Bayesian Survival Analysis

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

by

Keith Rowland Abrams

September 1992

Bayesian Survival Analysis

Keith Rowland Abrams

In cancer research the efficacy of a new treatment is often assessed by means of a clinical trial. In such trials the outcome measure of interest is usually time to death from entry into the study. The time to intermediate events may also be of interest, for example time to the spread of the disease to other organs (metastases). Thus, cancer clinical trials can be seen to generate multi-state data, in which patients may be in any one of a finite number of states at a particular time.

The classical analysis of data from cancer clinical trials uses a survival regression model. This type of model allows for the fact that patients in the trial will have been observed for different lengths of time and for some patients the time to the event of interest will not be observed (censored). The regression structure means that a measure of treatment effect can be obtained after allowing for other important factors.

Clinical trials are not conducted in isolation, but are part of an on-going learning process. In order to assess the *current* weight of evidence for the use of a particular treatment a Bayesian approach is necessary. Such an approach allows for the formal inclusion of prior information, either in the form of clinical expertise or the results from previous studies, into the statistical analysis.

An initial Bayesian analysis, for a single non-recurrent event, can be performed using non-temporal models that consider the occurrence of events up to a specific time from entry into the study. Although these models are conceptually simple, they do not explicitly allow for censoring or covariates.

In order to address both of these deficiencies a Bayesian fully parametric multiplicative intensity regression model is developed. The extra complexity of this model means that approximate integration techniques are required. Asymptotic Laplace approximations and the more computer intensive Gauss-Hermite quadrature are shown to perform well and yield virtually identical results.

By adopting counting process notation the multiplicative intensity model is extended to the multi-state scenario quite easily.

These models are used in the analysis of a cancer clinical trial to assess the efficacy of neutron therapy compared to standard photon therapy for patients with cancer of the pelvic region. In this trial there is prior information both in the form of clinical prior beliefs and results from previous studies. The usefulness of multi-state models is also demonstrated in the analysis of a pilot quality of life study.

Bayesian multi-state models are shown to provide a coherent framework for the analysis of clinical studies, both interventionist and observational, yielding clinically meaningful summaries about the current state of knowledge concerning the disease/treatment process.

Acknowledgements

I would like to thank my supervisors Deborah Ashby and Philip Brown for their encouragement and advice over the last three years.

I would also like to thank Philip Stell for advice, Doug Errington for permission to use the neutron therapy data and Kate Smith for permission to use the quality of life data. I would also like to thank Sheila Gore of the Medical Research Council Biostatistics Unit, Cambridge, for allowing me to use the results of the 'trial roulette' and meta analysis in the neutron therapy study.

Financial support was provided by the North West Cancer Research Fund to whom I am very grateful.

I am grateful to Mike West who originally sparked off my interest in Bayesian survival analysis whilst I was at Warwick University and David Clayton who sustained and encouraged this interest whilst I was at Leicester University.

I would also like to thank my colleagues at Liverpool University, most notably Mike Denham, Nancy Cleave, Bruno Sansó and especially Paul Manasse for reading and commenting on earlier drafts of this thesis.

I would like to thank all the patients and their relatives who took part in both the studies analysed in this thesis.

Finally, but by no means least, I would like to thank my family and friends for their support and encouragement.

To my husband
or
77 not out.

Blitzed by bomb and cancer,
Paralysed with fear;
Eyesight not very good,
Needs an aid to hear;
Not highly recommended,
Well past the sell-by date
That's my trade description
- Hard cheddar, my old mate.

So you're handy with the wheelchair?
But your cooking! Oh my dear!
And damn you for ignoring
My curses in your ear.

Lucky horses, so they say,
Are disposed of by a vet;
But give us a kiss, you great big twit
- I ain't going yet.

Ethel Dallin

BACUP News

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Aim

The aim of this thesis is to develop Bayesian parametric survival models that can be used to analyse phase III cancer clinical trials. Such models are required to be flexible to accommodate different types of events and multiple events. These models have a wide application in cancer clinical trials and observational studies.

1.2 Background

Cancer clinical trials are conducted to assess the efficacy of a new treatment compared to a standard one, with respect to a stated outcome measure. However, there are usually secondary outcome measures which are also of interest. In order that an overall picture is achieved of a new treatment's effect, statistical analysis needs to consider all outcome measures. For example, primary interest often focuses on survival, but secondary interest on treatment toxicity, tumour regression, metastatic disease and quality of life. Clinical cancer research is such that seldom are major advances made in terms of survival, and a new treatment may be expected to reduce toxicity, whilst maintaining survival. Thus statistical models are required that can be used not only for survival but also for the analysis of intermediate events, which may be recurrent.

Traditionally cancer clinical trials have been analysed in isolation, rather than as the current stage in the process of information accrual about a particular treatment. Before a cancer clinical trial is conducted there will have been evidence to suggest possible benefits of a new treatment. This process will have started with pharmacological or radiobiological theory, progressed to laboratory experiments and then to small phase II trials, (Pocock, 1983). Only if these studies have shown benefits will a larger phase III trial be conducted. In chemotherapy especially, 'new' treatments will be developed that are extensions of existing treatments, and there will be considerable information about the minimum likely benefit of a 'new' treatment from previous studies. Information will also be available in terms of clinical opinion, which to some extent will be influenced by previous work. Such clinical opinion about the possible direction and magnitude of treatment differences and clinically worthwhile differences are traditionally used in planning a clinical trial, but not in the subsequent analysis.

The existence of relevant prior information and the sequential nature of cancer clinical research make it an area to benefit from the use of Bayesian inference. Bayesian statistical models that could be used in a two-treatment cancer clinical trial are not new, Lindley (1965) and Box and Tiao (1973). Recent applications to clinical trials include Spiegelhalter and Freedman (1988), Pocock and Hughes (1990), Berry, Wolff and Sack (1992), Freedman and Spiegelhalter (1992), Dixon and Simon (1992), Greenhouse (1992) and Abrams, Ashby and Errington (Submitted). However, many of these methods do not address (i) explicit consideration of survival/censoring times, (ii) the inclusion of covariates,

and (iii) multiple events.

In the classical analysis of cancer clinical trials the main methods used have been the Kaplan-Meier survival curve proposed by Kaplan and Meier (1958) and the semi-parametric proportional hazards proposed by Cox (1972). The earliest Bayesian methods mirrored the development of the Kaplan-Meier survival curve in that they were non-parametric in nature, and considered patients to be homogeneous. It was not until the mid 1970s that regression methods similar to those of Cox (1972) were developed, in which the baseline hazard was considered to be non-parametric, Cornfield and Detre (1977) and Kalbfleisch (1978). More recently fully parametric models have been developed by Sweeting (1984, 1987) and Gamerman (1991).

Far from the Bayesian arena, interest in event history analysis in medical settings was spurred on by Aalen (1978) who proposed considering survival models, proportional and non-proportional hazards, both parametric and semi-parametric, as special cases of more general multiplicative or additive intensity models for multivariate counting processes. This meant that there was a general framework for event history data. Medical applications of such multi-state models have been described by Kay (1982, 1986). and Andersen (1988).

Therefore by adopting a counting process approach, Bayesian parametric multiplicative intensity models could be developed which could allow for the analysis of the multi-state data generated by cancer clinical trials, whilst also using the prior information which is often available.

1.3 Outline of Thesis

The outline of the thesis is as follows.

Chapter 2 describes two data sets which will be used throughout the thesis to illustrate various models that will be developed. The *neutron therapy data*, came from a cancer clinical trial to assess the efficacy of neutron therapy compared to conventional radiotherapy for the treatment of pelvic tumours. Details of the trial design and results are described. A *pilot quality of life study* was conducted at Clatterbridge hospital, Wirral. Details of the study are described and the data that was collected is displayed.

In Chapter 3 we consider the elicitation and quantification of prior information. We first consider the elicitation of clinical beliefs, and describe how this was performed in the neutron therapy trial described in Chapter 2. We also consider data based sources of prior information. Particular attention is paid to the use of the results from previous studies that are thought to be relevant to the current study. We describe the use of the results from six previous studies in neutron therapy. Finally we briefly consider Bayesian meta analyses.

Chapter 4 considers non-temporal models that may be used for analysis of survival data found in clinical trials. The first model is an odds model in which there is prior information about the twelve month mortality rates separately in two patient groups. Inferences about the ratio of odds of death in the two

groups can then be made. This model does not explicitly allow for differential follow-up or covariates. The odds model is special case of a logistic regression model, in which the response is death within twelve months. Finally, general methods based on assuming Normality are reviewed. These models still suffer from the deficiency that covariates cannot be allowed for.

Chapter 5 describes counting process notation and how this may be of use in describing and modelling event history type data. In particular the multiplicative intensity model is outlined.

Chapter 6 describes the development and application of fully parametric multiplicative intensity models when there are two states, the second of which is absorbing. This corresponds to failure-time data from cancer clinical trials. We consider the case in which the baseline intensity is constant, piecewise constant or has a Weibull parametric form. In the case when the baseline intensity is assumed to be constant over time, and there are only two patient groups, analytical results can be obtained. For other models we investigate the use of asymptotic approximation techniques and numerical integration methods in order to obtain parameter estimates. These models are compared with previously developed Bayesian survival models, and applied to the survival data in the neutron therapy study.

Chapter 7 considers situations where there are more than two states. After extending the the two state models of Chapter 6, two applications of multi-state models are described. For the neutron therapy data, we look at the development of metastases and their affect on subsequent survival. For the quality of life study, we consider patients moving between a finite number of quality of life states, with death as an absorbing state.

Finally, Chapter 8 summarises the techniques that have been developed and applied in the thesis, and outlines further work, including applications in epidemiology.

Chapter 2

Motivating Examples

2.1 Introduction

In this chapter we consider two motivating examples from clinical oncology. The first is a cancer clinical trial in which we require the comparison of two treatments in terms of morbidity as well as mortality, and the second is a pilot quality of life study in which a number of cancer patients have been followed up over a period of time and asked questions about their state of health, and other aspects of their lifestyle. In the quality of life study there is information on both survival and quality of life.

2.2 Cancer clinical trial for Neutron Therapy

2.2.1 Introduction

The objective of the trial was to compare high energy fast neutrons with conventional megavoltage x-rays (photons) for the treatment of pelvic carcinomas (cervix, bladder, rectum and prostate) in terms of death (without local control of symptoms), recurrence (after local control), radiation morbidity, and metastatic disease.

2.2.2 Background

Many tumours can be described as being radio-resistant. These are tumours that would require doses in excess of local tissue tolerances in order to produce any significant response. In this group of radio-resistant tumours are adenocarcinomas of the large bowel and carcinoma of the bladder. In these large tumour masses the central areas may be necrotic because of the absence of blood vessels and cells die. Towards the periphery of tumours the blood supply is usually good. In a zone between these two the degree of oxygenation may be just adequate for cells to survive, but low enough to reduce their sensitivity to standard radiotherapy. This concept has been used to explain the limited response of larger tumour masses, and the relatively high incidence of residual or recurrent tumour in them after radiation therapy.

However, the damaging effects of high energy (fast) neutrons are much less influenced by the degree of oxygenation of the tissues. This low dependence on oxygenation of cellular damage by neutrons is related to the mode of interaction with the atoms of the cells. Another factor favouring neutrons is the increased cellular damage at each exposure, and therefore the reduced possibility for cellular recovery.

