t_\alpha, a_h, t_{\alpha+f})$ where a_h are the elements of the initial time-grid a that lie between the limits of integration.

A form for $\int_{t_\alpha}^{t_{\alpha+f}} S(t)dt$ is obtained by first noting that for the j^th interval

$$\begin{aligned}\int_{a_j}^{a_{j+1}} S(t)dt &= \exp\Big\{-\sum_{g=1}^{j-1}\lambda_g(a_g - a_{g-1})\Big\}\int_{a_j}^{a_{j+1}} \exp\{-\lambda_j(t-a_j)\}dt \\ &= \exp\Big\{-\sum_{g=1}^{j-1}\lambda_g(a_g - a_{g-1})\Big\}\Big[\lambda_j^{-1}\exp\{-\lambda_j(t-a_j)\}\Big]_{a_j}^{a_{j+1}} \\ &= \lambda_j^{-1}\Big[\exp\Big\{-\sum_{g=1}^{j-1}\lambda_g(a_g - a_{g-1})\Big\} - \exp\Big\{-\sum_{g=1}^{j}\lambda_g(a_g - a_{g-1})\Big\}\Big]\end{aligned}$$

Defining $E_j = \exp\Big\{-\sum_{g=1}^{j}\lambda_g(a_g - a_{g-1})\Big\}$ and completing the integral by summing over all j intervals, define $\int_{t_\alpha}^{t_{\alpha+f}} S(t)dt = \sum_{j=1}^{J}\lambda_j^{-1}(E_{j-1} - E_j)$ where E_j are defined in terms of the time grid a^*. Noting that $E_j = E_{j-1}\exp\{-\lambda_j(a_j - a_{j-1})\}$, a set of recursive formula are produced, meaning that given the intervals and the estimated hazard rates are fixed, estimation of the full integral can be obtained relative to the integral of the first interval in the time-grid.

An estimation of the number of events based on a constant accrual rate is then given by

$$N\gamma\alpha^{-1}\Big[t_\alpha - \lambda_j^{-1}(E_{j-1} - E_j))\Big]$$

In practice, the rate of recruitment will rarely be considered constant from the start of a trial, neither will it be simple to apply some function of time to estimate recruitment. In practice therefore, it may be more straight forward to define recruitment in terms of the number of patients per month (say). Here, define the total number of months of recruitment as $(m = 1, ..., M)$ and n_m as the number of patients recruited in month m then the number of events is expressed as

$$\sum_{m=1}^{M} n_m\gamma(1 - S(t_{\alpha+m}))$$

which is straight forward to calculate for any form of survival function.

8.4 Optimal allocation ratio for the ViP trial

The design of the ViP trial is considered allowing for the allocation ratio to differ from 1:1. As with the design of ViP in Chapter 7, two approaches are taken to evaluate trial design; a fully Bayesian design using the Average Length Criterion (ALC) and Bayesian equivalents of Type I and Type II error rates.

As is shown from (8.7) the optimal allocation ratio for a trial depends both on the total sample size and the amount of information included through the prior distributions, here denoted by the effective number of prior events. Table 8.3 shows the estimated optimal allocation ratios using equation (8.4) for varying total sample sizes and effective number of events in the prior distributions for the control arm.

	Effective Prior Events			
Sample Size	10	20	30	50
60	0.34	0.26	0.18	0.01
70	0.36	0.28	0.21	0.07
80	0.36	0.30	0.24	0.11
90	0.37	0.32	0.26	0.15
100	0.38	0.33	0.28	0.18
110	0.38	0.34	0.29	0.20
120	0.39	0.34	0.30	0.22
130	0.39	0.35	0.31	0.23
140	0.39	0.36	0.32	0.25
150	0.39	0.36	0.33	0.26

Table 8.3: Table to show estimated allocation ratios under differing total sample sizes and effective prior events.

Bayesian sample size calculation

The total sample size for the study is determined using the simulation approach detailed in Chapters 6 and 7. Data are simulated from the design priors for both $\boldsymbol{\lambda}$ and β as detailed in Chapter 7. Only the number of patients allocated to each treatment arm differs based on table 8.3.

For each simulation, the length of the 90% credibility interval is estimated and the mean over all simulations is calculated to give the average length criterion (ALC). Prior distributions for data analysis the baseline hazard function are defined based on Normal, Step and Trapezium priors as detailed in Chapter 7. Uninformative vague prior for the baseline hazard function are included as a reference. All priors on the log hazard ratio are vague and uninformative.

The results are displayed graphically in Figure 8.9. Here four separate figures show the ALC for each of the four different prior distributions that are used. In each figure, the estimated ALC across varying sample sizes are displayed for four different effective event scenarios. The intersection of the vertical and horizontal lines shows the design conditions obtained under equal allocation in Chapter 7. Here 98 events are required to obtain an ALC of 0.6.

Considering initially the non-informative reference priors, it is shown that there is an increase in the ALC for each sample size for the larger 'effective prior events' used to determine the trial allocation ratio. This is misleading however as effective prior

events are included to calculate the optimal allocation ratio but are not included in the analysis of the trial data. The deterioration of the design parameters here is the cost of an unequal allocation when all information from a trial is taken from the data.

Considering the informative priors, all reduce the ALC in comparison to the results obtained under equal allocation in Chapter 7. The Normal distribution has the most accentuated effect on the posterior ALC. The Trapezium priors have the least effect which is consistent with previous findings.

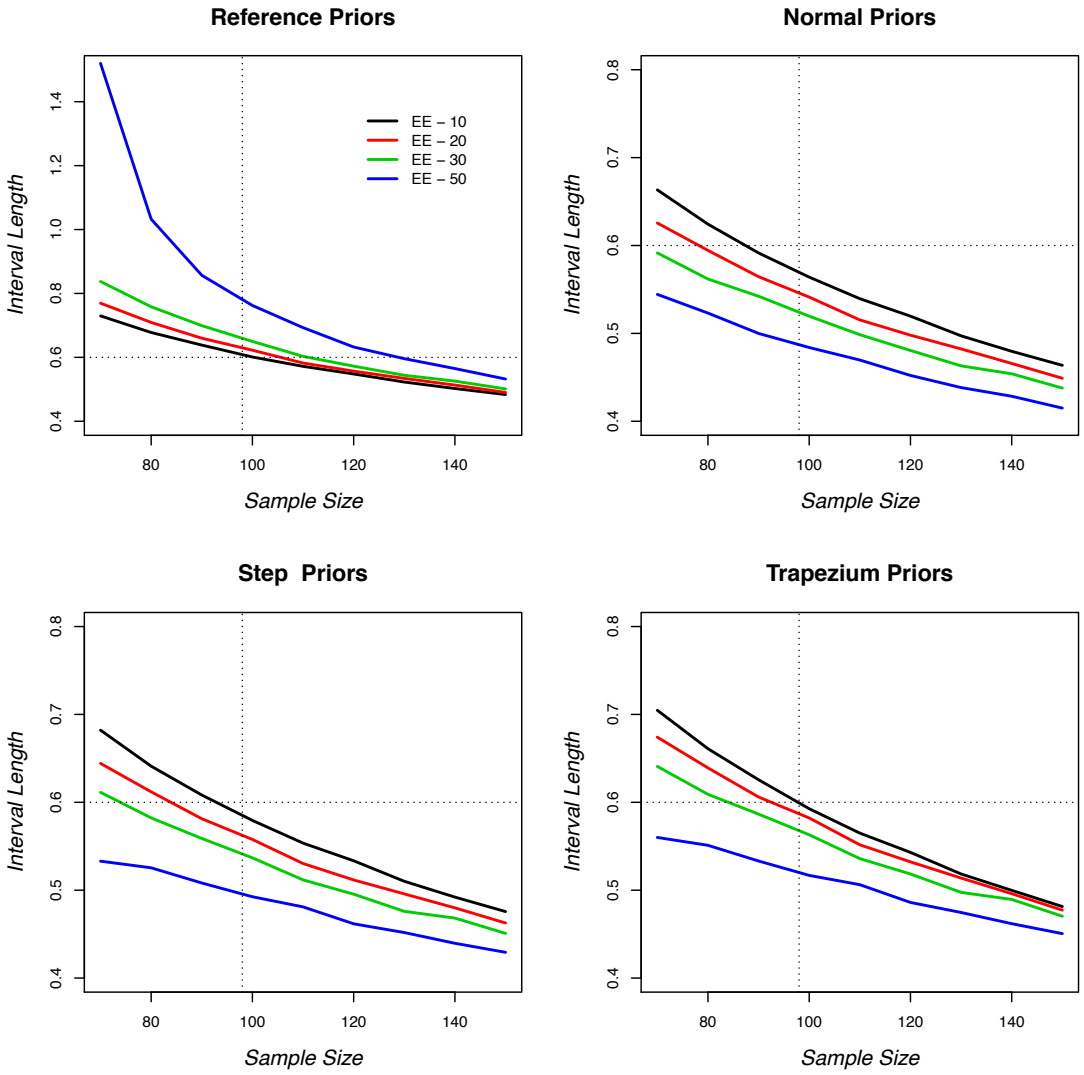

Figure 8.9: Figure to show the ALC for different types of priors and different number of effective events.

The sample sizes required to obtain, on average, an ACL of 0.6 for each of set of priors and set of effective prior event are given in Table 8.4. These can be compared directly to the sample sizes given for equal allocation ratios given in Table 7.3. Comparisons here show the normal priors have a greater effect over the locally flat priors

and also how the sample sizes with amended allocation ratio are consistently smaller than equivalent sample size calculations with equal allocation ratios.

	Effective Sample Size			
	$\zeta = 10$	$\zeta = 20$	$\zeta = 30$	$\zeta = 50$
Normal	89	78	74	67
Step	96	82	77	70
Trapezium	97	96	83	72

Table 8.4: Sample size estimates for the ViP trial under various differing priors on the baseline hazard function under the ALC.

Bayesian Type I and Type II error rates

Here design criteria are presented in terms of estimated Type I and Type II error rates. As in Chapter 7, sample sizes are fixed at 120 to replicate the initial design of the ViP trial. Table 8.5 gives the Bayesian Type I and Type II error rates that are obtained.

The results obtained here can be directly compared to those obtained in Figure 7.4 for equal allocation ratios. There are two points of note to be made, firstly there is a reduction in the Type I error rate as the effective prior events increase. This can be attributed to the unequal allocation ratios that are obtained. As is consistent throughout, the optimum design parameters are obtained when normal prior distributions are used. Differences in the Type I error rates between distributions are small however.

The second point to note is that the effect of the Type II error rate. Here for the reference models, where trials are designed based on informative priors but vague uninformative prior distributions are included in the analysis, there is a increase in the Type II error rate for larger effective prior events which is a consequence of the unequal allocation ratio. This shows that if informative prior distributions are used to inform a trial design, not including the same prior distributions in the analysis of a trial can have adverse effects.

For the normal, step and trapezium priors there is a notable decrease in the Type II rate, to a greater extent than what is observed in Table 7.4 for equal allocation.

8.5 Discussion

In this chapter, an investigation was carried out into the optimal design for clinical trials based on amendments to the proportion of patients allocated to each treatment arm in a two-arm trial. Clinical trials designs are proposed based on the assumption of attempting to maximise the precision about the main efficacy parameter of interest for normal, binomial and time-to-event endpoints in frequentist and Bayesian frameworks.

For a normal outcome in a frequentist framework, it is simple to show that the most information is obtained for a 1:1 allocation ratio. For both binary and time-to-event

Priors	Error	Effective Prior Events 10	20	30	50
Reference	Type I	0.11	0.09	0.07	0.08
	Type II	0.13	0.12	0.15	0.25
Normal	Type I	0.09	0.08	0.07	0.09
	Type II	0.10	0.06	0.07	0.06
Step	Type I	0.10	0.08	0.08	0.10
	Type II	0.10	0.07	0.09	0.06
Trapezium	Type I	0.11	0.09	0.07	0.08
	Type II	0.13	0.10	0.09	0.09

Table 8.5: The effect of different priors and effective prior events on Bayesian Type I and Type II error rates

endpoints however, an equal allocate ratio is only optimal when there is no difference between the control treatment and the experimental treatment, a scenario under which the precision of the efficacy parameter is of the least interest.

Optimal allocation ratios are initially obtained for binary and time-to-event outcomes in a frequentist framework based on providing an optimal precision for the efficacy parameter under the alternative hypothesis. Whilst there is some discussion on whether the optimal allocation should be based on some value between the null and alternative hypotheses, with the aim of providing the most precision under borderline results, this was not explored any further in this chapter.

As an extension, analytical expressions for the optimal allocation ratios are provided for each outcome in the presence of informative priors in a Bayesian framework. Expressions are based on being able to define prior distributions for the response in the control and experimental arms separately. No prior distributions are defined directly for the efficacy parameter. Examples are given whereby prior distributions are derived for the control arm of a trial with the aim of obtaining as much information as possible on the precision of an efficacy parameter. Examples are typically set in the context of a Phase II trial and aim to provide the most efficient and informative approach for determining whether a new therapy is suitable for expansion into a phase III trial.

Whilst trials designs that have deviated from the standard 1:1 have been criticised (see for example Korn et al. [221]), much of this criticism has been based on the arguments that non standard designs, such as adaptive trials, whilst ethically appealing can lead to a loss of statistical power. It has been shown that this may be misleading in both a Bayesian and frequentist framework, especially for binary and time-to-event endpoints as optimal precision, and by extension trial power, can be obtained from an unequal allocation ratio when there is some evidence of efficacy. In some cases ethical issues may be raised as it may be that increased allocation is given to a less effectual treatment as a means of maximising the statistical information. Where an unequal allocation ratio is both ethically and statistically appealing however, it should

be encouraged.

In a Bayesian framework, it is shown that design parameters can be improved by unequal allocation for all types of endpoint when there is reliable prior information available for one or both treatment arms.

Chapter 9

Discussion

9.1 Introduction

This chapter provides an overview of the research that has been carried out within this thesis as well as some suggestions for further work. Further discussions on the uses of methodology included in this thesis are also provided.

9.2 Topics covered

The focus of this thesis has been on the design and analysis of clinical trials with a primary endpoint of overall survival. In particular the aim has been to investigate methods which may influence the way trials are designed or data analysed which allow for greater information to be obtained than is provided under standard methodology.

The first approach taken is to introduce a new parametric form for the linear predictor in the analysis of proportional hazards models with unbounded covariates. This form is motivated by clinical application and whilst it doesn't provide an improvement over more complex methods such as the flexible parametric models approach, it does facilitate clinical interpretation and allows for intuitive extrapolation beyond the scope of the observed data.