To be of use in the treatment of tumours, neutron beams must be sufficiently penetrating and sufficiently dense to deliver the required dose in an acceptably short time. One method of producing such neutrons is via a cyclotron, in which positive ions are accelerated at a target of a low atomic number, which yields neutrons. The Medical Research Council neutron therapy study conducted at Clatterbridge hospital used a cyclotron to generate the neutron therapy beam.

The four sites were considered together principally because separately none of the trials would have developed sufficient statistical power due to lack of recruitment. Previous registry based evidence, (Griffin *et al.*, 1986), had shown that the potential benefits of fast energy neutrons over photons was in terms of morbidity, and therefore it was felt that although the tissue affected by treatment was different for the four sites the pooled data might give a clearer indication of whether morbidity was a serious problem.

2.2.3 Methods

For a detailed review of the methods used see Errington *et al* (1991) (Appendix C). As of 21st December 1990, 154 patients had been randomised to receive either neutron therapy or photon therapy. Randomisation took place in two phases. The first phase was from February 1986 to January 1988 when patients were randomised using a ratio of 3 to 1 in favour of neutrons, this was not stratified by site. It was performed using a block length of eight and sealed envelopes. The uneven randomisation ratio in favour of neutrons was designed to overcome the predicted problems in recruitment of patients into the trial. In the second phase of randomisation from January 1988 to February 1990 when the trial was stopped patients were randomised 1 to 1, but this was stratified by site, again using sealed envelopes and a permuted block length of four or six, determined by simple randomisation.

To be eligible for the study patients had to have histologically confirmed adenocarcinoma of the rectum or prostate, squamous cell carcinoma of the cervix or transitional cell carcinoma of the bladder and not to have been previously treated using either radiotherapy or chemotherapy. Patients who were over 80 years of age or had a Karnofsky index (Karnofsky and Burchenall, 1949) of less than 40 were also excluded, as were those having a previous history of a malignancy at another site or distant metastases. After randomisation all patients were staged using the TNM staging system (Harmer, 1978), and were T_{3a}, T_{3b} or T_4 and N_0, N_1, N_2 or N_x.

During the treatment period patients were followed-up weekly to assess their reactions to treatment. For the first year after treatment patients were recalled monthly, and in the following years they were recalled once every two to three months. In the statistical analysis reported in Errington *et al* (1991) the main outcome measures were mortality from all causes, and a secondary outcome measures were severe treatment toxicity, metastatic disease and tumour progression/regression. This analysis used Cox's proportional hazards model to allow for differential follow-up and to assess the effect of covariates on survival.

The detection of metastases at follow-up visits relied on clinical examination, CT scan, Magnetic resonance imaging, cystoscopy, sigmoidoscopy and biopsy. For this thesis a positive result using any one of the above detection methods was treated as metastatic disease.

In order to establish the ethicality of the trial, in terms of the individual clinicians involved, their prior beliefs about the possible benefits of high energy

neutrons were elicited using a 'trial roulette'. This aspect of the trial is reported in Errington *et al* (1991) and is also discussed in detail in Chapter 3.

A statistical overview of 5 previously published trials and one set of unpublished results was conducted. Errington *et al* (1991) report the results of this overview in detail, and these results are also presented and critiqued in Chapter 3, where we consider the use of such overviews as prior information that can be formally incorporated into a statistical analysis.

2.2.4 Results

Since the publication of the statistical analysis a further statistical analysis has been performed using a censoring date of 21st December 1990 rather than 26th January 1990 as in Errington *et al* (1991). Table 2.1 shows patient status at 21st December 1990.

	Photons	Neutrons	Total
Alive	37	46	83
Dead	25	46	71
Total	62	92	154

Table 2.1: Patient status at 12 month follow-up, as of 21st December 1990.

Figure 2.1 shows estimated Kaplan-Meier survival curves for neutrons and photons for all sites combined.

As reported in Errington *et al* using classical methods of inference, there is a statistically significant difference, $P = 0.02$, between the two treatment groups, not allowing for any covariates. This yields a hazard ratio of death for neutrons to photons of 1.71 with associated 95% confidence interval (1.10,2.92). This difference reduces slightly when other important covariates are included in the model. Preliminary cross-tabulations did not show the existence of patient imbalances with respect to covariates that were *a priori* thought to affect survival. Analysis using a censoring date of 21st December 1990 yields a still statistically significant difference, $P = 0.04$, but the hazard ratio of death for neutrons to photons reduces to 1.49 with associated 95% confidence interval (1.00,2.21). As before this difference reduces slightly when other important covariates are included in the model, Table 2.2.

Using Cox's proportional hazard model the hazard ratio for early severe reaction, within 3 months of treatment, of neutrons compared to photons is 1.30 with associated 95% confidence interval 0.49 to 3.45. For late severe reactions the hazard ratio is 1.18 with associated 95% confidence interval 0.53 to 2.56, (Errington *et al.*, 1991).

Of interest to clinicians is whether the two treatment groups develop metastases at different rates, and also whether having developed metastases the prognosis is different for the two treatment groups. One method for assessing the effect of developing metastases is to consider them as a time-dependent covari-

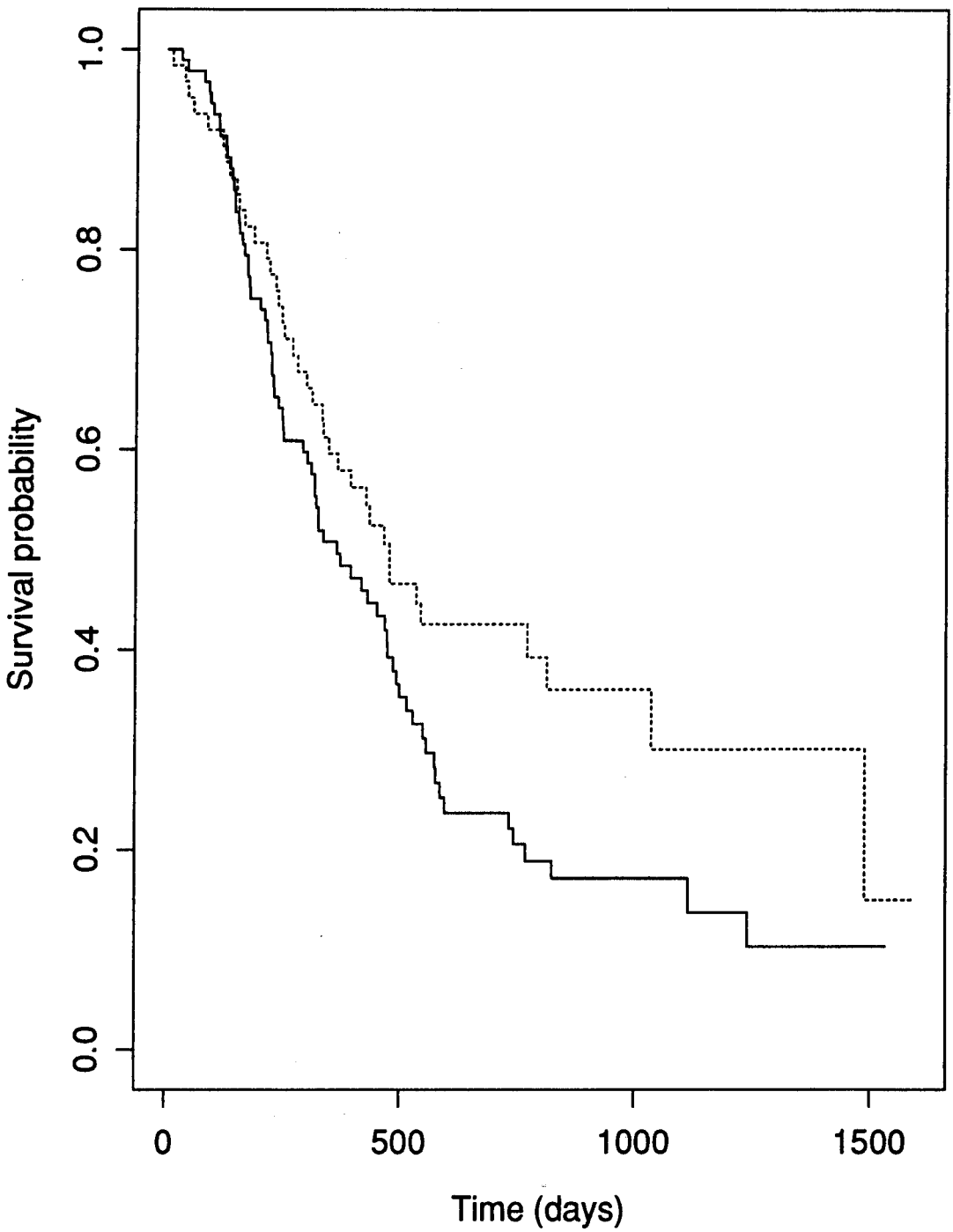


Figure 2.1: Kaplan-Meier estimated survival curves for neutron and photon patients in cancer clinical trial, with a censoring date of 21st December 1990.

Covariates	Hazard ratio (Neutrons to Photons)	95% CI[a]
None	1.49	(1.00,2.21)
Site	1.42	(0.95,2.12)
Phase	1.46	(0.97,2.18)
T stage + N stage + Karnofsky	1.44	(0.95,2.17)
Phase + T stage + N stage + Karnofsky	1.44	(0.95,2.20)

[a]CI denotes confidence interval for hazard ratio.

Table 2.2: Hazard ratios of death (neutrons to photons) allowing for various covariates using Cox's proportional hazards regression models, with censoring date 21st December 1990.

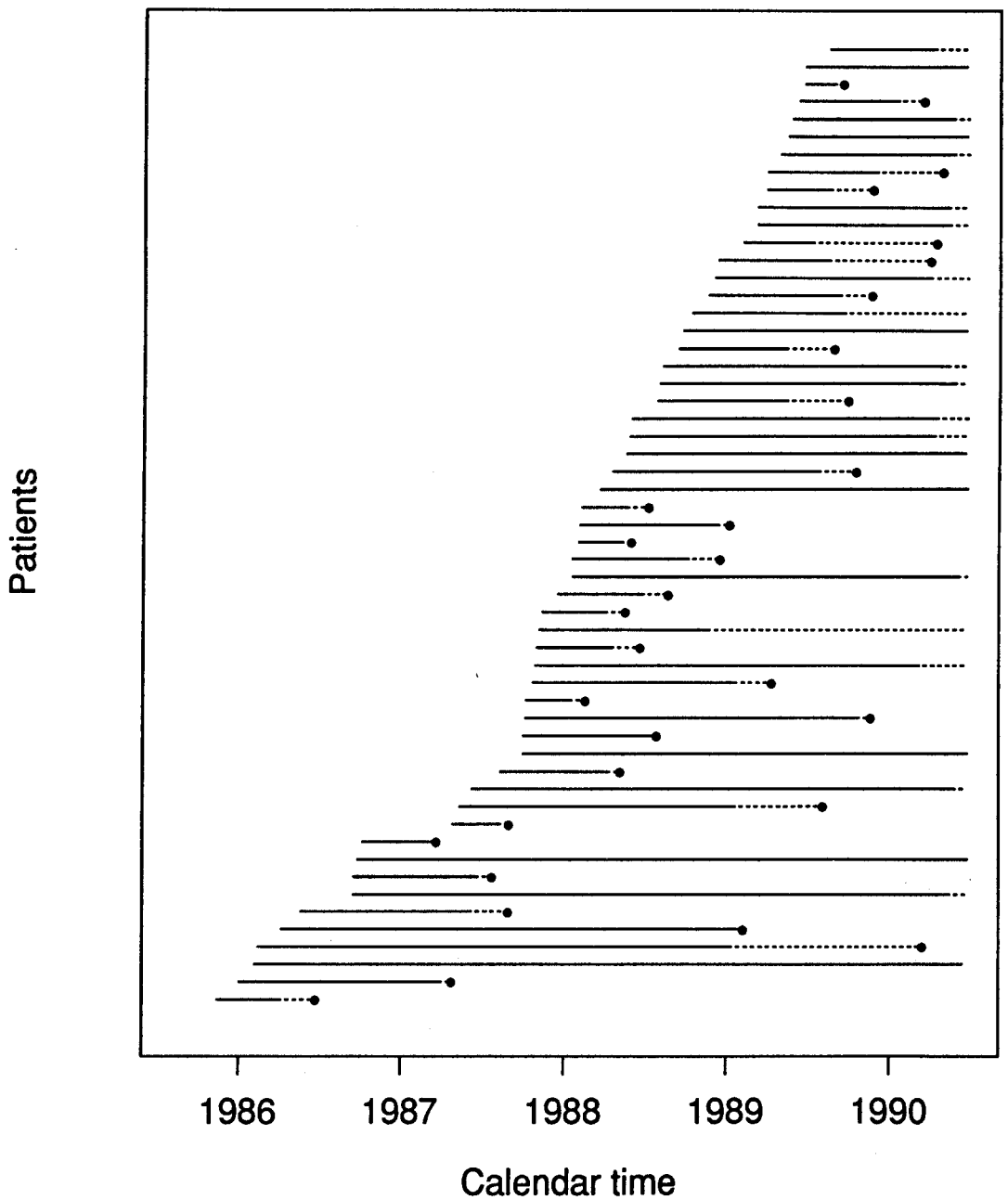
ate. This was done using the 2l program in BMDP[1]. Out of the original 154 patients 143 were at risk of developing a metastases before death. Of the eleven patients who were omitted for this purpose, 1 had a missing value, and 10 had not had sufficient follow-up to develop a metastases. Table 2.3 shows the results of fitting models involving metastases whilst Figure 2.2 and Figure 2.3 shows the dynamic evolution of metastatic disease for the 143 patients at risk of it for photon patients and neutron patients separately. These diagrams are a type of Lexis diagram (Lexis, 1875) and have been popularised by Keiding (1990). From Figure 2.2 and Figure 2.3 we can see that neutron patients would appear to develop metastases more often than photon patients, but also at an earlier point in the trial.

From Table 2.3 we can see that the hazard ratio of death for neutrons compared to photons is 1.73 with approximate 95% confidence interval not including one, indicating evidence for a detrimental effect compared to photons. Including an indicator variable for metastases as a time-independent variable slightly reduces the treatment effect. As we might expect the effect of metastases retrospectively on survival is to indicate that those patients who were known to have developed metastases had increased risk of dying. This analysis does not allow for the fact that those patients who did develop metastases did so at different points in their disease history, and in order to allow for this we could fit metastases as a time-dependent covariate. This has the effect of slightly reducing the treatment effect, but the main point to notice is that once a patient has developed metastases they have over four times the risk of dying than those patients who have not at that time point, regardless of treatment. There appears to be little evidence for the existence of a treatment-metastases interaction.

2.2.5 Clinically relevant questions

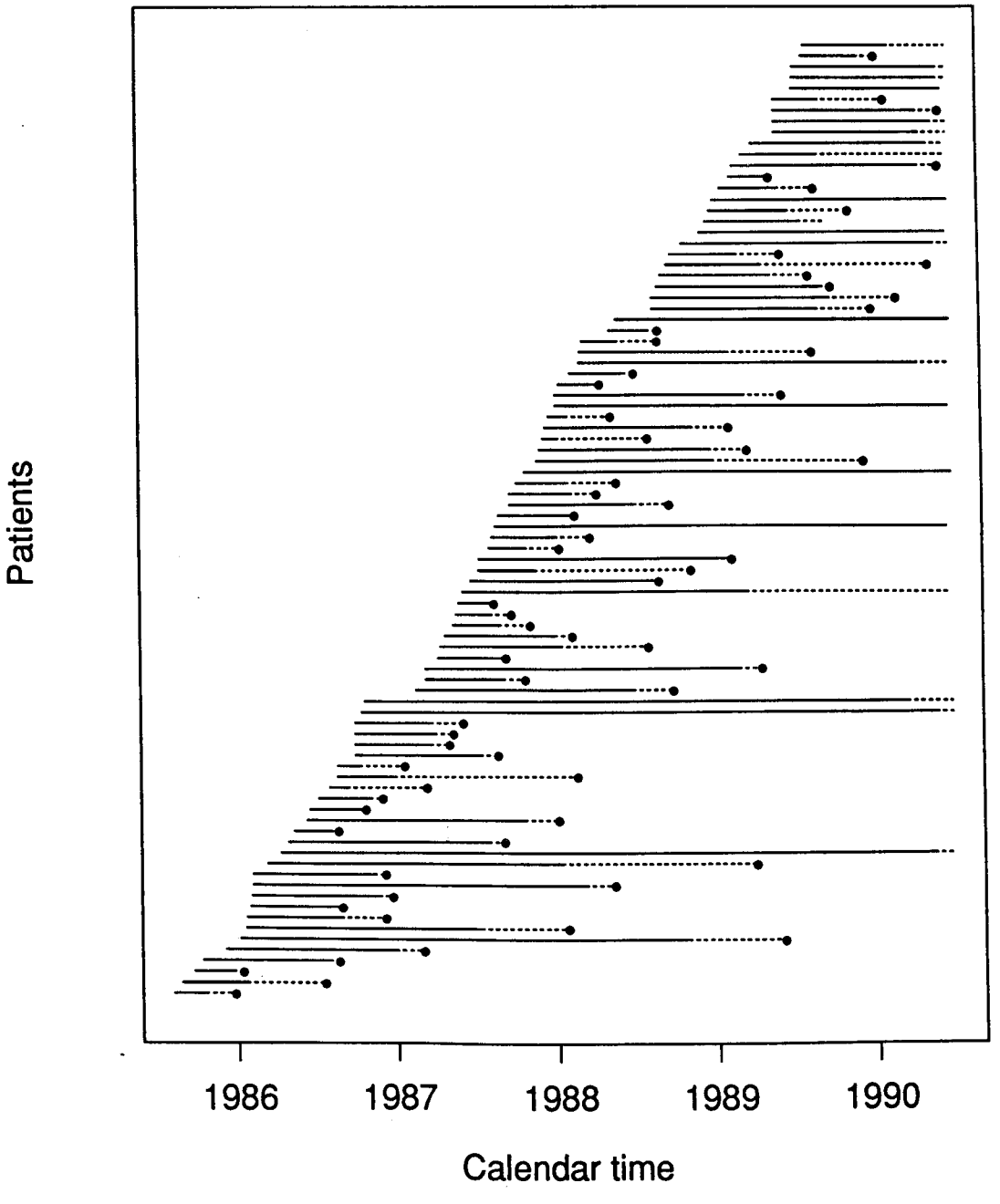
There are a number of clinical questions that need to be answered, which though posed with respect to this particular trial are common to cancer clinical trials

[1]BMDP is a registered trade mark of BMDP Statistical Software Inc.



Key : ——— Non-metastatic, ····· Metastatic, • Death.

Figure 2.2: Dynamic evolution of metastatic spread for photon patients, ordered by date of entry into trial and with 21st December 1990 as censoring date.



Key : ——— Non-metastatic, ····· Metastatic, • Death.

Figure 2.3: Dynamic evolution of metastatic spread for neutron patients, ordered by date of entry into trial and with 21st December 1990 as censoring date.

Covariate	β^a	SE[b]	Hazard ratio	95% CI[c]
Treatment	0.549	0.218	1.732	(1.120,2.678)
Treatment	0.481	0.221	1.618	(1.040,2.517)
Metastases	0.472	0.216	1.603	(1.041,2.469)
(Not time-dependent)				
Treatment	0.341	0.224	1.406	(0.899,2.201)
Metastases	1.430	0.221	4.177	(2.686,6.501)
(Time-dependent)				
Treatment	0.451	0.261	1.569	(0.931,2.646)
Metastases	1.485	0.225	4.413	(2.815,6.924)
(Time-dependent)				
Treatment × Metastases	-0.411	0.487	0.663	(0.250,1.756)

[a] β denotes log hazard ratio.
[b] SE denotes standard error for β.
[c] CI denotes confidence interval for the hazard ratio.

Table 2.3: Model results for hazard ratio of death (neutrons to photons) for 143 patients at risk of developing metastases, using censoring date of 21st December 1990.

generally.

The questions are;

- What is the current strength of evidence for the use of neutron therapy for tumours of the pelvic region?
- Is the rate of development of metastases different in the two treatment groups?
- How does the development of metastases affect subsequent survival? Is this the same for both treatment groups?
- How does the *time* at which metastases develop affect subsequent survival?
- What is the current strength of evidence of metastases being a major contributory factor to the differences in survival?
- How do covariates, other than treatment, affect survival?

We hope to show that all these questions may be answered within a coherent framework by adopting the Bayesian methods developed in subsequent chapters.

2.3 Quality of Life Study

2.3.1 Introduction

A pilot quality of life study was conducted at Clatterbridge hospital, Wirral, between early 1988 and mid 1989. A total of 35 patients were recruited to the study. These patients were the first 35 patients presenting for the first time at the Mersey Regional Radiotherapy Centre over the study period, and who were being treated for cancer of the testis, ovary, lung or skin (melanoma). All 35 patients were interviewed by the same research sister when they presented in clinic for follow-up visits.

The questionnaire used had two sections, a general quality of life section and a site/treatment specific section, being based around the European Organisation for Treatment of Cancer (EORTC) modular quality of life questionnaire (Aaronson *et al.*, 1988). Due to the relatively small number of patients recruited to the study only the general section is used in this analysis. This section of the questionnaire is replicated in Appendix B. The data was recorded using COMputer PAckage for Cancer Trials (COMPACT) (Chilvers *et al.*, 1988).

Apart from assessing the reliability and ease of collecting such data, there was clinical interest in whether different groups of patients could be identified as having different qualities of life at different times in their disease/treatment process. To assess this latter point the 12 lung cancer patients were compared with the remaining 23 patients whose tumours were of ovary, testis or skin.

In order to compare these two groups a global score was formed. The global score is the sum of the ordered responses to the 15 questions on the questionnaire. For each question a patient contributes a 0, 1, 2 or 3 to their overall score, the lower the number the better their quality of life on that particular question. Therefore the maximum score attainable was 45 and the minimum was 0.