Following this, the problem of analysing data with non-proportional hazards is explored. Proportional hazards being by far the most popular method of survival analysis. Non-proportional hazards do occur in practice however, and the appropriate method in this scenario is not always clear. Chapter 4 explores current methods for detecting and modelling non-proportional hazards and introduces the idea of modelling data with the inclusion of an asymmetry parameter. The difference between two levels of a covariate is here expressed by a divergence parameter (as opposed to a hazard ratio) which measurers the absolute difference and an asymmetry parameter which accounts for non proportionality. This approach is applied to a gastric cancer dataset and it is shown how a direct comparison can still be made despite the non-proportional nature of the data.

Chapters 5 - 7 concentrate on the application of Bayesian methodology to the design and analysis of a cancer clinical trial. Chapter 5 concentrates on analysing survival data from the ESPAC-3 trial using a piecewise exponential model (PEM), practical issues are discussed and some advantages of incorporating a Bayesian framework are explored when it comes to evaluating a treatment effect. Chapter 6 investigates the methods involved in the design of clinical trial based on Bayesian criteria and provides a comparison against the frequentist framework.

In Chapter 7, some investigation into the incorporation of historical control information is explored. Specifically, summary information is obtained from previously published clinical trials and this information is used to develop a set of informative prior distributions which describe the behaviour of the baseline hazard/survival function. It is shown that this approach can improve the design parameters of a clinical trial. This methodology is further applied to the ViP trial currently being run at the Liverpool Cancer Trials Unit.

Finally Chapter 8 investigates allocation ratios. Under standard trial clinical trials design, an allocation other than 1:1 is rare in a two arm trial. Optimal allocation ratios are investigated on the basis of minimising the expected size of the standard error for the key efficacy parameter. It is initially shown that in a frequentist framework, for binary and survival outcomes, a 1:1 allocation ratio is only optimal under the null hypothesis. Under successful or borderline successful trials therefore, a sub-optimal standard error is obtained which may inhibit the conclusions that can be drawn. The idea is extended to a Bayesian framework and the inclusion of informative historical controls considered. Depending on the precision of the prior distributions that are included, it is shown here that optimal trials can be designed with a larger proportion of patients being allocated to the experimental arm. In the most extreme cases with informative priors, no patients are recruited to the control arm of a trial. These methods are again incorporated into the design of the ViP trial.

9.3 Further work

There are a number of area on which further work may be carried out following the completion of this thesis. Considering the example of the asymmetry parameter introduced in Chapter 4, some difficulty was met with parameter estimation in large datasets. Specifically, under a counting process notation, large datasets required the inversion of a large matrix in order to obtain a standard error estimate. This step is computationally expensive and slows the procedure down. Efficient algorithms including a legrange estimator are available for proportional hazard models and applying these to the model with an asymmetry parameter would be a worthwhile step.

Considering the Bayesian survival data, there is already a large body of literature on both the Bayesian analysis of survival data and the inclusion of Bayesian methodology

for clinical trial design. The use of Bayesian methodology for trials with time-to-event endpoints is still rare in both practical applications and the literature however. Whilst there is scope for work to investigate further methods of defining prior distributions on the baseline hazard, the biggest impact would be through practical application of the design approaches.

Considering prior derivation, Chapter 7 introduces a method based on the effective number of events which is intended to be intuitive to clinicians. The number of events are then distributed proportional according to the assumed behaviour of the survival curve and a set of prior distributions formed from the results.

Lastly, considering allocation ratios, there has been a large portion of literature already attributed to adaptive allocation, these are typically based on assessing the treatment effect and allocating a greater proportion of patients to the therapy that is working. Should ethical issues allow, there may be some scope therefore in adapting the allocation ratio during the course of a trial to minimise the standard error associated with the efficacy parameter with the aim of maximising the information that can be obtained from a clinical trial.

9.4 Summary

Clinical trials never occur in a vacuum. They are a form of research which on their completion will contribute towards the general pool of clinical information available. Whilst Phase III trials may be designed with the intention of changing medical practice therefore, on their completion, they will not be evaluated in isolation. More that the information taken from any trial will be compared against other trials and other anecdotal evidence before practice is changed.

Furthermore, clinical trials are expensive in both time and money. It is an ethical imperative therefore that the statistician endeavours to make the most efficient use of any data that are obtained. Whilst efficiencies can be obtained in the accurate analysis of trial data, this process of efficient analysis begins at the design stage of any trial.

The inclusion of prior information to reflect this wider knowledge does not only hold advantages in that it better reflects the clinical environment, it allows for potentially large saving in both time and money and can provide answers to a greater level of precision than may have been previously feasible. As an example in the most extreme case, the ViP trial with a standard frequentist design required 120 patients which were recruited over a period of approximately 18 months. With a fully informative design only 70 patients would have been required which could potentially prevent 50 patients being given a potentially harmful treatment. Perhaps the greater impact however would be that the trial would have completed recruitment in a total of 13 months, saving 5 months. The trial would have completed recruitment and reported sooner and through predictive posterior distributions would in a better position to inform future design.

Finally the point is made that whilst this thesis concentrates on the design of phase II clinical trials the methodology proposed is not confined to this scenario. For example, the inclusion of priors on the baseline hazard may be useful in analyses of data where data are difficult or costly to obtain such as the analysis of rare disease data.

Appendices

Appendix A

Code

A.1 Piecewise Exponential Model

Included is a function written in R to fit a piecewise exponential model based on a generalised linear model assuming a poisson distribution, log link function and $\log(t)$ as a model offset. Required are a vector of survival times, 'T', a vector of censoring indicators, 'CEN', a set of partitions, 'part. and a formula for the explanatory covariates, 'form'.

```
PEM <- function(T,CEN, part , form){

        ### Setting structures
        n<-length(T)
        n.part<-length(part)-1

        t.mat<-matrix(0,n,n.part)
        w.mat<-t.mat

        ### Defining time, tt, and observation
        ### indicator, w.mat, in vector form
        for(i in 1:n){
                tt<-T[i]-part;tt
                tt[which(tt<0)]<-0
                t.mat[i,]<-tt[1:n.part]
                w.mat[i,max(which(tt>0))]<-1
        }

        for(i in 1:n.part){
        t.mat[which(t.mat[,i]>diff(part)[i]),i]<-diff(part)[i]}
        w.mat<-w.mat*CEN

        ### defining structures for log linear model
        resp<-c(w.mat)
        time<-log(c(t.mat)+0.000001)
        int<-gl(n.part,n)
```

```
        ### Setting formula
        if(form=="1") modmat<-matrix(1,n,1) else
        modmat<-model.matrix(form)

        n.co<-ncol(modmat)
        coef.mat<-matrix(NA,n*n.part,n.co)

        for(i in 1:n.co){
                coef.mat[,i]<-rep(modmat[,i],n.part)
        }

        colnames(coef.mat)<-colnames(modmat)

        ### Fitting models
        if(n.part==1) loglin<-glm(resp~coef.mat,
        family="poisson"(link=log),offset=time) else

        loglin<-glm(resp~-1+int+coef.mat,
        family="poisson"(link=log),offset=time)

        return(loglin)
}
```

A.2 PP plot

Included is a function for producing a PP plot based on the survival estimates obtained via the method of Kaplan and Meier. Required are a survival object (obtained using the 'Surv' function in the package 'survival') and an associated vector of covariates, z. Results can be presented using 'res = TRUE'. The output can be included on an already exciting plot using 'add=TRUE'.

```
PPplot <- function(s.ob,z,res=FALSE,add=FALSE){

        ### Organising Factor and getting survival estimates
        lev<-levels(factor(z))
        KM<-survfit(s.ob~z)

        ### Setting up required structures
        l1<-length(which(z==lev[1]))
        l2<-length(which(z==lev[2]))

        t0<-KM$time[1:l1]
        t1<-KM$time[(l1+1):(l1+l2)]
        s0<-KM$surv[1:l1]
        s1<-KM$surv[(l1+1):(l1+l2)]
```

```
    ### Sorting data structures
    T<-sort(c(t0,t1))
    surv0<-1
    surv1<-1

    for(i in 1:length(T)){
            surv0<-c(surv0,max(0,s0[which(t0==T[i])]))
            surv1<-c(surv1,max(0,s1[which(t1==T[i])]))

            if(surv0[i+1]==0) surv0[i+1]<-surv0[i]
            if(surv1[i+1]==0) surv1[i+1]<-surv1[i]
    }

    ### Plotting
    if(add==FALSE){
    plot(surv0,surv1,typ="s",xlim=c(0,1),ylim=c(0,1),main="PP plot",
    xlab="Treatment 0",ylab="Treatment 1")
    points(surv0,surv1,pch="+")
    abline(a=0,b=1,lty=2)
    }

    if(add==TRUE){
    lines(surv0,surv1,typ="s",xlim=c(0,1),ylim=c(0,1),main="PP plot",
    xlab="Treatment 0",ylab="Treatment 1")
    points(surv0,surv1,pch="+")
    }

    ### Returning results
    ret<-list(surv0,surv1)
    if(res==TRUE) return(ret)
}
```

A.3 Modelling non-proportional hazards using a non-parametric maximum likelihood estimation

Included is a function for fitting generalised survival model based on a non-parametric maximum likelihood estimation NPML of a counting process. Models can be fit on both a hazards and an odds scale with and without a asymmetry parameters. Required are a vector of survival times, 'time' and a vector of censoring indicators 'cen'. Both formulas for the inclusion of covariates as dispersion parameters 'z_p' and asymmetry parameters 'z_a' can be provided. Setting 'cf = TRUE' facilitates the estimation of a cure fraction. The model option may be set to 'hazard' or 'odds' (defaults to 'hazards') and dictates the scale on which models are fitted. The hessian option details whether the model hessian should be extracted and by extension, parameter standard errors are obtained. Large datasets with the hessian set to TRUE may take a while to fit.

```
CPgen <- function(time,cen,z_p=0,z_a=0,
cf=FALSE,model="hazard",hessian=FALSE){

### Setting necessary structures
n.ob<-length(time)
n.event<-sum(cen)

### Sorting covariates into matrix format
if(z_p!=0) Z_p<-as.matrix(model.matrix(formula(z_p))[,-1])
else Z_p<-matrix(0,n.ob,1)
if(z_a!=0) Z_a<-as.matrix(model.matrix(formula(z_a))[,-1])
else Z_a<-matrix(0,n.ob,1)

### Ordering data
ord.id<-order(time)
time<-time[ord.id]
cen<-cen[ord.id]
Z_p<-as.matrix(Z_p[ord.id,])
Z_a<-as.matrix(Z_a[ord.id,])

### Number of parameters
n.beta<-ncol(Z_p)
n.alpha<-ncol(Z_a)

### Fitting a cox model to get starting values
s.ob<-Surv(time,cen)
cox.mod<-coxph(s.ob~Z_p)
start<-log(coxph.detail(cox.mod)$hazard)
start<-c(start,log(2),rep(-0.4,n.alpha))

if(cf==TRUE){
        start[1:n.event]<-start[1:n.event]+log(0.9)
        start<-c(start,inv.logit(0.1))
}

### variable ID variables
beta.id<-(n.event+1):(n.event+n.beta)
alpha.id<-(n.event+n.beta+1):(n.event+n.beta+n.alpha)
if(cf==TRUE){gamma.id<-length(start)}


####### Likelihood function
Lik<-function(start,model=model,cf=cf){

        #### Defining parameter values
        b.line<-exp(start[1:n.event])
        beta<-start[(n.event+1):(n.event+n.beta)]
        alpha<-start[(n.event+n.beta+1):(n.event+n.beta+n.alpha)]
```

```
    if(cf==TRUE) gam<-logit(start[length(start)])

    ### Setting model structures
    phi<-exp(beta%*%t(Z_p))
    alp<-exp(alpha%*%t(Z_a))
    LAM<-cumsum(b.line)
    jump.id<-cumsum(cen)

    P<-t(matrix(t(phi),n.ob,n.ob))
    A<-t(matrix(t(alp),n.ob,n.ob))
    #pat.id<-cbind(jump.id,c(1:n.ob))

    G<-(P*(LAM[jump.id]^A))
    g<-diag(G)

    ### Defining Likelihood

    h1<-rbind(0,G)
    h1<-h1[-nrow(h1),]
    h1<-G-h1
    h<-diag(h1)


    if(toupper(model)=="HAZ"|toupper(model)=="HAZARD"
    |toupper(model)=="HAZARDS"){
            Su<-exp(-g)
            fu<-h*Su
    }

    if(toupper(model)=="ODD"|toupper(model)=="ODDS"){
            Su<-1/(1+g)
            fu<-h*(Su^2)
    }

    if(cf==TRUE){
            S<-gam+(1-gam)*Su
            f<-(1-gam)*fu
    }

    if(cf==FALSE){
            S<-Su
            f<-fu
    }
    f[which(f==0)]<-0.999

    -sum(log(f)*cen+(1-cen)*log(S))

}
```

```
### Maximising likelihood
op<-optim(start<-start, fn=Lik, cf=cf, model=model, method="BFGS",
control=list(trace=TRUE, maxit=100), hessian=hessian)

#### Presenting results
par<-op$par
sd<-NA
if(hessian==TRUE){
        hess<-op$hessian
        if(z_a==0) hess<-hess[-alpha.id,-alpha.id]
        sd<-sqrt(diag(solve(hess)))
}

BETA<-matrix(cbind(par[beta.id],exp(par[beta.id]),sd[beta.id]),n.beta,3)

BETA<-cbind(BETA,BETA[,1]/BETA[,3])
BETA<-cbind(BETA,2*(1-pnorm(abs(BETA[,4]))))
BETA<-as.data.frame(BETA)
names(BETA)<-c("coef","exp(coef)","se(coef)","Z","PR(>|Z|)")

ALPHA<-matrix(cbind(par[alpha.id],exp(par[alpha.id]),
sd[alpha.id]),n.alpha,3)
ALPHA<-cbind(ALPHA,ALPHA[,1]/ALPHA[,3])
ALPHA<-cbind(ALPHA,2*(1-pnorm(abs(ALPHA[,4]))))
ALPHA<-as.data.frame(ALPHA)
names(ALPHA)<-c("coef","exp(coef)","se(coef)","Z","PR(>|Z|)")

if(cf==TRUE) {
        GAMMA<-matrix(cbind(par[gamma.id],logit(par[gamma.id]),
        sd[gamma.id]),n.beta,2)
        GAMMA<-cbind(GAMMA,GAMMA[,1]/GAMMA[,2])
        GAMMA<-cbind(GAMMA,2*(1-pnorm(abs(GAMMA[,3]))))
        GAMMA<-as.data.frame(GAMMA)
        names(ALPHA)<-c("coef","logit(coef)","se(coef)","Z","PR(>|Z|)")
        }


if(z_a!=0){
        if(cf==TRUE) ret<-list("Beta"=BETA,"Alpha"=ALPHA,
        "Gamma"=GAMMA)
        if(cf==FALSE) ret<-list("Beta"=BETA,"Alpha"=ALPHA)
}

if(z_a==0){
        if(cf==TRUE) ret<-list("Beta"=BETA,"Gamma"=GAMMA)
        if(cf==FALSE) ret<-list("Beta"=BETA)
}
```

```
ret

}
```