2.3.2 Results

Figure 2.4 shows Kaplan-Meier survival curves for lung and non-lung cancer patients, using time to death, if it occurred within 30 days of last follow-up. Using Cox's proportional hazards model the hazard ratio for death of lung cancer patients compared to non-lung cancer patients is 1.56 with approximate 95% confidence interval (0.42,5.81). The wide confidence interval is because there are only 9 deaths, 4 lung cancer patients and 5 non-lung. If we just consider survival, and use all available data, i.e including times of death that are beyond the period of time when quality of life is measured, the hazard ratio of death increases to 6.83 with approximate 95% confidence interval (2.47,18.9).

Figure 2.5 shows the distribution of the quality of life scores for lung and non-lung cancer patients in the study. We can see that lung cancer patients distribution is more skew towards a higher score indicating lower quality of life.

From Table 2.4 we can see the numbers of transitions between states defined

		'good'	'medium'	'poor'	death
'good'	non-lung	8	5	0	0
	lung	1	4	0	0
'medium'	non-lung	2	8	3	2
	lung	3	5	4	3
'poor'	non-lung	0	0	1	2
	lung	0	1	3	0

Table 2.4: Transitional status for lung and non-lung cancer patients in pilot quality of life study.

using the quality of life score. These were defined as 'good' for scores less than 20, 'medium' for 20 to 30, and 'poor' for above 30. We can see from Table 2.4 that for some of the possible transitions there are very few or even zero transitions. This limits the level of sophistication of the models that we are able to fit to the data. Figure 2.6 shows the dynamic evolution of patients through the various quality of life states separately for the lung and non-lung cancer patients. From Figure 2.6 we can see that it would appear that lung cancer patients spend more time in 'medium' and 'poor' quality of life states than non-lung cancer patients.

2.3.3 Clinically relevant questions

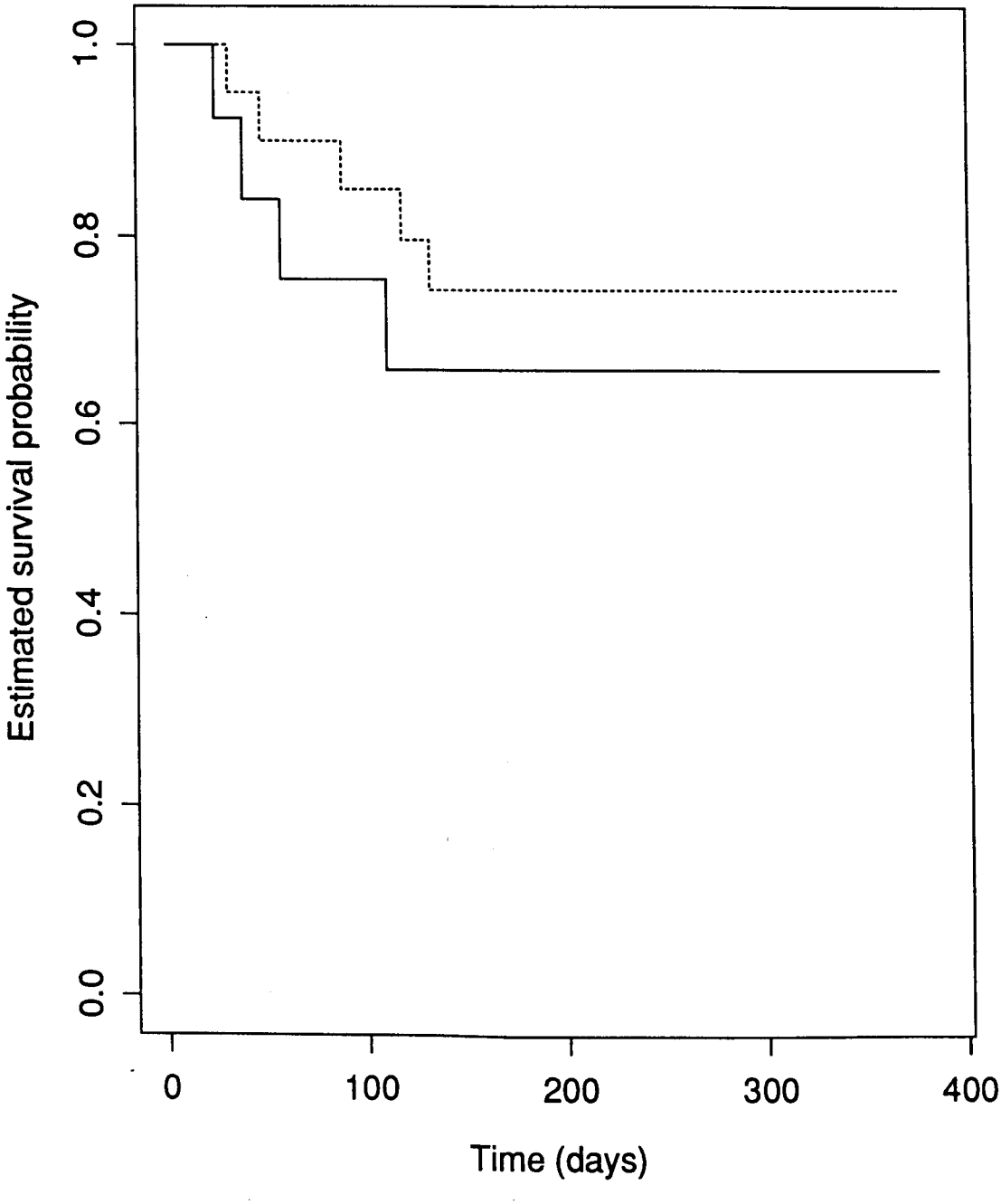
As with the neutron therapy trial there are a number of clinically relevant questions that need to be answered.

- Is the difference between lung and non-lung cancer patients in terms of survival also there in terms of quality of life?
- Are any differences between the two groups in the risk of deterioration/improvemen in quality of life constant?
- Is there changing risks of death with varying quality of life?
- Is there a different risk of death for the two groups at different levels of quality of life?

In this situation the questions of interest surround the comparison of the quality of life of the two patient groups over time. In order to do this multi-state models are used, which using a Bayesian approach to estimation, lead to clearly interpretable graphical summaries of the comparative risks of transitions between various quality of life states, including death.

2.4 Summary

In both examples there are aspects other than survival that need to be analysed. In the clinical trial we have morbidity events that may occur during the course



Key : ——— Lung, ····· Non-Lung

Figure 2.4: Kaplan-Meier estimated survival curves for Lung cancer and Non-Lung cancer patients in pilot quality of life study.

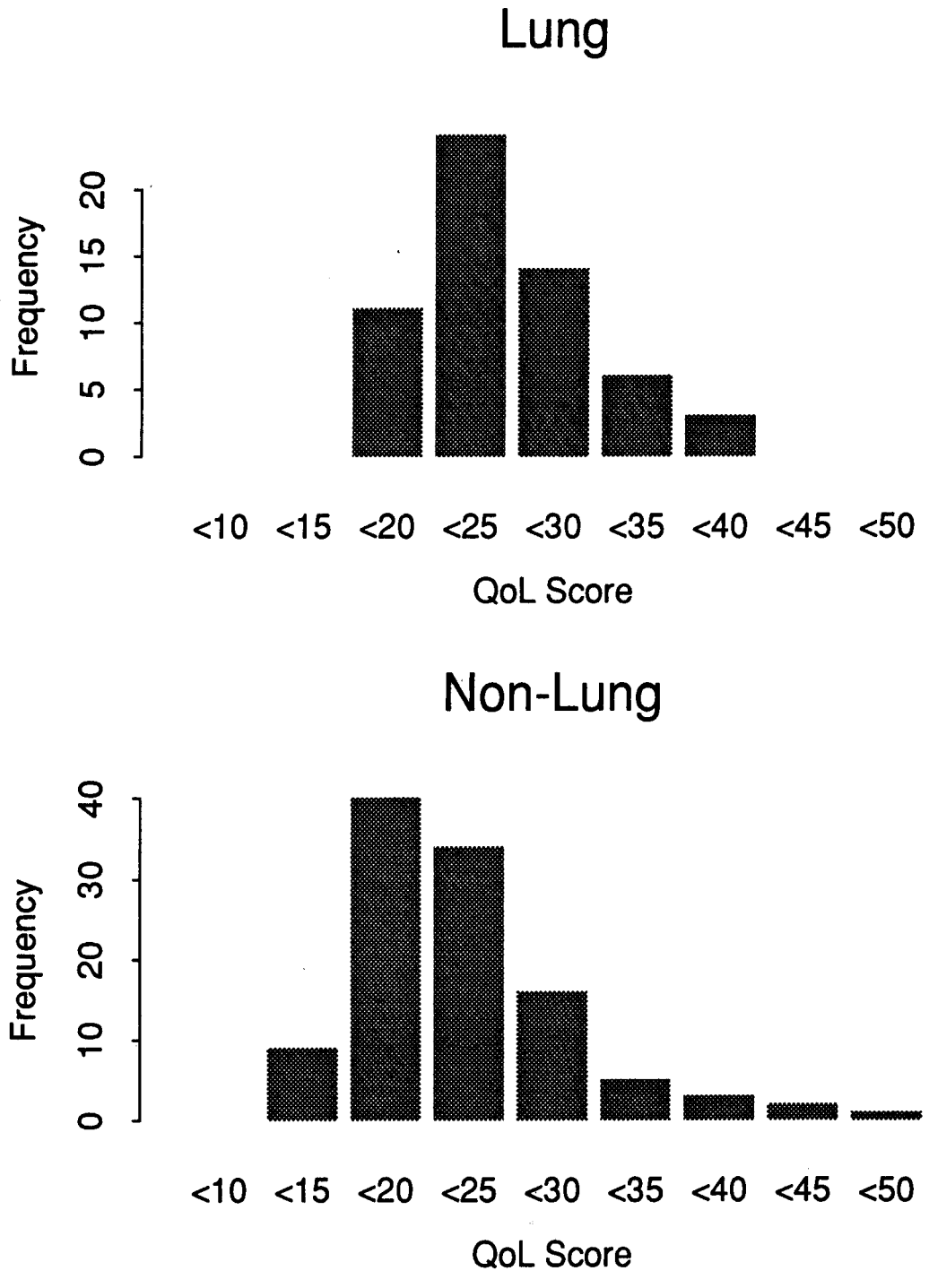


Figure 2.5: Histograms of global quality of life index for Lung cancer and Non-Lung cancer patients in pilot quality of life study.