A.4 Markov Chain Monte Carlo routine for fitting Bayesian piecewise exponential models

Included is a function for fitting Bayesian piecewise exponential model with a single covariate using batch sampling for the baseline hazard function. Code below includes vague normal priors for the baseline hazard parameters and the log hazard ratio. Required are the number of simulations desired 'SIMS', the number of models (chains) to be set 'n.mod', a list of starting parameters 'THETA'. The burn in 'burn' and thin, 'thin' are both required. The parameter 'CC' sets the variance parameter for the baseline hazard function. In this form, the variance parameters are inflated due to the number of partitions.

```
MCMCnorm <- function(SIMS,n.mod,THETA,burn,thin,CC){

### Setting structures and initiating procedure
n.parm<-length(THETA)
C<-CC*seq(1,2,length=(n.parm-1))
results<-array(NA,dim=c(SIMS,n.parm,n.mod))
results[1,,]<-THETA

        ### MCMC routine
        for(i in 2:SIMS){

        ###  Collecting results from previous iteration
        mcmc.res<-results[(i-1),,]
        prev.parm<-mcmc.res

        ######
        ### Baseline hazard parameters
        ######

        ### simulating proposal values
        prop.parm<-prev.parm
        prop.parm[1:(length(part)-1),] <-
        rnorm(n.mod*(length(part)-1),prev.parm[1:(length(part)-1),],0.1)

        ### Evaluating previous and proposal values
        prop.lik<-PEMlik(prop.parm) +
        colSums(log(dnorm(prop.parm[1:(n.parm-1),],pr.est,C)))
        prev.lik<-PEMlik(prev.parm) +
        colSums(log(dnorm(prev.parm[1:(n.parm-1),],pr.est,C)))
```

```
        ### Simulating acceptance probability and evaluating
        acc.prob<-exp(prop.lik-prev.lik)
        acc.check<-runif(n.mod)
        acc.id<-which(acc.prob>acc.check)

        ### Saving accepted results
        mcmc.res[1:(length(part)-1),acc.id] <- prop.parm[1:(length(part)-1),

        ######
        ### log hazard ratio parameters
        ######

        ### simulating proposal values
        prev.parm<-mcmc.res
        prop.parm<-prev.parm
        prop.parm[n.parm,]<-rnorm(n.mod,prev.parm[n.parm,],0.1)

        ### Evaluating previous and proposal values
        prop.lik<-PEMlik(prop.parm) +
        log(dnorm(prop.parm[n.parm,],0,1000))
        prev.lik<-PEMlik(prev.parm) +
        log(dnorm(prev.parm[n.parm,],0,1000))

        ### Simulating acceptance probability and evaluating
        acc.prob<-exp(prop.lik-prev.lik)
        acc.check<-runif(n.mod)
        acc.id<-which(acc.prob>acc.check)

        ### Saving accepted results
        mcmc.res[length(part),acc.id]<-prop.parm[length(part),acc.id]
        results[i,,]<-mcmc.res
        }

        ### Applying burn in and thin and retiring results
        id<-seq(burn,SIMS,by=thin)
        results[id,,]

}
```

Appendix B

Publications

Attached is the publication for "A robust parameterisation for the analysis of survival data in the presence of covariates with extreme value observations"

A Robust Parameterization for Unbounded Covariates Within the Cox Proportional Hazards Model

Richard J. Jackson* and Trevor F. Cox

Cancer Research UK Liverpool Cancer Trials Unit, University of Liverpool, Liverpool, UK

Abstract: The Cox proportional hazards model is widely used in the analysis of medical data either for survival or time to a particular event. Factors and continuous covariates can be easily incorporated into the model and hazard ratios calculated. The model can however be distorted when extreme value observations occur within a continuous covariate and the hazard ratio can become extremely large. To overcome this, transformations of the covariate are often made, which can be simple, e.g. log, or more sophisticated such as the fitting of a fractional polynomial. This paper takes a different approach and makes a transformation based on the logistic function that has the property that the hazard ratio is bounded. The models are introduced and discussed. Model diagnostics based on Schoenfeld residuals and the influence function are established and then data from a pancreatic cancer trial are used to illustrate the model.

Keywords: ESPAC 3 trial, hazard ratio, influence function, logistic function, Schoenfeld residuals.

INTRODUCTION

Cox proportional hazards modeling is in widespread use in medical and other contexts [1]. Here robust hazard models are proposed that extend the Cox model. Survival data will be the main focus but the robust models proposed can be used in any time-to-event situation. Survival data from a pancreatic cancer randomized controlled trial will be used for illustration.

For the Cox model, the hazard function, λ, is modeled as

$$\lambda = \lambda_0 \exp\left(\beta^T \mathrm{x}\right)$$

where *x* is a vector of covariates and β a vector of coefficients. For a factor, x_i, the associated coefficient, β_i, leads to the hazard ratio, $\gamma_i = \exp(\beta_i)$, whilst for a continuous variable β_i gives the increase in hazard ratio per unit increase in the value of the continuous variable. One problem with continuous variables is that modeling the hazard in such a linear way can mean large or extreme values of x are associated with very high hazards when this may be unrealistic in practice. The presence of only a single extreme value observation can be enough to violate any model assumptions of proportionality [2] and biased estimates produced. Also, the interpretation of the hazard ratio is awkward or inappropriate. To overcome these problems, sometimes a simple transformation can be made, for instance, log(x), or x^{-1} and this can be sufficient for obtaining a good model fit. A more sophisticated model is achieved by using a fractional polynomial approach [3], which uses a mixture of transformations, but both simple transformations and fractional polynomials can still be influenced by extreme observations. In this paper a transformation based on the logistic function is proposed that can improve the model fit and guard against extreme observations having undue influence on the overall model.

Some previous methods to account for extreme value observations have concentrated on amendments to the likelihood formulation. A good overview is given by Farcomeni and Ventura [4] with two approaches in particular given specific attention: an approach based on a weighted likelihood formulation for the Cox model, notably proposed by Bednarski [5] and Minder and Bednarski [6] and secondly, an approach using `trimmed' likelihoods given by Viviani and Farcomeni [2]. For weighted Cox regression, a likelihood is proposed in the form:

$$1(\beta) = \log\left(\mathrm{L}(\beta)\right) = \sum_{i=1}^{N} A\left(t_i, z_i\right)\left[z_i - \frac{\sum_{R(j)} A\left(t_i, z_i\right) z_j \exp\left(\beta^T z_j\right)}{\sum_{R(j)} A\left(t_i, z_i\right) \exp\left(\beta^T z_j\right)}\right]$$

Here $A\left(t_i, z_i\right)$ is a smooth non-negative function which takes a value zero for either large values of t or $\beta^T z$ and R (j) the usual risk set at time t_i. This method down-weights or completely ignores patients who either have large covariate values or who live longer than may be expected. The second approach uses a trimmed likelihood by excluding observations that give the smallest contribution to the likelihood. Whilst either procedure may produce more robust hazard ratios they do not cure the problem of non-proportionality. More troublesome may be that the model is explicitly treating some data as less valuable than others and a possible

*Address correspondence to this author at the Cancer Research UK Liverpool Cancer Trials Unit, University of Liverpool, Block C Waterhouse Building, 1-3 Brownlow Street, Liverpool, L69 3GL, UK; Tel: +44 (0)151 7948834;
E-mail: richj23@liv.ac.uk

E-ISSN: 1929-6029/14

criticism is that the methods can be seen as trying to amend the data to fit a model as opposed to producing a model to fit the data.

A ROBUST PARAMERIZATION

Let the hazard function be modelled as

$$\lambda = \lambda_0 f(\theta, x)$$

where $\theta=(\alpha, \delta, \omega, \beta)$ are parameters to be estimated. The family of transformations proposed here has the form

$$f(\theta, x) = \frac{\delta + \alpha \exp(\beta x)}{\omega + \exp(\beta x)} \quad \text{(model 1)}$$

This is an adaption of the logistic function and has asymptotes α and δ/ω. Restrictions are needed on the parameters: $\alpha \geq 0$, $\delta \geq 0$, $\omega \geq 0$, in order for $f(\theta,x)$ to be non-negative. The first derivative of $f(\theta,x)$ is $(\omega\alpha-\delta)\beta\exp(\beta x)/\{\omega+\exp(\beta x)\}^2$ and in order for a positive β to have positive slope for f, and correspondingly, a negative β to have a negative slope for f, then $\delta<\omega\alpha$. Also, $f(\theta,x)$ is monotonically increasing in x which is usually a useful property in practice. A particular fractional polynomial might not possess this property. Model 1 has the property that the hazard function is symmetric regarding the baseline hazard, i.e. $f(\theta,x)$ and $1/f(\theta,x)$ have the same functional form for the two reciprocal models $\lambda=\lambda_0 f(\theta,x)$, and $\lambda_0=\lambda/f(\theta,x)$.

A desirable property for $f(\theta,x)$ is that when $\beta = 0$, implying that the covariate has no effect on survival, then $f(\theta,x)$ should have the value unity. This implies $\omega=\delta+\alpha-1$ and leads to model 2.

$$f(\theta, x) = \frac{\delta + \alpha \exp(\beta x)}{\delta + \alpha - 1 + \exp(\beta x)} \quad \text{(model 2)}$$

The asymptotes for model 2 are α and $\delta/(\delta+\alpha-1)$ and it still retains baseline hazard symmetry. For $\alpha > 1$, positive β will give a positive slope for f and negative β a negative slope. The value of x which has no effect on the baseline hazard is x=0. If this should be a different value then the variable x should be adjusted accordingly with a linear transformation. Note if a parameter is entered into the model, replacing x by $x-\zeta$ with estimation of ζ, then it can be shown that, by rearranging parameters, model 2 reverts back to model 1.

The slopes of the logistic function at +x and –x are identical. For model 2 to have this property, then $\delta=2-\alpha$ and hence

$$f(\theta, x) = \frac{2 - \alpha + \alpha \exp(\beta x)}{1 + \exp(\beta x)} \quad \text{(model 3)}$$

where the asymptotes are α and $2-\alpha$. This model loses its baseline hazard symmetry.

For model 2 to have reciprocal asymptotes, α and $1/\alpha$, then $\delta=1$ giving

$$f(\theta, x) = \frac{1 + \alpha \exp(\beta x)}{\alpha + \exp(\beta x)} \quad \text{(model 4)}$$

This model retains baseline hazard symmetry.

Lastly the standard Cox model is obtained by letting α become infinite in model 4, or by letting $\alpha = 0$ which will negate the β coefficient. If $\alpha = 1$ then $f(\theta, x) = 1$ with x having no effect on the hazard function.

In this paper, concentration is on model 4 although the other models could be used and fitted to data in a similar manner to that for model 4.

Fitting The Model

Suppose the explanatory variables consist of p-1 binary variables, $x_1,\ldots,x_{p-1}$, representing various factors and one continuous covariate, x_p, to be fitted in the proportional hazards models. The hazard for the i*th* observation is

$$\lambda_i = \lambda_0 \exp\left(\beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}\right)\left\{1 + \alpha \exp\left(\beta_p x_p\right)\right\} / \left\{\alpha + \exp\left(\beta_p x_p\right)\right\}$$

and then the partial likelihood is given by

$$\prod_{i=1}^{N}\left\{\frac{\lambda_i}{\sum_{j \in R(i)} \lambda_i}\right\}^{\delta i}$$

where *R(i)* is the risk set at the*i*th survival time and δ_i is the censoring value (1 – the event occurred, 0 – the time is censored). Here, the partial likelihood was maximized using the "optim" package within the statistical package R.

A Simulation Study

A small simulation study was carried out to investigate model fitting and accuracy. Survival times were simulated from an exponential distribution with baseline hazard set at 0.5 and with 5% of observations

Table 1: Results of Fitting the Standard and the Robust Model to Data Simulated from the Standard Model (α=0)

N	Param.	Standard model					New model				
		Est. (s.e.)	Bias	Acc.	Cov.	ACIL	Est. (s.e.)	Bias	Acc.	Cov.	ACIL
100	β_{trt}	0.128 (0.231)	0.022	0.054	0.94	0.848	0.129 (0.227)	0.021	0.052	0.93	0.844
	β_{cov}	0.05 (0.005)	0	0	0.99	0.022	0.05 (0.005)	0	0	0.92	0.02
250	β_{trt}	0.127 (0.134)	0.023	0.018	0.69	0.518	0.126 (0.133)	0.024	0.018	0.68	0.516
	β_{cov}	0.05 (0.003)	0	0	0.68	0.013	0.049 (0.004)	0.001	0	0.63	0.012
500	β_{trt}	0.151 (0.093)	-0.001	0.009	0.97	0.361	0.144 (0.097)	0.006	0.009	0.94	0.36
	β_{cov}	0.05 (0.002)	0	0	0.96	0.009	0.049 (0.002)	0.001	0	0.92	0.009
1000	β_{trt}	0.152 (0.063)	-0.002	0.004	0.95	0.253	0.148 (0.067)	0.002	0.005	0.93	0.252
	β_{cov}	0.05 (0.002)	0	0	0.93	0.006	0.049 (0.002)	0.001	0	0.78	0.006

randomly censored. Explanatory variables were a two-level factor representing treatment within a two-arm clinical trial alongside a continuous covariate. Parameters for the robust model are denoted by (β_{trt}, β_{cov}, α_{cov}). Firstly, a check was made on whether the new model could be fitted adequately to data arising from the standard Cox model by fixing the value of α_{cov} to be zero when simulating the data. The parameters (β_{trt}, β_{cov}) were given the values (0.15, 0.05). Patients were split equally between the two arms. The covariate was simulated as log (x) ~ N (3.5, 1.5). Sample sizes of 100, 250, 500 and 1000 were used, each time simulating 1000 datasets. Each fitted model was assessed in terms of bias, accuracy, coverage and average confidence interval length (ACIL) [7]. Table **1** shows the results of fitting the standard Cox model and the new model. It can be seen that there is very good agreement between the estimates for two models. The estimated values of α_{cov} (not shown) for the robust model were large enough to essentially make the model equivalent to the standard Cox model.

Next, data were simulated from the robust model formulation with the parameters (β_{trt}, β_{cov}, α_{cov}) and given the values (0.15, 0.05, 5). Table **2** shows the results of the simulations where the new model, the standard Cox model, the standard Cox model with log-transformed covariate values and a fractional polynomial model are fitted. The new model fits the data well and recovers the true parameter values accurately. The bias in the treatment coefficient is very small. There is some small reduction in the coverage for β_{cov} in the robust models and upon further inspection, this can be attributed to some skewness in the distribution of β_{cov}. The standard model underestimates the treatment coefficient even for a sample size of 1000. Note, β_{cov} cannot be compared across the two models. The log-transformed model achieves similar bias to the robust model but the fractional polynomial model offers little improvement over the standard model.

MODEL DIAGNOSTICS

Two model diagnostics are explored for the new model, (i) residuals based on standard Schoenfeld residuals [8] and (ii) an analytical form of an influence function following the method of Reid and Crapeau [9].

Residuals

Schoenfeld residuals for a particular covariate are calculated using the partial derivative of the partial log-likelihood function with respect to the covariate's associated parameter and evaluating this at the maximum likelihood estimate. For robust model 4, there are two parameters α and β and so two Shoenfeld type residuals will be calculated. Differentiating the log-likelihood with respect to α gives

$$\sum_{i=1}^{N}\delta_i\left[\frac{\exp\left(\beta x_i\right)-1}{\alpha\left\{\alpha-1+\exp\left(\beta x_i\right)\right\}}-\frac{\sum_{j\in R}\frac{\left(\exp\left(\beta x_j\right)-1\right)\left(\exp\left(\beta x_j\right)\right)}{\left\{\alpha-1+\exp\left(\beta x_j\right)\right\}^2}}{\sum_{j\in R}\frac{\alpha\exp\left(\beta x_j\right)}{\alpha-1+\exp\left(\beta x_j\right)}}\right]$$

and with respect to β,

$$\sum_{i=1}^{N}\delta_i\left[\frac{\left(\alpha-1\right)x_i}{\alpha-1+\exp\left(\beta x_i\right)}-\frac{\sum_{j\in R}\frac{\alpha\left(\alpha-1\right)x\exp\left(\beta x_j\right)}{\left\{\alpha-1+\exp\left(\beta x_j\right)\right\}^2}}{\sum_{j\in R}\frac{\alpha\exp\left(\beta x_j\right)}{\alpha-1+\exp\left(\beta x_j\right)}}\right].$$

Table 2: Simulation Results to Assess Fitting of the Robust Model

N	Param.	Standard model					New model				
		Est. (s.e.)	Bias	Acc.	Cov.	ACIL	Est. (s.e.)	Bias	Acc.	Cov.	ACIL
100	β_{trt}	0.127 (0.203)	0.023	0.042	0.954	0.811	0.154 (0.204)	–0.004	0.042	0.9522	0.820
β_{cov}		0.003 (0.002)					0.056 (0.028)	–0.006	0.001	0.916	0.094
α_{cov}							6.35 (3.682)	–1.35	15.381	0.968	13.206
250	β_{trt}	0.132 (0.130)	0.018	0.017	0.95	0.515	0.152 (0.135)	–0.002	0.018	0.938	0.507
β_{cov}		0.002 (0.001)					0.052 (0.016)	–0.002	0	0.932	0.055
α_{cov}							5.313 (1.120)	–0.313	1.353	0.968	4.33
500	β_{trt}	0.135 (0.092)	0.015	0.009	0.946	0.355	0.154 (0.091)	–0.004	0.008	0.938	0.356
β_{cov}		0.002 (0.001)					0.051 (0.010)	–0.001	0	0.932	0.037
α_{cov}							5.197 (0.725)	–0.197	0.565	0.966	2.911
1000	β_{trt}	0.131 (0.061)	0.019	0.004	0.954	0.25	0.150 (0.061)	0	0.004	0.964	0.25
β_{cov}		0.001 (0.001)					0.050 (0.007))	0	0	0.914	0.026
α_{cov}							5.067 (0.493)	–0.067	0.247	0.962	1.984
N	**Param.**	**log transformed models**					**Fractional polynomial model**				
		Est. (s.e.)	**Bias**	**Acc.**	**Cov.**	**ACIL**	**Est. (s.e.)**	**Bias**	**Acc.**	**Cov.**	**ACIL**
100	β_{trt}	0.156 (0.220)	-0.006	0.048	0.952	0.813	0.122 (0.249)	0.028	0.063	0.914	0.777
β_{cov}		0.357 (0.083)									
α_{cov}											
250	β_{trt}	0.146 (0.129)	0.004	0.017	0.95	0.505	0.136 (0.169)	0.014	0.029	0.932	0.503
β_{cov}		0.344 (0.049)									
α_{cov}											
500	β_{trt}	0.147 (0.092)	0.003	0.008	0.948	0.355	0.14 (0.131)	0.01	0.017	0.928	0.353
β_{cov}		0.344 (0.035)									
α_{cov}											
1000	β_{trt}	0.145 (0.062)	0.005	0.004	0.95	0.25	0.137 (0.079)	0.013	0.006	0.95	0.249
β_{cov}		0.341 (0.025)									
α_{cov}											

The individual terms to the right in the overall sums give the Shoenfeld type residuals, a pair for each observed survival time. The residuals are not linked to x directly, but to the terms to the left within the overall sums.

Influence Function

The influence function measures the rate of change in a statistical functional when there is a small amount of contamination from another distribution and is defined as

$$I(x) = \lim_{\in \to 0}\left[\frac{T\{(1-\varepsilon)F + \varepsilon\delta_x - T(F)\}}{\in}\right],$$

where T is the statistical functional giving the parameter of interest, F is the underlying distribution of the data and δ_x is the contamination introduced into the distribution. Replacing F by F_n, the empirical distribution function, $T(F_n)$, will be the estimate of $T(F)$ and the corresponding empirical influence function will measure the dependence of the estimate on particular data values.

Read and Crepeau establish the influence function for the proportional hazards model. They show this to be

$$\mathrm{I} = \mathrm{A}^{-1}\left(\hat{\beta}\right)\delta_{\mathrm{i}}\left\{\mathrm{z}_{\mathrm{i}} - \sum_{\mathrm{Ri}} \mathrm{z}_j \exp\left(\beta^T \mathrm{z}_j\right) / \sum_{\mathrm{Ri}} \exp\left(\beta^T \mathrm{z}_j\right)\right\} + \mathrm{A}-1\left(\hat{\beta}\right) C_i\left(\hat{\beta}\right)$$

where

$$\mathrm{A}\left(\hat{\beta}\right) = \mathrm{n}^{-1}\sum_{\mathrm{i}=1}^{N}\delta_{\mathrm{i}}\left[\sum_{Ri} \mathrm{z}_j \mathrm{z}_j{}^{\mathrm{T}} \exp\left(\beta^T \mathrm{z}_j\right) / \sum_{Ri} \exp\left(\beta^T \mathrm{z}_j\right) - \left\{\sum_{Ri} zj \exp\left(\beta^T \mathrm{z}_j\right) / \sum_{Ri} \exp\left(\beta^T \mathrm{z}_j\right)\right\}\left\{\sum_{Ri} zj \exp\left(\beta^T \mathrm{z}_j\right) / \sum_{Ri} \exp\left(\beta^T \mathrm{z}_j\right)\right\}^{\mathrm{T}}\right]$$

and

$$Ci\left(\beta\right) = \exp\left(\beta^T z_i\right)\begin{pmatrix} \sum_{tj \le ti} \delta_j \\ \left[\sum_{Rj} \mathrm{z}_k \exp\left(\beta^T \mathrm{z}_k\right) / \left\{\sum_{Rj} \exp\left(\beta^T \mathrm{z}_k^T\right)\right\}^2\right] \\ -\mathrm{z}_i \sum_{tj \le ti} \delta_j \left\{1 / \sum_{Rj} \exp\left(\beta^T \mathrm{z}_k\right)\right\} \end{pmatrix}$$

The algebra involved to arrive at this result is heavy and not particularly informative. A similar result was found for the robust models 1, 2, 3 and 4 where the algebra was even more involved and lengthy and so is not repeated here. Details are available from the authors and also will appear in a PhD thesis written by Jackson.

APPLICATION TO DATA FROM THE ESPAC 3 TRIAL

Robust model 4 was applied to data from the ESPAC-3 trial set up to investigate the effect of adjuvant chemotherapy on patients with resectable pancreatic cancer. Of particular interest are the group of patients who had pancreatic ductal adenocarcinomas (PDAC) and for whom a value of post operative CA19.9 was recorded (n=759). It is reasonably assumed that information for this covariate is missing completely at random and no bias is introduced by considering a complete case analysis. Previously published analyses [10] are followed, forcing the terms `Resection Margin' (Negative vs. Positive) and `Treatment Arm' (5FU vs. Gemcitabine) into the model as stratification factors. Also identified as important are `Lymph Nodes' (Negative vs. Positive), `Tumour Differentiation' (Poor vs. Moderate vs. Well) and `Smoking Status' (Never vs. Past vs. Present vs. Missing).

Figure **1** gives a histogram of CA19.9 values which is seen to have a very skewed distribution and prone to extreme value observations. The median (inter quartile range) is 24 (10, 63) but there are a number of observations greater than 1,000; only values up to 2,000 are displayed, the largest recorded being 37,000.

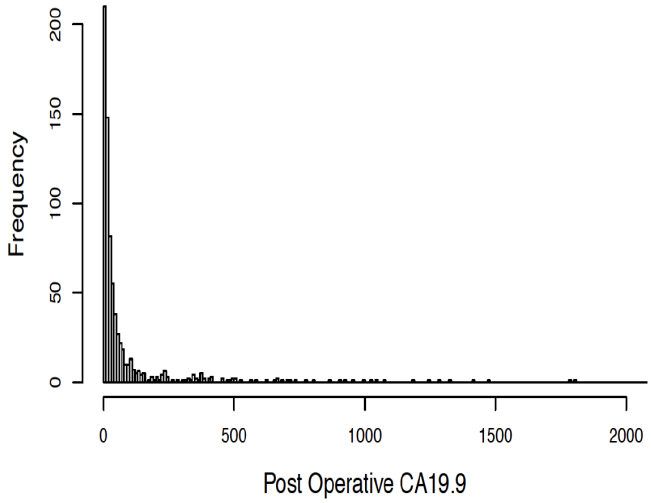

Figure 1: Histogram of post operative CA19.9 values.

Four models were fitted to the data, (i) standard Cox proportional hazards with raw CA19.9 data (Reference model), (ii) standard Cox proportional hazards with log transformed CA19.9 (Log model), (iii) a fractional polynomial model for CA19.9 (Frac. polyn. model) and (iv) robust model 4(Robust model). Table **3** shows the log-likelihood, Akaiki's information criterion (AIC), the model coefficients and their estimated standard errors. For the reference model, small estimates of β and for the estimated standard error are obtained. This is a consequence of the large extreme values observed. Taking as an example, the median value for CA19.9 as 24, a hazard ratio of 1.02 is obtained showing very modest increases in the baseline hazard. For extreme values of 2,000, 5,000 and 37,000, hazard ratios of 1.17, 1.49 and 18.9 are obtained. A clinician, however, may find it difficult to believe that a patient with CA19.9 value of 37,000 has an instantaneous risk of death of almost 20 times that of a patient with a zero value. The log-transformed model gives an improved model fit as shown by an AIC of 6882 compared to 6912 for the reference model. The hazard ratios for the reference values of 24, 2,000, 5,000 and 37,000 for CA19.9 are 1.95, 4.93, 5.90 and 9.11 respectively. Here, extreme hazard ratios are avoided to a small extent. Patients with a median value of CA19.9 are almost twice more likely to die at any given time point as those with a zero value.

Table 3: Results of Fitting Four Models to the ESPAC 3 Pancreatic Adenocarcinoma Survival Data

		Model							
		(i) Reference		(ii) Log		(iii) Frac. polyn.		(iv) Robust model 2	
		Log-lik.	AIC	Log-lik.	AIC	Log-lik.	AIC	Log-lik.	AIC
		-3447	6912	-3432	6882	-3412	6847	-3420	6861
Factor	Level	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Resec. Margin		Neg.							
	Pos.	0.21	0.09	0.19	0.09	0.20	0.09	0.18	0.09
Treatment		5FU							
	Gem.	-0.12	0.08	-0.10	0.08	-0.11	0.08	-0.09	0.08
Lymph N.		Neg.							
	Pos.	0.55	0.10	0.48	0.10	0.46	0.10	0.46	0.10
Tumour Diff.		Poor							
	Mod.	-0.29	0.10	-0.29	0.10	-0.27	0.10	-0.30	0.10
	Well	-0.64	0.15	-0.62	0.15	-0.69	0.15	-0.63	0.15
Smoke		Never							
	Past	0.09	0.10	0.08	0.10	0.07	0.10	0.08	0.10
	Present	0.24	0.12	0.26	0.12	0.27	0.12	0.27	0.12
	Missing	0.22	0.18	0.22	0.18	0.21	0.18	0.17	0.18
CA19.9		α				β_1=0.02	3.87e-3	3.77	0.70
	β	7.95e-5	1.54e-5	0.21	0.03	β_2=0.32	0.03	0.01	0.00

The fractional polynomial fitted was $\beta_1 \times 100/(CA19.9+1) + \beta_2 \times \log\{(CA19.9+1)/100$.

The fractional polynomial model had the lowest AIC with a value of 6847. The fractional polynomial that was produced was

$$\beta_1 \times \frac{100}{CA19.9+1} + \beta_2 \times \log\left\{\frac{(CA19.9+1)}{100}\right\}$$

This pair of transformations chosen by the fractional polynomial software might not be interpretable by clinicians and one problem with fractional polynomial regression is that a new data set generated under the same conditions can easily give rise to different transformations. As an illustration, the ESPAC 3 data were randomly split into two equal sized subsets of the data and fractional polynomial models fitted separately to both. The functional form of the two fractional polynomials differed. They were

$$\beta_1 \times \{(CA19.9+1)/1000\}^{-0.5} + \beta_2 \times \log\{(CA19.9+1)/1000\}$$

and

$$\beta_1 \times \{(CA19.9+1)/100\}^{-2} + \beta_2 \times \log\{(CA19.9+1)/100\}$$

Returning to the fractional polynomial model fitted to the whole dataset, for the reference points of 24, 2,000, 5,000 and 37,000,hazard ratios of 1.73, 3.64, 4.55 and 7.79 are obtained.