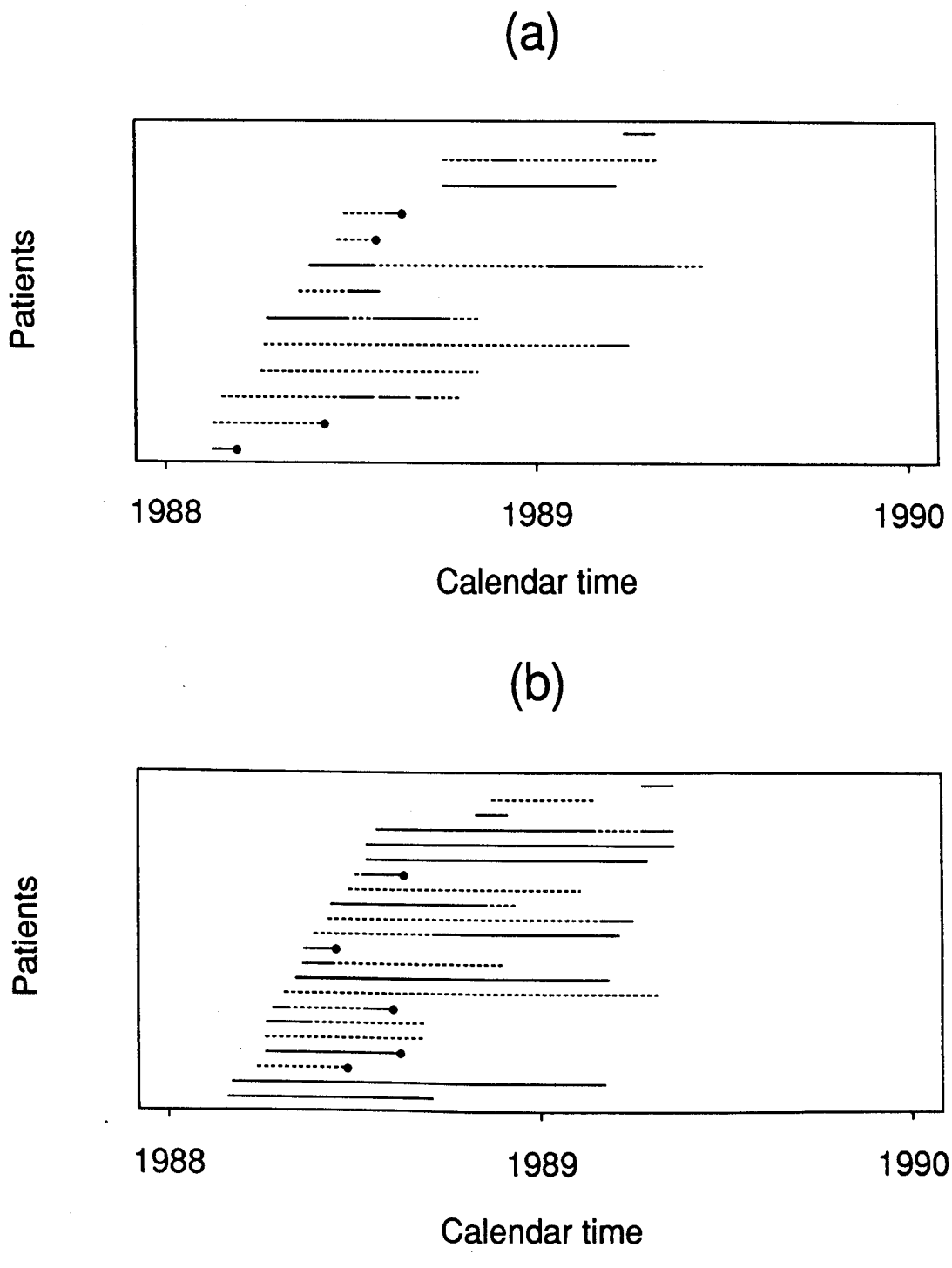


Figure 2.6: Dynamic evolution of global quality of life score for (a) lung cancer and (b) non-lung cancer patients in pilot quality of life study.

of a patient's treatment and subsequent recovery. In the quality of life study patients may be seen to move between a finite number of quality of life states. In both examples, the use of multi-state models is advocated as being desirable.

For the clinical trial there is substantial prior information both in the form of a statistical overview of six previous trials, and in the form of quantified clinical prior beliefs of ten clinicians. This aspect of clinical trials and observational studies is dealt with in more detail in Chapter 3, and in particular the existence of prior information for the neutron therapy study. In order to make use of this prior information, regardless of statistical models used we need to adopt Bayesian methodology.

Hence these examples between them strongly suggest the use of Bayesian survival models, or models that may be used for survival situations, and that may also allow for movement between disease states.

Chapter 3

Quantifying Prior Information

3.1 Introduction

In this chapter we consider the possible sources of prior information and the ways in which it may be quantified so that it can be included formally in a statistical analysis. This will be illustrated with prior information available for the Medical Research Council neutron therapy trial described in Chapter 2.

The general process is evident in clinical trials (Pocock, 1983). Initially a drug may be developed based on pharmacological grounds, and tested using animals. Afterwards the drug is tested on healthy volunteers to assess allergic reactions, a small pilot study is then conducted, and only after this is a 'large' Phase III trial conducted. At each step in this process information about the possible benefits of the drug are accrued. This argument that clinical trials, especially Phase III trials, should be seen as just another piece of evidence about the possible benefits has been advocated by both Peto (1987) and Spiegelhalter *et al.* Peto has also argued, as have others including Bulpitt (1988) and Thompson and Pocock (1991), that as many Phase III trials are relatively small, the true picture of the possible benefits of a drug or treatment can only be achieved by considering all relevant phase III trials that have been conducted in a formal statistical overview, or what is also termed a *meta analysis.* We shall return to the idea of meta analyses in more detail later in this chapter.

Two of the possible sources of prior information, are clinical beliefs about the current study, and the results of previous studies that are considered relevant to the current one. We may expect these two sources of prior information to yield similar results since clinicians should be aware of the previous studies that have been conducted in their area of research. However, their prior beliefs about the current study will not only be based on the results of the previous studies but also on their personal experience.

Section 3.2 reviews the elicitation and modelling of clinical prior beliefs, with particular attention to the MRC neutron therapy trial, in which clinical prior beliefs were obtained.

Section 3.3 first considers the possible sources of data that may be used to construct prior densities. There are essentially two sources; previous trial results, and routine data. We consider the use of previous study results, how they can be combined and how these combined results may be used to obtain a density for a quantity of interest. This process is described in detail for the MRC neutron therapy trial for which previous trial results were available. A number of authors have developed Bayesian methods for meta analyses, and we review some of these approaches. Section 3.3 also reviews the use of historical controls in clinical trials. Finally, Section 3.4 summarises the chapter.

3.2 Clinical Prior Beliefs

3.2.1 General

Several authors including Savage (1971), Smith (1985) and O'Hagan (1988) have considered various methods of eliciting prior information from individuals. However, there are few examples detailed in medical statistics literature. Three exceptions are, are Freedman and Spiegelhalter (1983), the results of a collaborative multi centre study on the use of artificial surfactant for premature babies (Lloyd *et al.*, 1987) and Chaloner *et al* (1992).

Freedman and Spiegelhalter (1983) consider 18 clinicians involved in a Medical Research Council Study for the treatment of superficial bladder cancer. They were comparing surgery, the standard treatment, to a new drug treatment, thiotepa. The prior beliefs were elicited by 'face to face' interviews. In order to elicit beliefs about *clinical demands* each clinician was asked 'If the real benefit was $x\%$ would you use thiotepa as your routine treatment?', with real benefit taken to mean the results of a very large impeccably conducted trial. By varying x they were able to establish upper and lower limits for routine use of thiotepa for each clinician. Between these limits the clinician would consider the two treatments equivalent, and outside this range would use one or other of the treatments routinely.

Clinical beliefs were elicited in a similar manner, with each clinician being asked first for the most likely improvement to be gained from thiotepa. They were then asked for upper and lower limits outside which they thought the difference was very unlikely to lie. They were then asked for the chance of the difference exceeding a selection of intermediate points between these limits. Pooling all this information prior distributions for the individual clinician's beliefs about the treatment difference were drawn. Freedman and Spiegelhalter comment on the diversity of not only the location of the prior distributions but also their shape, raising doubts on the validity of combining them to form a consensus prior.

Chaloner *et al* (1992) describe the use of graphical elicitation of prior information in a trial of prophylactics for tomoplasmosis in a population of HIV positive people. This is based on the `XLISP-STAT` environment (Tierney, 1990).

Savage (1971), Smith (1985) Chapter 4 and O'Hagan (1988) Chapter 2. have commented on the elicitation of prior beliefs, often though from a more mathematical point of view. Smith (1985) suggests that in a medical context the most promising method of eliciting subjective probabilities is via a relative frequency approach, which has the advantage of being easier to identify with, and also avoids the hypothetical betting structure, e.g 'If 100 patients were given treatment A how many do you think would react in x hours?'. O' Hagan (1988) considers elicitation from a betting odds point of view, e.g 'Would you use treatment A rather than treatment B if the reduction in mortality was $x\%$, but that there was only $y\%$ chance of this occurring?'.

In clinical trials the question of interest is often straight forward and elicita-

tion via question and answer sessions appears to be the most appropriate. An important consideration though is whether in a two treatment comparative trial to elicit beliefs about the 'failure' rates in the two groups separately or whether to elicit beliefs about a relative measure of efficacy.

3.2.2 MRC Neutron Therapy Trial

In March 1988 it was decided to elicit the beliefs clinicians involved in the Medical Research Council Neutron Therapy Study. This was done primarily to establish the ethical nature of the randomisation of further patients into what was considered by some to be a controversial trial.

In order to elicit these beliefs a 'trial roulette' was conducted by Dr Sheila Gore. This method had been previously used in a number of clinical trials, Gore (1987) and Lloyd *et al* (1987). Ten clinicians were willing to participate in the present exercise. The trial roulette was conducted by post, and used the form in Figure 3.1. The clinicians were asked to express their beliefs about the effectiveness of neutron therapy compared to a failure rate of 50% for photon therapy. Failure was defined to be

- death without having achieved local control of symptoms
- recurrence after local control
- grade 4 or 5 radiation morbidity (using EORTC grading system)
- death due to metastatic disease

The clinicians were allowed twenty counters for use with the betting streets on the form. They first had to place one counter at the lower end of their beliefs, and another one at the upper end of their beliefs. The remaining 18 counters were then place so as represent their belief distribution between these two points.

The form asked the clinicians for their beliefs about the failure rate on neutrons compared to a 50% failure rate on photons, and they were also asked what the failure rate on neutrons would have to be before they would use them in routine clinical practice. The average failure rate demanded by the ten clinicians for neutrons was 38.5% compared to a 50% failure rate on photons.

Figure 3.2 shows the individual beliefs for the ten clinicians. All but one of the clinicians placed the majority of their counters below a failure rate of 50%, indicating that they thought that neutrons would have a beneficial effect on failure. One clinician, (f), acted quite differently, and placed the bulk of his counters above 50% indicating that he thought in terms of failure, that photons would be more beneficial that neutrons. He did however place one counter below 50%, indicating that he believed that there was a 5% chance that there would be a small improvement for neutrons.

In order for us to be able to use these elicited clinical beliefs in any type of statistical model, we need to determine a prior distribution or density function.

We also need to consider whether to combine them in some way, or to treat them individually. If we decide to combine them, then we could combine the raw data, and then estimate a density. Alternatively we can estimate a prior density for each clinician, and then combine the prior densities.

Genest and Zidek (1986) consider the combination of prior densities for a number of individuals, according to both a linear opinion pool and a logarithmic opinion pool. Under a linear opinion pool a weighted linear combination of the individual prior densities is taken as a consensus prior density. Problems arise in assigning the weights to the individual densities and also in the fact that if there are diverse prior beliefs then the consensus prior density can be multi-modal and quite dispersed. A possible solution to both of these problems is to use logarithmic pooling in which the consensus prior density is a weighted product of the individual densities. Although more likely to be uni-modal and less dispersed than under linear pooling, the choice of weights still remains.