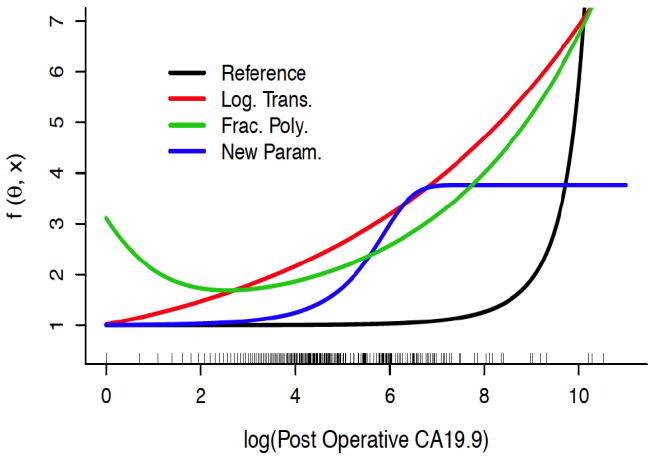

Figure 2: Hazard ratios plotted against log (CA19.9) for the four fitted models.

Model 4 has an AIC value of 6861 which is less than that for the log transformed model but more than that for the fractional polynomial model. The upper asymptote is 3.77 which corresponds to a maximum

hazard ratio of also 3.77. A hazard ratio of 1.19 is obtained for the median CA19.9 value of 24. The other reference values of 2,000, 5,000 and 37,000 all have a hazard ratio of 3.77 obtained from the upper asymptote. From a clinical perspective, this is the most attractive model with modest small increases in the CA19.9 resulting in modest increases in the hazard ratio and larger values curtailed to ensure that unrealistically large hazard ratios are not obtained. This is highlighted in Figure **2** where the hazard ratio is plotted against log (CA19.9) for all four models. The hazard ratio for the log transformed model follows an exponential curve while for the standard Cox model the hazard ratio follows a curve exp (exp(x)) because the x-axis is on the log-scale. For the fractional polynomial model the hazard ratio first decreases and then increases which is unrealistic in practice and could make clinical interpretations troublesome. Furthermore, there is no value of CA19.9 that has zero effect on the baseline hazard function within the observed range of data and this may affect confounding in other covariates as the baseline hazard function is amended to account for this. This can be seen somewhat in the analysis of the ESPAC-3 dataset with some amendments in the point estimates, especially for the Tumour Differentiation covariates. Model 2 has the desired shape of curve for the hazard function and is bounded whereas all other models can have the hazard ratio increase indefinitely.

Model diagnostics

Model diagnostics were calculated for the four models, firstly residuals and then influence measures. Figure **3** shows Shoenfeld residuals for the parameters associated with CA19.9 for the four models fitted. The scales on the graphs are not comparable. The residuals from the extreme values can be seen in the plot for the reference model and also the two plots for the fractional polynomial model. The variance of residuals for the log model decreases as survival time increases. The two plots for robust model **4** show how the asymptote controls the residuals and also shows that the variability of the residuals with time is much less than for the other models.

Figure **4** shows the influence measures obtained for the four models, plotted against log (CA19.9) and where crosses mark observed events and circles censored events. For the reference model, there is no obvious relationship of the influence measures with

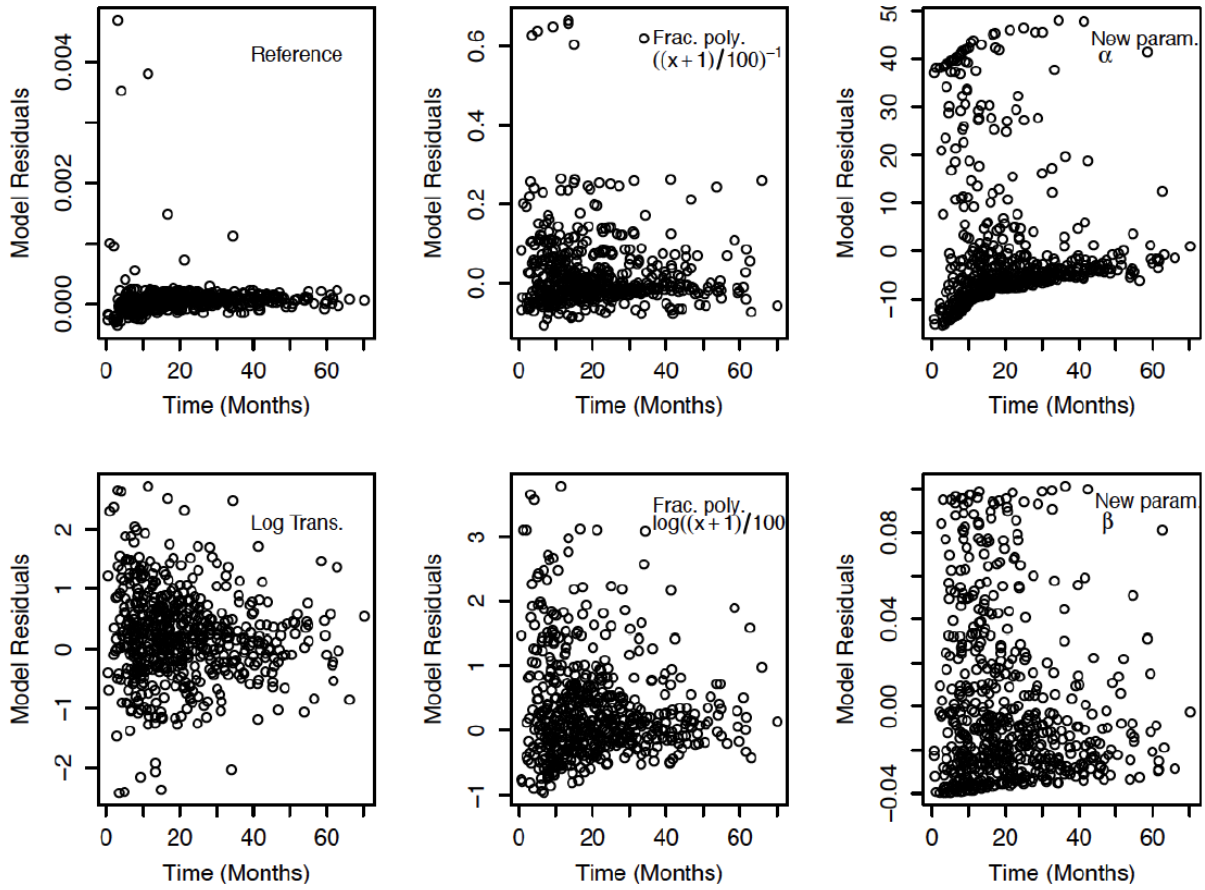

Figure 3: Schoenfeld residuals for the four models.

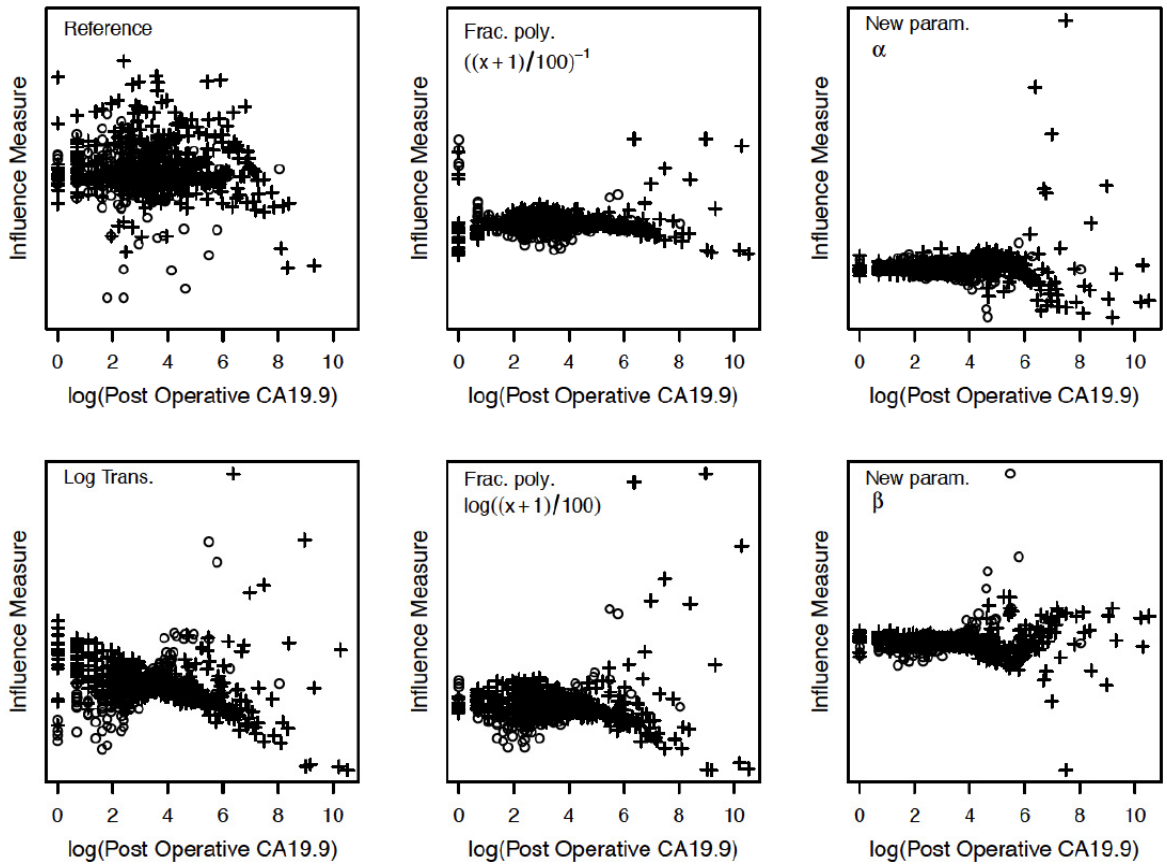

Figure 4: Influence measures for the four models.

CA19.9, whereas for the log-transformed model there is a central point of CA19.9 around the value 4. Either side of this point, there is a general divergence with both small and large values of CA19.9 having relatively large effects upon parameter estimation. For the fractional polynomial model, the influence measures associated with the $\{(x+1)/100\}^{-1}$ term show that very small values of CA19.9 can have a disproportionally large influence upon parameter estimation. There are also some large positive influence measures associated with log (CA19.9) values greater than 6. For the term given by log {(x+1)/100}, there is a similar relationship to that seen for the log transformed model although here the divergence from some central point is less pronounced. Large values of CA19.9 are again associated with typically large, positive influence measures. For robust model 4, there is neither the divergence away from some central point, nor any large influence measures associated with small values of CA19.9. However, the parameter α is associated with some large positive influence measures. This is to be expected, as this is the parameter associated with setting the upper asymptote. The estimate of the parameter is driven by the amount of `large' data that are observed and any single data value can have a relatively large effect on the estimate. Upon first inspection of the plot for β, apart from two large influence measures, there is fairly flat relationship with log(CA19.9), even at large values. Upon closer inspection, there is some change in the relationship between log (CA19.9) values of 4 and 6, and again between 6 and 8. This change, although small, can be seen to correspond to the points in the function that are chiefly concerned with the growth of the functional relationship and immediately afterwards as the asymptote is approached.

DISCUSSION

A method has been given to robustly model a continuous covariate in the Cox proportional hazards situation that automatically guards against extreme values and sets asymptotes for the minimum and maximum hazard ratios. This can be very useful in the clinical context. The model was successfully demonstrated on survival data following resection for pancreatic adenocarcinoma where CA19.9 is used as a biomarker. The distribution of CA19.9 is highly skewed and so was a good candidate for the robust parameterization.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. JP Neoptolemos for permission to use the ESPAC 3 trial data.

REFERENCES

[1] Cox DR. Regression Models and Life-Tables. JR Stat Soc Series B 1972; 34: 187-220.

[2] Viviani S, Farcomeni A. Robust estimation for the Cox regression model based on trimming. Biometrical Journal 2011; 53: 956-973. http://dx.doi.org/10.1002/bimj.201100008

[3] Royston P, Lambert PC. Flexible parametric survival analysis using Stata: Beyond the Cox model. Texas: Stata Press 2011.

[4] Farcomeni A, Ventura L. An overview of robust methods in medical research. Stat Methods Med Res 2010; 21: 111-133. http://dx.doi.org/10.1177/0962280210385865

[5] Bednarski T. Robust estimation in Cox's Regression models. Scand J Stat 1993; 20: 213-225.

[6] Minder CE, Bednarski T. A robust method for proportional hazards regression. Statist Med 1996; 15: 1033-1047. http://dx.doi.org/10.1002/(SICI)1097-0258(19960530)15:10<1033::AID-SIM215>3.0.CO;2-Y

[7] Burton A, Altman DG, Royston P, *et al*. The design of simulation studies in medical statistics. Statist Med 2006; 25: 4279-4292. http://dx.doi.org/10.1002/sim.2673

[8] Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika1982; 69, 239-241. http://dx.doi.org/10.1093/biomet/69.1.239

[9] Reid N, Crepeau H. Influence function for proportional hazards regression. Biometrika1985; 72: 1-9. http://dx.doi.org/10.1093/biomet/72.1.1

[10] Neoptolemos JP, Stocken DD, Bassi C, *et al*. Adjuvant Chemotherapy With Fluorouracil Plus Folinic Acid vs Gemcitabine Following Pancreatic Cancer Resection: A Randomized Controlled Trial. JAMA 2010; 304: 1073-1081. http://dx.doi.org/10.1001/jama.2010.1275

Received on 14-07-2014 Accepted on 03-09-2014 Published on 06-11-2014

http://dx.doi.org/10.6000/1929-6029.2014.03.04.1

Bibliography

[1] J. Whitehead, "The case for frequentism in clinical trials.," *Statistics in medicine*, vol. 12, no. 15-16, pp. 1405–1413; discussion 1415–1419, 1993.

[2] L. Amiri-Kordistani and T. Fojo, "Why do phase III clinical trials in oncology fail so often," *Journal of the National Cancer Institute*, vol. 104, no. 8, pp. 1–2, 2012.

[3] D. A. Berry, "Bayesian Statistics and the Efficiency and Ethics of Clinical Trials," *Statistical Science*, vol. 19, no. 1, pp. 175–187, 2004.

[4] D. A. Berry, "Bayesian clinical trials.," *Nature reviews. Drug discovery*, vol. 5, no. 1, pp. 27–36, 2006.

[5] T. V. Perneger and D. S. Courvoisier, "Interpretation of evidence in data by untrained medical students: a scenario-based study.," *BMC medical research methodology*, vol. 10, no. 1, p. 78, 2010.

[6] R. Christensen, "Testing Fisher, Neyman, Pearson, and Bayes," *The American Statistician*, vol. 59, no. 2, pp. 121–126, 2005.

[7] S. Senn, "Bayesian, likelihood and frequentist approaches to statistics," *Applied Clinical Trials*, pp. 35–38, 2003.

[8] J. O. Berger, "Could Fisher, Jeffreys and Neyman Have Agreed on Testing?," *Statistical Science*, vol. 18, no. 1, pp. 1–32, 2003.

[9] T. R. Fleming and D. Y. Lin, "Survival analysis in clinical trials: past developments and future directions.," *Biometrics*, vol. 56, no. December, pp. 971–983, 2000.

[10] M. Gönen, "Bayesian clinical trials: no more excuses.," *Clinical trials (London, England)*, vol. 6, no. 3, pp. 203–204, 2009.

[11] G. Howard, C. S. Coffey, and G. R. Cutter, "Is Bayesian analysis ready for use in phase III randomized clinical trials? Beware the sound of the sirens.," *Stroke; a journal of cerebral circulation*, vol. 36, no. 7, pp. 1622–1623, 2005.

[12] R. J. Lewis and R. L. Wears, "An introduction to the Bayesian analysis of clinical trials.," *Annals of emergency medicine*, vol. 22, no. 8, pp. 1328–1336, 1993.

[13] J. Herson, "Bayesian analysis of cancer clinical trials: An introduction to four papers," *Statistics in Medicine*, vol. 11, no. 1, pp. 1–3, 1992.

[14] D. O. Dixon and R. Simon, "Bayesian subset analysis in a colorectal cancer clinical trial," *Statistics in Medicine*, vol. 11, no. 1, pp. 13–22, 1992.

[15] L. S. Freedman and D. J. Spiegelhalter, "Application of Bayesian statistics to decision making during a clinical trial.," *Statistics in medicine*, vol. 11, no. 1, pp. 23–35, 1992.

[16] J. B. Greenhouse, "On some applications of Bayesian methods in cancer clinical trials.," *Statistics in medicine*, vol. 11, no. 1, pp. 37–53, 1992.

[17] S. Wieand and S. Cha, "Description of the statistical aspects of a study for advanced colorectal cancer patients.," Tech. Rep. 1, Cancer Center Statistics, Mayo Clinic, Rochester, MN 55905., 1992.

[18] D. Ashby, "Bayesian statistics in medicine: a 25 year review," *Statistics in medicine*, vol. 25, no. 21, pp. 3589–3631, 2006.

[19] A. P. Grieve, "25 years of Bayesian methods in the pharmaceutical industry: A personal, statistical bummel," *Pharmaceutical Statistics*, vol. 6, no. 4, pp. 261–281, 2007.

[20] J. Berger, "The case for objective Bayesian analysis," *Bayesian Analysis*, vol. 1, no. 3, pp. 385–402, 2006.

[21] D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles, *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, 2004.

[22] K. R. Abrams, *Bayesian Survival Analysis*. PhD thesis, University of Liverpool, 1992.

[23] K. Abrams, D. Ashby, and D. Errington, "Simple Bayesian analysis in clinical trials: a tutorial.," *Controlled clinical trials*, vol. 15, no. 5, pp. 349–359, 1994.

[24] S. N. Goodman, "Introduction to Bayesian methods I: measuring the strength of evidence," *Clinical Trials*, vol. 2, no. 4, pp. 282–290, 2005.

[25] T. a. Louis, "Introduction to Bayesian methods II: fundamental concepts," *Clinical Trials*, vol. 2, no. 4, pp. 291–294, 2005.

[26] D. A. Berry, "Introduction to Bayesian methods III: use and interpretation of Bayesian tools in design and analysis.," *Clinical trials (London, England)*, vol. 2, no. 4, pp. 295–300; discussion 301–304, 364–378, 2005.

[27] L. a. Moyé, "Bayesians in clinical trials: Asleep at the switch," *Statistics in Medicine*, vol. 27, no. 4, pp. 469–482, 2008.

[28] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics.," *CA: a cancer journal for clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[29] N. Alexakis, C. Halloran, M. Raraty, P. Ghaneh, R. Sutton, and J. P. Neoptolemos, "Current standards of surgery for pancreatic cancer," *British Journal of Surgery*, vol. 91, no. 11, pp. 1410–1427, 2004.

[30] E. L. Kaplan and P. Meier, "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958.

[31] M. Greenwood, "A Report on the Natural Duration of Cancer.," *JAMA: The Journal of the American Medical Association*, vol. 88, no. 7, pp. 507–507, 1927.

[32] J. P. Neoptolemos, D. D. Stocken, C. Bassi, P. Ghaneh, D. Cunningham, D. Goldstein, R. Padbury, M. J. Moore, S. Gallinger, C. Mariette, M. N. Wente, J. R. Izbicki, H. Friess, M. M. Lerch, C. Dervenis, A. Oláh, G. Butturini, R. Doi, P. a. Lind, D. Smith, J. W. Valle, D. H. Palmer, J. a. Buckels, J. Thompson, C. J. McKay, C. L. Rawcliffe, and M. W. Büchler, "Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: a randomized controlled trial.," *JAMA : the journal of the American Medical Association*, vol. 304, no. 10, pp. 1073–1081, 2010.

[33] J. P. Neoptolemos, J. a. Dunn, D. D. Stocken, J. Almond, K. Link, H. Beger, C. Bassi, M. Falconi, P. Pederzoli, C. Dervenis, L. Fernandez-Cruz, F. Lacaine, a. Pap, D. Spooner, D. J. Kerr, H. Friess, and M. W. Büchler, "Adjuvant chemoradiotherapy and chemotherapy in resectable pancreatic cancer: A randomised controlled trial," Tech. Rep. 9293, Department of Surgery, Liverpool University, Liverpool, UK. j.p.neoptplemos@liverpool.ac.uk, 2001.

[34] M. J. Moore, "Effect of Adjuvant Chemotherapy With Fluorouracil Plus Folinic Acid or Gemcitabine vs Observation on Survival in Patients With Resected Periampullary Adenocarcinoma¡subtitle¿The ESPAC-3 Periampullary Cancer Randomized Trial¡/subtitle¿¡alt-title¿Adjuvant T," *JAMA: The Journal of the American Medical Association*, vol. 308, no. 2, p. 147, 2012.

[35] J. W. Valle, D. Palmer, R. Jackson, T. Cox, J. P. Neoptolemos, P. Ghaneh, C. L. Rawcliffe, C. Bassi, D. D. Stocken, D. Cunningham, D. O'Reilly, D. Goldstein, B. a. Robinson, C. Karapetis, A. Scarfe, F. Lacaine, J. Sand, J. R. Izbicki, J. Mayerle, C. Dervenis, A. Oláh, G. Butturini, P. a. Lind, M. R. Middleton, A. Anthoney, K. Sumpter, R. Carter, and M. W. Büchler, "Optimal duration and timing of adjuvant chemotherapy after definitive surgery for ductal adenocarcinoma of the pancreas: Ongoing lessons from the ESPAC-3 study," *Journal of Clinical Oncology*, vol. 32, pp. 504–512, 2014.

[36] P. S. Schein and G. T. S. Group, "A comparison of combination chemotherapy and combined modality therapy for locally advanced gastric carcinoma," *Cancer*, vol. 49, no. 9, pp. 1771–1777, 1982.

[37] D. R. Cox, "Regression models and life tables," *Royal Statistical Society. Series B*, vol. 34, no. 2, pp. 187–220, 1972.

[38] M. Friedman, "Piecewise Exponential Models for Survival Data with Covariates," *The Annals of Statistics*, vol. 10, no. 1, pp. 101–113, 1982.

[39] S. T. Liu, *Counting Processes and Survival Analysis*, vol. 49. Wiley, 2007.

[40] P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes : A large sample study," *The Annals of Statistics*, vol. 10, no. 4, pp. 1100–1120, 1982.

[41] A. Gelfand, "Sampling-Based Approaches to Calculating Marginal Densities," *Journal Of The American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[42] M. Bayes and M. Price, "An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.," *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370–418, 1763.

[43] G. Rodriguez, "Parametric survival models," in *Regression Modeling Strategies*, Statistics for Biology and Health, pp. 1–14, Springer New York, 2010.

[44] D. Collett, *Modelling Survival Data in Medical Research*, vol. 46. Chapman & Hall, 2004.

[45] T. P. Ryan and W. H. Woodall, "The most-cited statistical papers," 2005.

[46] B. Efron, "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, vol. 72, no. 359, pp. 557–565, 1977.

[47] N. Breslow, "Covariance analysis of censored survival data.," *Biometrics*, vol. 30, no. 1, pp. 89–99, 1974.

[48] M. S. Goodman, Y. Li, and R. C. Tiwari, "Detecting multiple change points in piecewise constant hazard functions," *Journal of Applied Statistics*, vol. 38, no. 11, pp. 2523–2532, 2011.

[49] D. Gamermant, "Dynamic Bayesian Models for Survival Data," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 40, no. 1, p. 63˜79, 1991.

[50] D. Gamerman, "Bayes estimation of the piece-wise exponential distribution," *IEEE Transactions on Reliability*, vol. 43, no. 1, pp. 128–131, 1994.

[51] D. Zelterman, P. M. Grambsch, C. T. Le, J. Z. Ma, and J. W. Curtsinger, "Piecewise exponential survival curves with smooth transitions.," *Mathematical biosciences*, vol. 120, no. 2, pp. 233–250, 1994.

[52] G. Malla and H. Mukerjee, "A new piecewise exponential estimator of a survival function," *Statistics and Probability Letters*, vol. 80, pp. 1911–1917, 2010.

[53] D. Sinha, M. H. Chen, and S. K. Ghosh, "Bayesian analysis and model selection for interval-censored survival data.," *Biometrics*, vol. 55, no. 2, pp. 585–590, 1999.

[54] M. C. Koissi and G. Högnäs, "Using WinBUGS to study family frailty in child mortality, with an application to child survival in Ivory Coast," *Etude de la Population Africaine*, vol. 20, no. 1, pp. 1–17, 2005.

[55] Y. Li, M. H. Gail, D. L. Preston, B. I. Graubard, and J. H. Lubin, "Piecewise exponential survival times and analysis of case-cohort data," *Statistics in Medicine*, vol. 31, no. 13, pp. 1361–1368, 2012.

[56] M. Crowther, R. D. Riley, J. Staessen, J. Wang, F. Gueyffier, and P. C. Lambert, "Individual patient data meta-analysis of survival data using Poisson regression models," *BMC Medical Research Methodology*, vol. 12, no. 1, p. 34, 2012.

[57] J. D. Kalbfleisch, "Non-parametric Bayesian Analysis of Survival Time Data," *Royal Statistical Society. Series B*, vol. 40, no. 2, pp. 214–221, 1978.

[58] J. Yan, *Bayesian Survival Analysis*, vol. 99. Springer, 2004.

[59] N. Breslow, "Covariance analysis of censored survival data.," *Biometrics*, vol. 30, no. 374, pp. 89–99, 1974.

[60] T. R. Fleming and D. P. Harrington, *Counting processes and survival analysis*, vol. 169. John Wiley & Sons, 2011.

[61] K. Andersen, *Statistical Models Based on Counting Processes (Google eBook)*. Springer, 1997.

[62] D. Clayton, "Some approaches to the analysis of recurrent event data.," *Statistical methods in medical research*, vol. 3, no. 3, pp. 244–262, 1994.

[63] D. G. Clayton, "A Monte Carlo method for Bayesian inference in frailty models.," *Biometrics*, vol. 47, no. 2, pp. 467–485, 1991.

[64] P.-S. Shen, "Semiparametric analysis of transformation models with doubly censored data," *Journal of Applied Statistics*, vol. 38, no. 3, pp. 675–682, 2011.

[65] S. C. Cheng, L. J. Wei, and Z. Ying, "Analysis of transformation models with censored data," *Biometrika*, vol. 82, no. 4, pp. 835–845, 1995.

[66] O. Aalen, "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, vol. 6, no. 4, pp. 701–726, 1978.

[67] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, 1984.

[68] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[69] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State by Fast Computing Machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[70] a. Gelman, D. Rubin, H. Stern, and J. Garlin, *Bayesian data analysis*. Chapman & Hall, 2014.

[71] P. Dellaportas and A. Smith, "Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling," *Applied Statistics*, vol. 42, no. 3, pp. 443–459, 1993.

[72] P. Royston, "Flexible parametric alternatives to the Cox model , and more," *Stata Journal*, vol. 1, no. 1, pp. 1–28, 2001.

[73] P. Royston and M. K. B. Parmar, "Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects," *Statistics in Medicine*, vol. 21, no. 15, pp. 2175–2197, 2002.

[74] S. Viviani, "Trimmed Cox regression for Robust Estimation in Survival Studies," *Dss.Uniroma1.It*, 2010.

[75] a. Farcomeni and L. Ventura, "An overview of robust methods in medical research," *Statistical Methods in Medical Research*, vol. 21, no. October, pp. 111–133, 2012.

[76] T. Bednarski, "Robust estimation in Cox's Regression models," *Scandinavian Jounal of Statistics*, vol. 20, no. 3, pp. 213–225, 1993.

[77] C. E. Minder and T. Bednarski, "A robust method for proportional hazards regression," *Statistics in Medicine*, vol. 15, no. 10, pp. 1033–1047, 1996.

[78] A. Farcomeni and S. Viviani, "Robust estimation for the Cox regression model based on trimming.," *Biometrical journal. Biometrische Zeitschrift*, vol. 53, no. 6, pp. 956–973, 2011.

[79] P. Royston and D. G. Altman, "Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling," *Applied Statistics*, pp. 429–467, 1994.

[80] D. G. Altman and P. Royston, "The cost of dichotomising continuous variables.," *BMJ (Clinical research ed.)*, vol. 332, p. 1080, 2006.

[81] P. Royston, D. G. Altman, and W. Sauerbrei, "Dichotomizing continuous predictors in multiple regression: A bad idea," *Statistics in Medicine*, vol. 25, no. 1, pp. 127–141, 2006.

[82] A. Burton, D. Altman, P. Royston, and R. Holder, "The design of simulation studies in medical statistics.," *Statistics in Medicine*, vol. 25, no. 24, pp. 4279–4292, 2006.