Note that under linear pooling if we assume fixed weights, the posterior mixture distribution is the same as if we updated each individual's prior separately and then combined the posterior densities. Another important point is the validity of pooling when there are serious discrepancies between individuals. Unless a consensus prior density is reached by an iterative process such the *delphi* technique, then there appears to be more weight to the argument that the prior densities should be considered separately, and the resulting posterior densities compared.

Apart from clinician (f) there is considerable agreement that neutrons should be beneficial. One simple way would be to aggregate all the data to form a consensus distribution of beliefs. This is in fact what was done for the purposes of analyses in subsequent chapters. Figure 3.3 shows the aggregated beliefs. We see that the clinician who had different beliefs to the majority has been swamped, and that the consensus belief is that there is a beneficial effect of neutrons. Another option would have been to combine the individual beliefs according to a weighting scheme. However, as in the case of combining prior densities this would introduce the problem of choosing the weights.

Assume that we wish to work with the histogram of aggregate beliefs in Figure 3.3. In estimating densities for quantities of interest there are two possible approaches; to estimate the failure rates for the two treatment groups separately or to estimate a density for some measure of relative efficacy, e.g relative risk of failure. We shall see in subsequent chapters that depending on which statistical model we use, we may require the prior information in either of these forms.

First consider the case when we wish to estimate densities separately. For the neutron therapy patients we wish to summarise the histogram in Figure 3.3. Simply taking the average of the histogram yields 0.464, and a variance of 0.018. These are relative to a mean failure rate of 0.5 for photons. In the odds model that is developed in Chapter 4 and also reported in Abrams, Ashby and Errington (Submitted) we require the prior density for the failure rate in each treatment group to have a Beta density as a prior density. We can use the 'Method of Moments' described in Maritz and Lwin (1989) to do this.

Assume that we are interested in the failure rate of neutrons, θ_n, and this has a Beta distribution with parameters α_n and β_n. For this to be a proper density $\alpha_n > 0$ and $\beta_n > 0$. The expectation of θ_n depends on the ratio α_n/β_n. If this ratio remains constant but α_n and β_n increase, the variance decreases and the distribution tends to the Normal distribution, (Rothschild and Logothetis, 1985). Thus the mean of θ_n, denoted by $E(\theta_n)$, is given by

$$E(\theta_n) = \frac{\alpha_n + 1}{\alpha_n + \beta_n + 2}$$

and the variance of θ_n, denoted $V(\theta_n)$, is given by

$$V(\theta_n) = \frac{(\alpha_n + 1)(\beta_n + 1)}{(\alpha_n + \beta_n + 2)^2(\alpha_n + \beta_n + 3)}$$

Since we can elicit beliefs about the mean and variance of θ_n, we can find the corresponding estimates of α_n and β_n from

$$\begin{aligned} \alpha_n &= \frac{-E(\theta_n)^3 + E(\theta_n)^2 - E(\theta_n)V(\theta_n) - V(\theta_n)}{V(\theta_n)} \\ \beta_n &= \frac{-2E(\theta_n)^2 + E(\theta_n) - 2V(\theta_n) + E(\theta_n)^3 + E(\theta_n)V(\theta_n)}{V(\theta_n)} \end{aligned} \tag{3.1}$$

Exactly analogous expressions yield α_p and β_p parameters of a Beta prior distribution for θ_p, the failure rate for photon therapy.

Thus, using $E(\theta_n) = 0.464$ and $V(\theta_n) = 0.018$, we can obtain estimates for α_n and β_n. These are $\alpha_n = 4.78$ and $\beta_n = 5.68$. Figure 3.4 shows the histogram of aggregate beliefs superimposed this Beta density. Although the mode of the density reflects the location of the consensus prior beliefs accurately, clinician (f) has caused the density to give more weight to higher neutron failure rates. For the purposes of analysis we assume that for the photon treatment, the mean of the density is 0.5, and the variance is *assumed* to be 0.005. The variance is assumed to be 0.005, as photons are the standard treatment and their effect is known more accurately. The choice of a variance of 0.005, a quarter of that for neutrons, serves to place emphasis on the effect of the new treatment. Analogously α_p and β_p are both found to be 23.5.

For the purposes of the statistical survival models developed in Chapter 6 we require a prior density for the log-relative risk of failure, for neutrons compared to photons. The histogram in Figure 3.3 can be thought as relating to relative risk as in specifying the failure rate on neutrons the clinicians have been asked to assume that the failure rate on photons is 50%. Therefore we may calculate the mean relative risk as 0.928, and the variance as 0.0735.

If we think of a survival model in which e^{β}, the hazard ratio, is assumed to be an instantaneous relative risk, then we may be interested in specifying a prior for β, the log relative risk. Since β is a regression coefficient and can therefore take both negative and positive values, a Normal distribution would

seem a sensible choice as a prior distribution. Therefore, if we assume a Normal distribution for β, e^{β} has a Log-Normal distribution. The relationship between the Normal and Log-Normal means that if β has a Normal distribution with mean μ and variance σ^2, then the mean of e^{β} is $e^{\mu+0.5\sigma^2}$, and the variance of e^{β} is $e^{2\mu}\left(e^{2\sigma^2}-e^{\sigma^2}\right)$, see Johnson and Kotz (1970) pages 115-117. Solving

$$\begin{aligned} e^{\mu+0.5\sigma^2} &= 0.928 \\ e^{2\mu}\left(e^{2\sigma^2}-e^{\sigma^2}\right) &= 0.0735 \end{aligned}$$

yields a mean of β of -0.116 and a variance of 0.082. In Chapter 6 we shall use this prior distribution in a survival model. The corresponding Log-Normal density for the relative risk is superimposed on the histogram of aggregate prior beliefs in Figure 3.4 (b).

An alternative method for specifying a prior for β would have been to work with the hazard ratio directly. Assuming that the hazards in each of the two treatment groups were constant, we can obtain an expression for the hazard ratio in terms of the 12 month failure rates. The effect of using such a method for the neutron therapy study, would have been to obtain a prior density that had 'heavier tails' than that shown in Figure 3.4 (b). We will return to deriving a prior density for the hazard ratio directly in Section 6.5.3.

As with estimating the Beta prior density for the failure rate on neutron therapy, the prior density seems an adequate approximation to the histogram. Part of the 'lack of fit' is due to clinician (f) who believed *a priori* that neutrons were not likely to be beneficial, and therefore considerably affected the mean and variance. This problem could be overcome if we used a mixture distribution or a kernal based density estimator (Silverman, 1986). However, there is a trade-off between gaining a 'good' representation of the histogram and having a density that is more mathematically convenient.

An important point to notice is that by combining the individual beliefs is that we may obtain a distribution of consensus beliefs that underestimates the variability amongst the individuals.

Both the Beta prior densities and the Normal prior densities derived above will be referred to as *clinical prior* densities in subsequent chapters.

3.3 Data Based Priors

3.3.1 Sources of Data

There are essentially two sources of data which may be used to construct prior densities; previously published trials and routine sources of data such as cancer registries and specialist databases.

In recent years there has been a growing interest in the collection and possible combination of previous trial results, in what are termed *meta analyses* or statistical overviews, (Peto, 1987, Thompson and Pocock, 1991). In many ways this reflects the belief that clinical trials should not be seen in isolation, but rather

Neutron Therapy Pelvic Irradiation : Trial Roulette

Prepared by Dr Sheila Gore

Tumour locally advanced viz. T3,T4; N1—NX; M0
Sites : rectum, cervix, prostate, bladder

The diagram shows betting 'streets' (as on a gaming table) and identifies the 'street' (centred on 50% treatment failure rate by 12 months) which represents photon therapy failures.

By treatment failure, we mean :-

death without having achieved local control of symptoms
recurrence after local control
grade 4 or 5 radiation morbidity
death due to metastatic disease

How effective do **YOU** think neutron therapy might be in terms of reducing the 12-month failure rate of patients with pelvic cancer?

Please, place your 20 gaming tokens (X) in some or all of the 'streets' to represent your current belief and uncertainty (i.e bets) about treatment failure by 12 months of patients randomised to neutron therapy for pelvic cancer.

Betting 'Streets' for NEUTRON FAILURE RATE at 12 months

25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%

For reference : photon failure rate at 12 months = 50%

What would the 12-month failure rate need to be with neutrons before you would **advise neutron therapy routinely** (rather than photons) for pelvic cancer?

		%

Signature

Figure 3.1: 'Trial Roulette' Form, Gore, (1988).

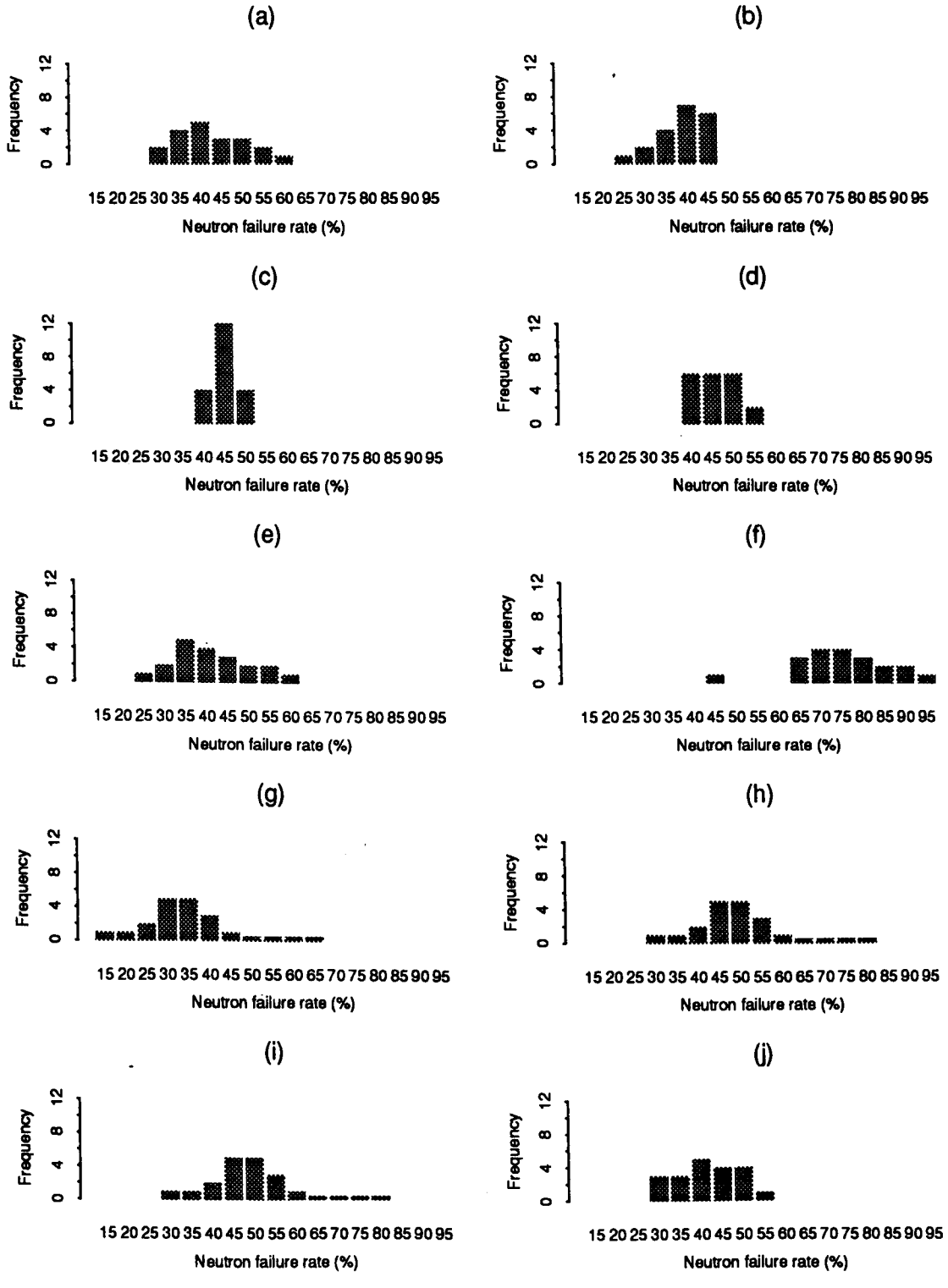


Figure 3.2: Individual elicited prior beliefs of ten clinicians for neutron therapy compared to a 50% failure rate on photon therapy.

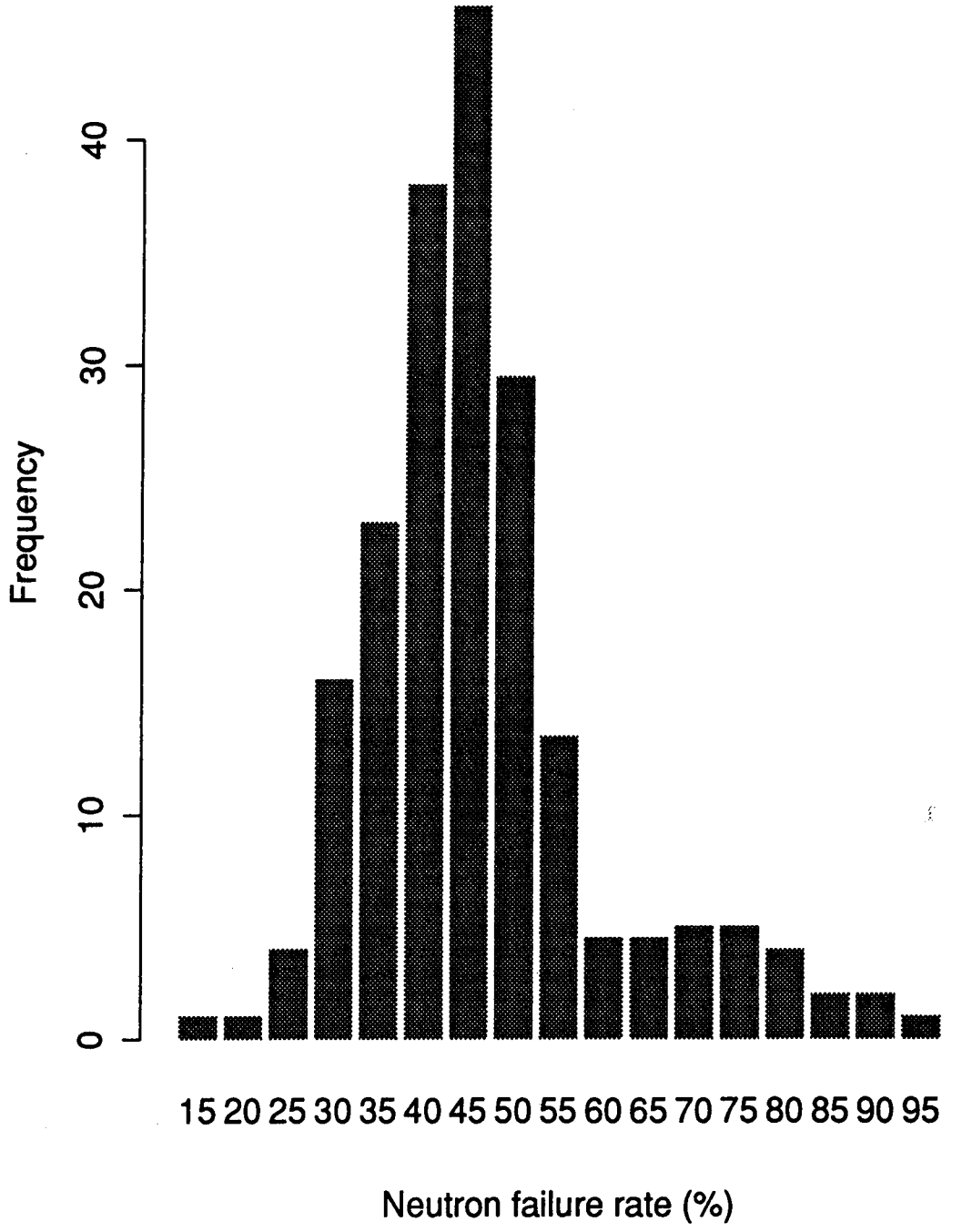


Figure 3.3: Elicited aggregate prior beliefs of ten clinicians for neutron therapy compared to a 50% failure rate on photon therapy.

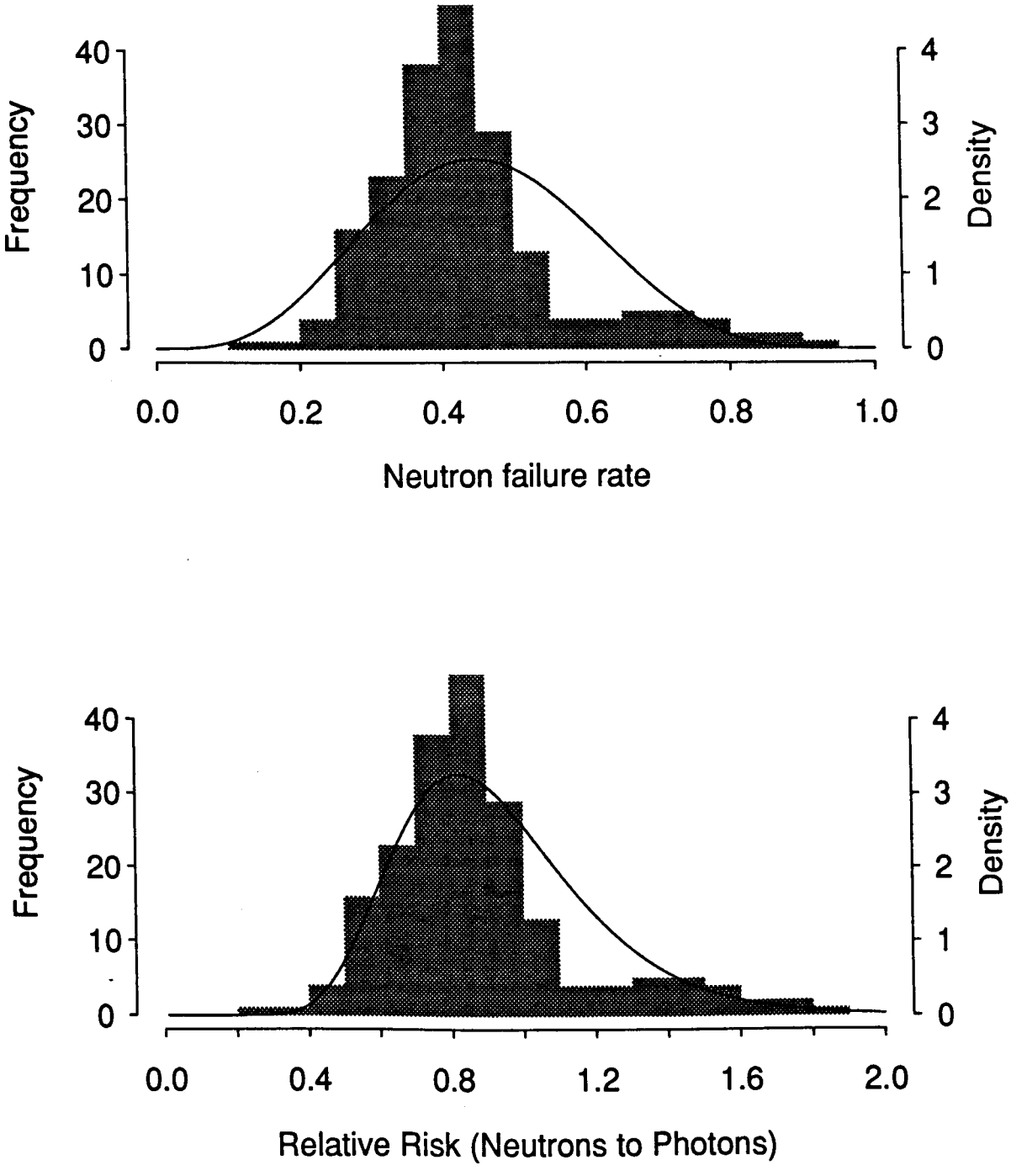


Figure 3.4: Elicited aggregate prior beliefs of ten clinicians for neutron therapy compared to a 50% failure rate on photon therapy, (a) Beta density, (b) Log-Normal density.

as adding to, or updating, our current beliefs about a particular treatment, (Spiegelhalter *et al.*, In press). Several authors have differentiated between what have been termed 'qualitative overviews' and 'quantitative overviews', (Bulpitt, 1988, Thompson and Pocock, 1991). The former relates to studies that look at the differences between trials, their protocols, patients accrual etc, as well as treatment differences, and on the basis of these makes statements about the effectiveness or otherwise of the treatment in question. 'Quantitative overviews' on the other hand aim at a statistical amalgamation of the results of studies that are deemed sufficiently alike, and on the basis of this make a quantitative assessment of treatment effect. Clearly, a 'qualitative overview' may precede a quantitative one, so as to assess *which* previous studies to include in a 'quantitative overview'.

One possible source of routine prior information, for cancer clinical trials, is from cancer registries. Cancer registries record details of the new cases of various cancers together with various patient characteristics, (Office of Population Censuses and Surveys (OPCS), 1990). They are also notified of patients that have died, via the Office of Population and Censuses (OPCS). The main problems with using this source of information are that the quality of the data that is recorded is variable, even within registries, and there is also variability in what data is actually collected between registries. There are other complications such as no information of clinical staging, using systems such as TNM (Harmer, 1978), which has severe implications for prognosis.

3.3.2 MRC Neutron Therapy Trial

In the previous section we outlined the rationale for conducting meta analyses. Here we consider the mechanics of quantitative overviews, and describe the meta analysis that was performed in the neutron therapy trial of Chapter 2, which was reported in Errington *et al* (1991).

There are essentially two methods that have been advocated for statistical overviews (meta analyses). The first suggested by Mantel and Haenszel (1959) assumes that there are no differences in the underlying true treatment effects for the individual trials. This assumption is sometimes known as the *fixed effects* assumption. Prior to combining a group of trials in such a manner there are a number of significance tests that may be used to assess the strength of trial heterogeneity. Two such tests are Cochran's Test and Woolf's Test. See Armitage and Berry (1987) for further details. If either a test for heterogeneity fails or graphical methods indicate that the true effects in the individual trials are different then a second approach that allows for such eventualities is applicable, this approach is known as a *random effects* model. We allow for such heterogeneity by considering the trials to have different *true* treatment effects. One such *random effects* model has been proposed by DerSimonian and Laird (1986).