[83] D. Schoenfeld, "Partial Residuals for The Proportionnal Hazards Regression Model," *Biometrika*, vol. 69, no. 1, pp. 239–241, 1982.

[84] N. Reid and H. Crepeau, "Influence function for proportional hazards regression," *Biometrika*, vol. 72, no. 1, pp. 1–9, 1985.

[85] P. M. Grambsch and T. M. Therneau, "Proportional hazards test and diagnostics based on weighted residuals," *Biometrika*, vol. 81, no. 3, pp. 515–526, 1994.

[86] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, 1974.

[87] T. S. Mok, Y.-L. Wu, S. Thongprasert, C.-H. Yang, D.-T. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, Y. Nishiwaki, Y. Ohe, J.-J. Yang, B. Chewaskulyong, H. Jiang, E. L. Duffield, C. L. Watkins, A. a. Armour, and M. Fukuoka, "Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma.," *The New England journal of medicine*, vol. 361, no. 10, pp. 947–957, 2009.

[88] D. M. Stablein, W. H. Carter, and J. W. Novak, "Analysis of survival data with nonproportional hazard functions.," *Controlled clinical trials*, vol. 2, pp. 149–159, 1981.

[89] E. Arjas, "A graphical method for assessing goodness of fit in {Cox's} proportional hazards model," *J Am Stat Assoc*, vol. 83, no. 401, pp. 204–212, 1988.

[90] N. H. Ng'andu, "An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model," *Statistics in Medicine*, vol. 16, no. 6, pp. 611–626, 1997.

[91] F. Harrell, "The PHGLM procedure, SAS Supplemental Library User's Guide, Version 5," *Cary, NC: SAS Institute Inc*, 1986.

[92] T. M. Therneau, P. M. Grambsch, and T. R. Fleming, "Martingale based residuals for survival models," *Biometrika*, vol. 77, no. 1, pp. 147–160, 1990.

[93] T. Moreau, J. O'quigley, and M. Mesbah, "A global goodness-of-fit statistic for the proportional hazards model," *Applied Statistics*, pp. 212–218, 1985.

[94] T. Moreau, J. O. Quigley, and J. Lellouch, "On D . Schoenfeld ' s approach for testing the proportional hazards assumption," *Biometrika*, vol. 73, no. 2, pp. 513–515, 1986.

[95] D. M. Stablein, W. H. Carter, and J. W. Novak, "Analysis of survival data with nonproportional hazard functions.," *Controlled clinical trials*, vol. 2, no. 2, pp. 149–159, 1981.

[96] L. D. Fisher and D. Y. Lin, "Time-dependent covariates in the Cox proportional-hazards regression model.," *Annual review of public health*, vol. 20, no. 1, pp. 145–157, 1999.

[97] D. M. Zucker and A. F. Karr, "Nonparametric Survival Analysis with Time-Dependent Covariate Effects: A Penalized Partial Likelihood Approach," *The Annals of Statistics*, vol. 18, pp. 329–353, 1990.

[98] L. Tian, D. Zucker, and L. J. Wei, "On the Cox Model With Time-Varying Regression Coefficients," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 172–183, 2005.

[99] N. Ata and M. Sözer, "Cox regression models with nonproportional hazards applied to lung cancer survival data," *Hacettepe Journal of Mathematics and ...*, vol. 36, no. 2, pp. 157–167, 2007.

[100] C. a. Bellera, G. MacGrogan, M. Debled, C. T. de Lara, V. Brouste, and S. Mathoulin-Pélissier, "Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer.," *BMC medical research methodology*, vol. 10, p. 20, 2010.

[101] E. D. Lustbader, "Time-dependent covariates in survival analysis," *Biometrika*, vol. 67, no. 3, pp. 697–698, 1980.

[102] S. Murphy and P. Sen, "Time-dependent coefficients in a Cox-type regression model," *Stochastic Processes and their Applications*, vol. 39, no. 1, pp. 153–180, 1991.

[103] Z. Cai and Y. Sun, "Local Linear Estimation for TimeâĂŘDependent Coefficients in Cox's Regression Models," *Scandinavian Journal of Statistics*, vol. 30, no. 1, pp. 93–111, 2003.

[104] R. Giorgi, M. Abrahamowicz, C. Quantin, P. Bolard, J. Esteve, J. Gouvernet, and J. Faivre, "A relative survival regression model using B-spline functions to model non-proportional hazards," *Statistics in Medicine*, vol. 22, no. 17, pp. 2767–2784, 2003.

[105] R. Henderson, P. Diggle, and a. Dobson, "Joint modelling of longitudinal measurements and event time data.," *Biostatistics (Oxford, England)*, vol. 1, no. 4, pp. 465–480, 2000.

[106] M. D. Anastasios A. Tsiatis, "Joint modeling of longitudinal and time-to-event data: an overview," *Statistica Sinica*, vol. 14, no. 3, pp. 809–834, 2004.

[107] Y. K. Tseng, P. Hsieh, and J. L. Wang, "Joint modelling of accelerated failure time and longitudinal data," *Biometrika*, vol. 92, no. 3, pp. 587–603, 2005.

[108] Y.-y. Chi and J. G. Ibrahim, "Bayesian Approaches To Joint Longitudinal and Survival Models Accommodating Both Zero and Nonzero Cure Fractions," *Statistica Sinica*, vol. 17, no. 2, pp. 445–462, 2007.

[109] J. Herson, *The statistical analysis of failure time data*, vol. 2. John Wiley & Sons, 1981.

[110] P. Royston and M. K. B. Parmar, "The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt," *Statistics in Medicine*, vol. 30, no. 19, pp. 2409–2421, 2011.

[111] R. Xu and J. O'Quigley, "Estimating average regression effect under non-proportional hazards.," *Biostatistics (Oxford, England)*, vol. 1, no. 4, pp. 423–439, 2000.

[112] P. C. Lambert, J. R. Thompson, C. Weston, and D. P.W., "Estimating and modelling the cure fraction in population-based cancer survival analysis," *Biostatistics (Oxford, England)*, vol. 8, no. 3, pp. 576–594, 2009.

[113] G. Yin and J. G. Ibrahim, "A general class of Bayesian survival models with zero and nonzero cure fractions," *Biometrics*, vol. 61, no. 2, pp. 403–12, 2005.

[114] S. N. U. a. Kirmani and R. C. Gupta, "On the proportional odds model in survival analysis," *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 2, pp. 203–216, 2001.

[115] S. Bennett, "Analysis of survival data by the proportional odds model.," *Statistics in medicine*, vol. 2, no. 2, pp. 273–277, 1983.

[116] Y. Q. Chen, N. Hu, S.-C. Cheng, P. Musoke, and L. P. Zhao, "Estimating Regression Parameters in an Extended Proportional Odds Model.," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 318–330, 2012.

[117] S. A. Murphy, A. J. Rossini, and A. van der Vaart, "Maximum Likelihood Estimation in the Proportional Odds Model," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 968–976, 1997.

[118] O. Aalen, "A Model for Nonparametric Regression Analysis of Counting Processes," in *Mathematical Statistics and Probability Theory*, vol. 2 of *Lecture Notes in Statistics*, pp. 1–25, Springer New York, 1980.

[119] D. W. Hosmer and P. Royston, "Using AalenâĂŹs linear hazards model to investigate time-varying effects in the proportional hazards regression model," *The Stata Journal*, vol. 2, no. 4, pp. 331 – 350, 2002.

[120] D. E. Schaubel and G. Wei, "Fitting semiparametric additive hazards models using standard statistical software," *Biometrical Journal*, vol. 49, no. 5, pp. 719–730, 2007.

[121] A. Abadi, S. Saadat, P. Yavari, C. Bajdik, and P. Jalili, "Comparison of AalenâĂŹs Additive and Cox Proportional Hazards Models for Breast Cancer Survival: Analysis of Population Based Data from British Columbia, Canada," *Asian Pacific Journal of Cancer Prevention*, vol. 12, pp. 3113–3116, 2011.

[122] W. Conley, "Multivariate Analysis: Overview," *Ecology*, vol. 54, no. 7, p. 1409, 1973.

[123] T. F. Cox, "Testing the Equivalence of Survival Distributions using PP- and PPP-plots," *International Journal of Statisitcs in Medical Research*, vol. 3, pp. 161–173, 2014.

[124] C. Quantin, T. Moreau, B. Asselain, J. Maccario, and J. Lellouch, "A regression survival model for testing the proportional hazards hypothesis.," *Biometrics*, vol. 52, pp. 874–885, 1996.

[125] G. Yin and D. Zeng, "Efficient Algorithm for Computing Maximum Likelihood Estimates in Linear Transformation Models," *Journal of Computational and Graphical Statistics*, vol. 15, no. 1, pp. 228–245, 2006.

[126] S. a. Murphy and A. W. Van der Vaart, "On profile likelihood," *Journal Of The American Statistical Association*, vol. 95, no. 450, pp. 449–465, 2000.

[127] R. D. C. Team and R. R Development Core Team, "R: A language and environment for statistical computing," 2009.

[128] P. C. Austin, "Generating survival times to simulate Cox proportional hazards models with time-varying covariates," *Statistics in Medicine*, vol. 31, no. 11, pp. 3946–3958, 2012.

[129] J. G. Ibrahim, M. Chen, and D. Sinha, *Bayesian survival analysis*. Wiley Online Library, 2005.

[130] M. J. Symons, *Bayesian Inference for survival data with nonparametric hazards and vague priors*. PhD thesis, University of North Carolina, 1985.

[131] E. Arjas and D. Gasbarra, "Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler," *Statistica Sinica*, vol. 4, no. 1, pp. 505–524, 1994.

[132] S. K. Sahu, D. K. Dey, H. Aslanidou, and D. Sinha, "A Weibull regression model with gamma frailties for multivariate survival data.," *Lifetime data analysis*, vol. 3, no. 2, pp. 123–137, 1997.

[133] Z. Qiou, N. Ravishanker, and D. K. Dey, "Multivariate survival analysis with positive stable frailties.," *Biometrics*, vol. 55, no. 2, pp. 637–644, 1999.

[134] M. C. Paik, W. Y. Tsai, and R. Ottman, "Multivariate survival analysis using piecewise gamma frailty.," *Biometrics*, vol. 50, no. 4, pp. 975–988, 1994.

[135] M.-H. Chen, J. G. Ibrahim, and D. Sinha, "A New Bayesian Model for Survival Data with a Surviving Fraction," *Journal of the American Statistical Association*, vol. 94, no. 447, pp. 909–919, 1999.

[136] P. Gustafson, "A Bayesian analysis of bivariate survival data from a multicentre cancer clinical trial.," Tech. Rep. 23, Department of Statistics, University of British Columbia, Vancouver, Canada., 1995.

[137] G. Gómez, M. L. Calle, and R. Oller, "Frequentist and Bayesian approaches for interval-censored data," *Statistical Papers*, vol. 45, pp. 139–173, 2004.

[138] Y. Kim and J. Lee, "Bayesian analysis of proportional hazard models," *Annals of Statistics*, vol. 31, no. 2, pp. 493–511, 2003.

[139] D. Sinha and D. Dey, "Semiparametric Bayesian analysis of survival data," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 1195–1212, 1997.

[140] S. M. Berry, D. a. Berry, K. Natarajan, C.-S. Lin, C. H. Hennekens, and R. Belder, "Bayesian Survival Analysis With Nonproportional Hazards," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 36–44, 2004.

[141] J. He, D. L. McGee, and X. Niu, "Application of the Bayesian dynamic survival model in medicine," *Statistics in Medicine*, vol. 29, no. 3, pp. 347–360, 2010.

[142] I. K. Omurlu, K. Ozdamar, and M. Ture, "Comparison of Bayesian survival analysis and Cox regression analysis in simulated and breast cancer data sets," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11341–11346, 2009.

[143] M. K. B. Parmar, V. Torri, and L. Stewart, "Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints," *Statistics in Medicine*, vol. 17, no. 24, pp. 2815–2834, 1998.

[144] M. J. S. Gang Han and J. Kim, "Improved Survival Modeling Using A Piecewise Exponential Approach," *Statistics in medicine*, vol. 33, no. 1, pp. 1–21, 2009.

[145] F. N. Demarqui, R. H. Loschi, D. K. Dey, and E. a. Colosimo, "A class of dynamic piecewise exponential models with random time grid," *Journal of Statistical Planning and Inference*, vol. 142, pp. 728–742, 2012.

[146] F. N. Demarqui, R. H. Loschi, and E. a. Colosimo, "Estimating the grid of time-points for the piecewise exponential model," *Lifetime Data Analysis*, vol. 14, no. 3, pp. 333–356, 2008.

[147] M. J. Crowther and P. C. Lambert, "Simulating biologically plausible complex survival data," *Statistics in Medicine*, vol. 32, no. 23, pp. 4118–4134, 2013.

[148] R. J. Sylvester, "A Bayesian approach to the design of phase II clinical trials.," *Biometrics*, vol. 44, no. 3, pp. 823–836, 1988.

[149] S. Biswas, D. D. Liu, J. J. Lee, and D. a. Berry, "Bayesian clinical trials at the University of Texas M. D. Anderson Cancer Center.," *Clinical trials (London, England)*, vol. 6, no. 3, pp. 205–216, 2009.

[150] D. A. Berry, "A case for Bayesianism in clinical trials.," *Statistics in medicine*, vol. 12, no. 15-16, pp. 1377–1393; discussion 1395–1404, 1993.

[151] L. S. Freedman, D. J. Spiegelhalter, and M. K. B. Parmar, "Bayesian Approaches to Randomized Trials," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 157, no. 3, pp. pp. 357–416, 1994.

[152] M. D. Hughes, "Reporting Bayesian analyses of clinical trials.," *Statistics in medicine*, vol. 12, no. 18, pp. 1651–1663, 1993.

[153] D. A. Berry, "Statistical Innovations in Cancer Research," in *Cancer Medicine*, ch. 33, pp. 465–478, London: BC Decker, 6 ed., 2003.

[154] C. Gatsonis and J. B. Greenhouse, "Bayesian methods for phase I clinical trials.," *Statistics in medicine*, vol. 11, no. 10, pp. 1377–1389, 1992.

[155] J. Whitehead, Y. Zhou, A. Mander, S. Ritchie, A. Sabin, and A. Wright, "An evaluation of Bayesian designs for dose-escalation studies in healthy volunteers," *Statistics in Medicine*, vol. 25, no. 3, pp. 433–445, 2006.

[156] S. M. Berry, W. Spinelli, G. S. Littman, J. Z. Liang, P. Fardipour, D. a. Berry, R. J. Lewis, and M. Krams, "A Bayesian dose-finding trial with adaptive dose expansion to flexibly assess efficacy and safety of an investigational drug.," *Clinical trials (London, England)*, vol. 7, no. 2, pp. 121–135, 2010.

[157] S. K. Fan, Y. Lu, and Y. G. Wang, "A simple Bayesian decision-theoretic design for dose-finding trials," *Statistics in Medicine*, vol. 31, no. August 2011, pp. 3719–3730, 2012.

[158] S. Chevret, "The continual reassessment method in cancer phase I clinical trials: a simulation study.," *Statistics in medicine*, vol. 12, no. 12, pp. 1093–1108, 1993.

[159] P. F. Thall and E. H. Estey, "A Bayesian strategy for screening cancer treatments prior to phase II clinical evaluation.," *Statistics in medicine*, vol. 12, no. 13, pp. 1197–1211, 1993.

[160] S. B. Tan and D. Machin, "Bayesian two-stage designs for phase II clinical trials," *Statistics in Medicine*, vol. 21, no. 14, pp. 1991–2012, 2002.

[161] R. J. Lewis and D. a. Berry, "Group Sequential Clinical-Trials - a Classical Evaluation of Bayesian Decision-Theoretic Designs," 1994.

[162] U. Bandyopadhyay, A. Biswas, and R. Bhattacharya, "A Bayesian adaptive design for two-stage clinical trials with survival data," *Lifetime Data Analysis*, vol. 15, no. 4, pp. 468–492, 2009.

[163] L. Zhao, J. M. G. Taylor, and S. M. Schuetze, "Bayesian decision theoretic two-stage design in phase II clinical trials with survival endpoint," *Statistics in Medicine*, vol. 31, pp. 1804–1820, 2012.

[164] M. K. Parmar, G. O. Griffiths, D. J. Spiegelhalter, R. L. Souhami, D. G. Altman, and E. van der Scheuren, "Monitoring of large randomised clinical trials: a new approach with Bayesian methods.," *Lancet*, vol. 358, no. 9279, pp. 375–381, 2001.

[165] F. S. Resnic, K. H. Zou, D. V. Do, G. Apostolakis, and L. Ohno-Machado, "Exploration of a bayesian updating methodology to monitor the safety of interventional cardiovascular procedures.," *Medical decision making : an international journal of the Society for Medical Decision Making*, vol. 24, no. 4, pp. 399–407, 2004.

[166] D. A. Berry, P. Mueller, a. P. Grieve, M. Smith, T. Parke, R. Blazek, N. Mitchard, and M. Krams, "Adaptive Bayesian designs for dose-ranging drug trials," *Case studies in Bayesian statistics*, vol. 6, pp. 99–181, 2002.

[167] D. A. Berry, "Adaptive clinical trials: The promise and the caution," 2011.

[168] G. Yin, N. Chen, and J. J. Lee, "Phase II trial design with Bayean adaptive randomization and predictive probability," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 61, no. 2, pp. 219–235, 2012.

[169] L. Y. T. Inoue, P. F. Thall, and D. a. Berry, "Seamlessly expanding a randomized phase II trial to phase III.," *Biometrics*, vol. 58, no. 4, pp. 823–831, 2002.

[170] J. J. Lee and D. D. Liu, "A predictive probability design for phase II cancer clinical trials.," *Clinical trials (London, England)*, vol. 5, no. 2, pp. 93–106, 2008.

[171] S. Hong and L. Shi, "Predictive power to assist phase 3 go/no go decision based on phase 2 data on a different endpoint," *Statistics in Medicine*, vol. 31, no. 9, pp. 831–843, 2012.

[172] C. J. Adcock, "Sample size determination: a review," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 46, no. 2, pp. 261–283, 1997.

[173] L. Y. Inoue, D. a. Berry, and G. Parmigiani, "Relationship Between Bayesian and Frequentist Sample Size Determination," *The American Statistician*, vol. 59, no. 1, pp. 79–87, 2005.

[174] C. E. M'Lan, L. Joseph, and D. B. Wolfson, "Bayesian sample size determination for binomial proportions," 2008.

[175] L. Joseph and P. Belisle, "Bayesian sample size determination for normal means and differences between normal means," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 46, no. 2, pp. 209–226, 1997.

[176] a. L. Gould, "Sample sizes for event rate equivalence trials using prior information.," *Statistics in medicine*, vol. 12, no. 21, pp. 2009–2023, 1993.

[177] S. Gubbiotti and F. De Santis, "A Bayesian method for the choice of the sample size in equivalence trials," *Australian and New Zealand Journal of Statistics*, vol. 53, no. 4, pp. 443–460, 2011.

[178] F. Wang and A. E. Gelfand, "A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models," *Statistical Science*, vol. 17, no. 2, pp. 193–208, 2002.

[179] D. B. Rubin and H. S. Stern, "Sample Size Determination Using Posterior Predictive Distributions," *The Indian Journal of Statistics*, vol. 60, no. 1, pp. 161–175, 1998.

[180] D. V. Lindley, "The choice of sample size," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 46, pp. 129–138, 1997.

[181] T. Pham-Gia, "On Bayesian analysis, Bayesian decision theory and the sample size problem," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 46, no. 2, pp. 139–144, 1997.

[182] P. Bacchetti and J. M. Leung, *Sample size calculations in clinical research.*, vol. 97. Chapman & Hall, 2nd ed., 2002.

[183] D. Santis and R. La, "Using historical data for Bayesian sample size determination," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 170, no. 1, pp. 95–113, 2014.

[184] D. A. Schoenfeld, "Sample-size formula for the proportional-hazards regression model.," *Biometrics*, vol. 39, no. 2, pp. 499–503, 1983.

[185] L. S. Freedman, "Tables of the number of patients required in clinical trials using the logrank test.," *Statistics in medicine*, vol. 1, no. 2, pp. 121–129, 1982.

[186] D. Cunningham, I. Chau, D. D. Stocken, J. W. Valle, D. Smith, W. Steward, P. G. Harper, J. Dunn, C. Tudur-Smith, J. West, S. Falk, A. Crellin, F. Adab, J. Thompson, P. Leonard, J. Ostrowski, M. Eatock, W. Scheithauer, R. Herrmann, and J. P. Neoptolemos, "Phase III randomized comparison of gemcitabine versus gemcitabine plus capecitabine in patients with advanced pancreatic cancer," Tech. Rep. 33, Royal Marsden National HealthService (NHS) Foundation Trust, London and Surrey, United Kingdom. david.cunningham@rmh.nhs.uk, 2009.

[187] R. Herrmann, G. Bodoky, T. Ruhstaller, B. Glimelius, E. Bajetta, J. Schüller, P. Saletti, J. Bauer, A. Figer, B. Pestalozzi, C.-H. Köhne, W. Mingrone, S. M. Stemmer, K. Tàmas, G. V. Kornek, D. Koeberle, S. Cina, J. Bernhard, D. Dietrich, and W. Scheithauer, "Gemcitabine plus capecitabine compared with gemcitabine alone in advanced pancreatic cancer: a randomized, multicenter, phase III trial of the Swiss Group for Clinical Cancer Research and the Central European Cooperative Oncology Group.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 25, no. 16, pp. 2212–2217, 2007.

[188] E. Van Cutsem, W. L. Vervenne, J. Bennouna, Y. Humblet, S. Gill, J. L. Van Laethem, C. Verslype, W. Scheithauer, A. Shang, J. Cosaert, and M. J. Moore, "Phase III trial of bevacizumab in combination with gemcitabine and erlotinib in patients with metastatic pancreatic cancer," Tech. Rep. 13, University Hospital Gasthuisberg/Leuven, Digestive Oncology Unit, Herestraat 49, B-3000 Leuven, Belgium. eric.vancutsem@uz.kuleuven.ac.be, 2009.

[189] R. Simon, "Bayesian design and analysis of active control clinical trials.," *Biometrics*, vol. 55, no. 2, pp. 484–487, 1999.

[190] S. J. Pocock, "The combination of randomized and historical controls in clinical trials.," *Journal of chronic diseases*, vol. 29, pp. 175–188, 1976.

[191] H. Sacks, T. C. Chalmers, and H. Smith, "Randomized versus historical controls for clinical trials.," *The American journal of medicine*, vol. 72, no. 2, pp. 233–240, 1982.

[192] R. E. Tarone, "The use of historical control information in testing for a trend in Poisson means.," *Biometrics*, vol. 38, pp. 457–462, 1982.

[193] Y. Ma, J. Guo, N.-Z. Shi, and M.-L. Tang, "On the use of historical control information for trend test in carcinogenesis.," *Biometrics*, vol. 58, no. 4, pp. 917–927, 2002.

[194] Y. Ma, J. Guo, N.-Z. Shi, and M.-L. Tang, "On the use of historical control information for trend test in carcinogenesis.," *Biometrics*, vol. 58, no. 4, pp. 917–927, 2002.

[195] J. K. Haseman, "Data analysis: statistical analysis and use of historical control data.," *Regulatory toxicology and pharmacology : RTP*, vol. 21, no. 1, pp. 52–59; discussion 81–86, 1995.

[196] P. F. Thall and R. Simon, "Incorporating historical control data in planning phase II clinical trials.," *Statistics in medicine*, vol. 9, no. 3, pp. 215–228, 1990.

[197] R. E. Tarone, "The use of historical control information in testing for a trend in Poisson means.," *Biometrics*, vol. 38, no. 2, pp. 457–462, 1982.

[198] Y. Kikuchi and T. Yanagawa, "Incorporating historical information in testing for a trend in Poisson means," *Annals of the Institute of Statistical Mathematics*, vol. 40, no. 2, pp. 367–379, 1988.

[199] D. G. Chen, "Incorporating historical control information into quantal bioassay with Bayesian approach," *Computational Statistics and Data Analysis*, vol. 54, no. 6, pp. 1646–1656, 2010.

[200] A. Racine, A. P. Grieve, and H. Fluhler, "Bayesian methods in practice: experiences in the pharmaceutical industry with discussion.," *Applied Statistics*, vol. 35, no. 2, pp. 93–150, 1986.

[201] J. Ibrahim, L. Ryan, and M. Chen, "Using historical controls to adjust for covariates in trend tests for binary data," *Journal of the American Statistical Association*, vol. 93, no. 444, pp. 1282–1293, 1998.

[202] J. L. French and J. G. Ibrahim, "Bayesian methods for a three-state model for rodent carcinogenicity studies.," *Biometrics*, vol. 58, no. 4, pp. 906–916, 2002.

[203] M. H. Chen, J. G. Ibrahim, P. Lam, A. Yu, and Y. Zhang, "Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data," *Biometrics*, vol. 67, no. 3, pp. 1163–1170, 2011.

[204] F. De Santis, "Sample Size Determination for Robust Bayesian Analysis," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 278–291, 2006.

[205] B. Neuenschwander, G. Capkun-Niggli, M. Branson, and D. J. Spiegelhalter, "Summarizing historical information on controls in clinical trials.," *Clinical trials (London, England)*, vol. 7, no. 1, pp. 5–18, 2010.

[206] S. Gsteiger, B. Neuenschwander, F. Mercier, and H. Schmidli, "Using historical control information for the design and analysis of clinical trials with overdispersed count data," *Statistics in Medicine*, vol. 32, no. 21, pp. 3609–3622, 2013.

[207] J. D. Cook, J. M. Fúquene, and L. R. Pericchi, "Skeptical and Optimistic Robust Priors for Clinical Trials," *Revista Colombiana de Estadística*, vol. 34, no. 2, pp. 333–345, 2011.

[208] J. A. Fúquene, J. D. Cook, and L. R. Pericchi, "A case for robust bayesian priors with applications to clinical trials," *Bayesian Analysis*, vol. 4, no. 4, pp. 817–846, 2009.

[209] B. P. Hobbs, B. P. Carlin, S. J. Mandrekar, and D. J. Sargent, "Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials," *Biometrics*, vol. 67, no. 3, pp. 1047–1056, 2011.

[210] K. Viele, S. Berry, B. Neuenschwander, B. Amzal, F. Chen, N. Enas, B. Hobbs, J. G. Ibrahim, N. Kinnersley, S. Lindborg, S. Micallef, S. Roychoudhury, and L. Thompson, "Use of historical control data for assessing treatment effects in clinical trials," *Pharmaceutical Statistics*, vol. 13, no. 1, pp. 41–54, 2013.

[211] F. De Santis, "Power Priors and Their Use in Clinical Trials," *The American Statistician*, vol. 60, no. 2, pp. 122–129, 2006.

[212] K. F. Schulz and D. a. Grimes, "Unequal group sizes in randomised trials: Guarding against guessing," *Lancet*, vol. 359, no. 9310, pp. 966–970, 2002.

[213] J. C. Dumville, S. Hahn, J. N. V. Miles, and D. J. Torgerson, "The use of unequal randomisation ratios in clinical trials: A review," *Contemporary Clinical Trials*, vol. 27, no. 1, pp. 1–12, 2006.

[214] R. Vozdolska, M. Sano, P. Aisen, and S. D. Edland, "The net effect of alternative allocation ratios on recruitment time and trial cost.," *Clinical trials (London, England)*, vol. 6, no. 2, pp. 126–132, 2009.

[215] A. L. Avins, "Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials.," *Journal of medical ethics*, vol. 24, no. 6, pp. 401–408, 1998.

[216] S. W. Raudenbush, "Statistical analysis and optimal design for cluster randomized trials.," *Psychological Methods*, vol. 2, no. 2, pp. 173–185, 1997.

[217] S. W. Raudenbush and X. Liu, "Statistical power and optimal design for multisite randomized trials.," *Psychological methods*, vol. 5, no. 2, pp. 199–213, 2000.

[218] W. F. Rosenberger, N. Stallard, a. Ivanova, C. N. Harper, and M. L. Ricks, "Optimal adaptive designs for binary response trials.," *Biometrics*, vol. 57, no. 3, pp. 909–913, 2001.

[219] R. Brooks, "Optimal allocation for Bayesian inference about an odds ratio," *Biometrika*, vol. 74, no. 1, pp. 196–199, 1987.

[220] V. Sambucini, "Sample size determination for inferences on the odds ratio," *Journal of the Italian Statistical Society*, vol. 9, no. 1-3, pp. 219–243, 2000.

[221] E. L. Korn and B. Freidlin, "Outcome-adaptive randomization: Is it useful?," *Journal of Clinical Oncology*, vol. 29, no. 6, pp. 771–776, 2011.