Errington *et al* (1991) describe a classical meta analysis of trials for low energy neutrons. The results of this meta analysis are summarised in Table 3.2

and are displayed graphically in Figure 3.5 for the odds ratio and Figure 3.6 for the relative risk. Figures 3.5 and 3.6 indicate that, with the exception of the unpublished study, there is little evidence for heterogeneity between the studies. Applying Woolf's Test for homogeneity yields a test statistic of 6.7, which under the null hypothesis of homogeneity has a chi squared distribution with 5 degrees of freedom, also indicating virtually no evidence for heterogeneity. Hence using a fixed effects model based on the Mantel-Haenszel estimate of the overall odds ratio is appropriate.

	Dead	Alive	Total
Neutrons	a_i	b_i	$a_i + b_i$
Photons	c_i	d_i	$c_i + d_i$
Total	$a_i + c_i$	$b_i + d_i$	n_i

Table 3.1: A 2 × 2 table for an ith hypothetical clinical trial.

Consider a sequence of m 2×2 tables, each one of the same form as Table 3.1. The Mantel-Haenzel estimate of the odds ratio, $\mathrm{OR}_{\mathrm{M-H}}$, is defined,

$$\mathrm{OR}_{\mathrm{M-H}} = \frac{\sum_{i=1}^{m} \frac{a_i d_i}{n_i}}{\sum_{i=1}^{m} \frac{b_i c_i}{n_i}}$$

Using the information in Table 3.2 we can calculate the Mantel-Haenzel estimate of the pooled odds ratio, which is 2.13. A method for calculating the variance of the Mantel-Haenszel estimate of the odds ratio is described in Armitage and Berry (1987). Define the quantities P_i, Q_i, R_i and S_i as follows

$$P_i = \frac{a_i+d_i}{n_i} \quad Q_i = \frac{b_i+c_i}{n_i}$$

$$R_i = \frac{a_i d_i}{n_i} \quad S_i = \frac{b_i c_i}{n_i}$$

The variance of the logarithm of the Mantel-Haenszel estimate is then given by

$$\frac{\sum P_i R_i}{2(\sum R_i)^2} + \frac{\sum (P_i S_i + Q_i R_i)}{2 \sum R_i \sum S_i} + \frac{\sum Q_i S_i}{2(\sum S_i)^2}$$

where summation is over the m trials. This method has the advantage over other methods which have been proposed in that it is suitable when there are a large number of strata and when some of these strata contain small frequencies. Applying this method to the data in Table 3.2 yields 0.044 for the variance of the log odds ratio, and therefore we can calculate an approximate 95% confidence interval for the odds ratio, which is (1.36,3.31).

We may wish to have an overall estimate of the relative risk rather than the odds ratio. Fleiss (1981) (page 69) describes how an estimate of the relative risk can be derived from the Mantel-Haenzel estimate of the odds ratio. If we

Trial	Site of tumour	Treatment	Outcome at one year Alive	Dead	Total
Batterman	Bladder, rectum	Neutrons	23	34	57
(1982)		Photons	16	18	34
Pointon	Bladder	Neutrons	39	16	55
(1985)		Photons	42	11	53
Duncan	Bladder	Neutrons	27	26	53
(1987)		Photons	43	17	60
Duncan	(inoperable) Rectum	Neutrons	3	17	20
(1987)		Photons	10	6	16
Duncan	(recurrent) Rectum	Neutrons	5	10	15
(1987)		Photons	11	5	16
Duncan	Bladder	Neutrons	2	4	6
(Unpublished, 1987)		Photons	1	3	4
		Total	222	167	389

Table 3.2: Meta analysis summary, Errington *et al* (1991).

assume that the marginal totals are fixed, and the Mantel-Haenzel estimate of the odds ratio is ω, then the expected frequencies of the main body of the table can be calculated. If A_i, B_i, C_i and D_i are the expected frequencies, then

$$A_i = \frac{X - Y}{2(\omega - 1)}$$

where

$$\begin{aligned} X &= \omega(2a_i + b_i + c_i) + (d_i - a_i) \\ Y &= \sqrt{X^2 - 4\omega(1-\omega)(a_i + b_i)(a_i + c_i)} \end{aligned}$$

having found A_i the other expected frequencies can be found by subtraction. Applying this method to the meta analysis, the table of expected frequencies, and fixed marginal totals is as in Table 3.3. and the relative risk is $\frac{106 \times 183}{61 \times 206} = 1.54$. Using the approximate 95% confidence interval that we have calculated for the odds ratio, we can apply the same technique that we have used on the point estimate on the upper and lower bounds of this interval to transform it to the relative risk scale. An approximate 95% confidence interval on the relative risk scale is (1.19,2.0). It may be that we require a prior density on the log relative risk scale. A relative risk of 1.54 corresponds to a log relative risk of 0.43 . An approximate 95% confidence interval for the log relative risk is $\log_e(1.19)$ to $\log_e(2.0)$, i.e 0.17 to 0.69 . Using a Normal prior distribution

	Dead	Alive	Total
Neutrons	106	100	206
Photons	61	122	183
Total	167	222	389

Table 3.3: Expected cell frequencies, assuming fixed marginals and $\omega_{M-H} = 2.13$ for neutron therapy meta analysis.

for the log relative risk assumes that this interval is symmetric about the mean, 0.43. Under this assumption we can calculate the standard deviation for the log relative risk, i.e 2 × standard deviation = 0.43 − 0.17. Therefore an estimate for the standard deviation is 0.13. The derivation of this Normal distribution of the log relative risk can be justified along Fiducial lines (Cox and Hinkley, 1974), i.e that beliefs about the relative risk can be translated into beliefs about the log relative risk, as the transformation is one-to-one.

In deriving a prior density for either the odds ratio or the relative risk based on the six studies, we have worked with the individual measures in each trial. The advantage of doing this is that we can incorporate variation within each trial into the overall estimate. This is not the case if we were to simply use the raw data in order to obtain an overall estimate of a quantity of interest. However, it might be the case that we require a prior density on the two treatment failure rates separately, and with the proviso that this ignores trial heterogeneity, we can calculate such a prior. From Table 3.2 we see that 61 out of 183 photon patients were dead at 12 months, i.e 32.8%, and 106 out of 206 neutron patients were dead at 12 months, i.e 51.9%. If we assume that the failure rate on photons is θ_p and that on neutrons is θ_n, then $E(\theta_p) = 0.328$ and $Var(\theta_p) = 0.328(1 - 0.328)/183 = 0.0012$, and $E(\theta_n) = 0.519$ and $Var(\theta_n) = 0.519(1 - 0.519)/206 = 0.0012$. Therefore, if as in the case of the clinical prior we wish to specify the prior distribution on each of the treatments for 12 month failure as Beta distributions we can use the 'Method of Moments' as in Section 3.2.2. The Beta distribution for photon patients has parameters 58.9 and 121.8, whilst the Beta distribution for neutron patients has parameters 106.4 and 98.6. Note that as both parameters of the two Beta distributions are relatively large this indicates that the two prior distributions are approximately symmetric.

Whilst we feel able to apply the results of trials for bladder and rectum patients to trials of other pelvic sites, when we use a relative comparison, we do not feel that we are justified in doing this when we consider the rates in the two treatment groups separately.

Both the Normal prior distribution and the two Beta prior distributions will be referred to as the *meta prior* in subsequent chapters.

An alternative method for deriving the hazard ratio directly from the results of a number studies is described by Peto, Pike, Armitage *et al* (1976). This method treats each study as a separate stratum, and then compares the ratio of the overall observed and expected number of deaths in the two treatment groups.

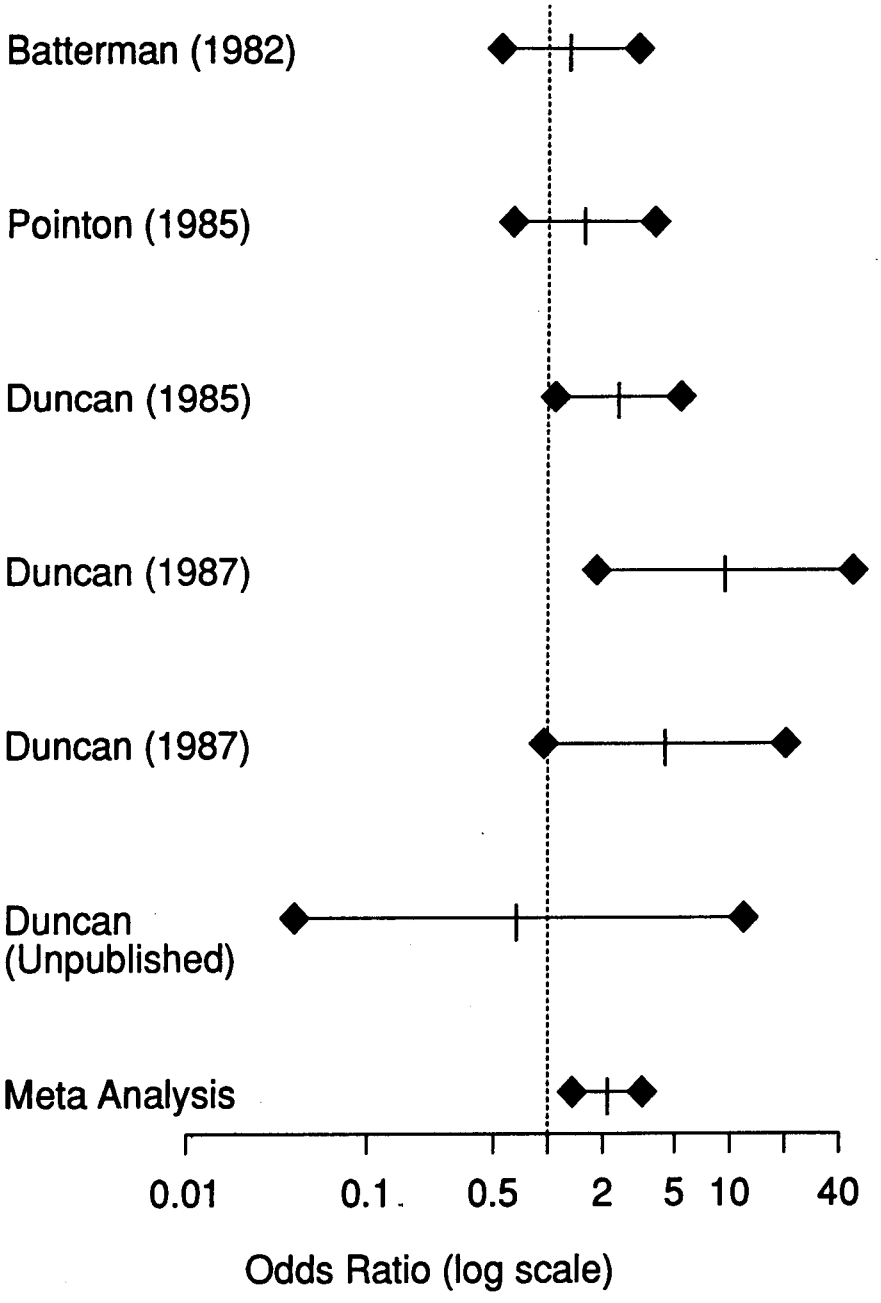


Figure 3.5: Statistical overview of six studies for low energy neutrons compared to photon therapy using odds ratio scale.

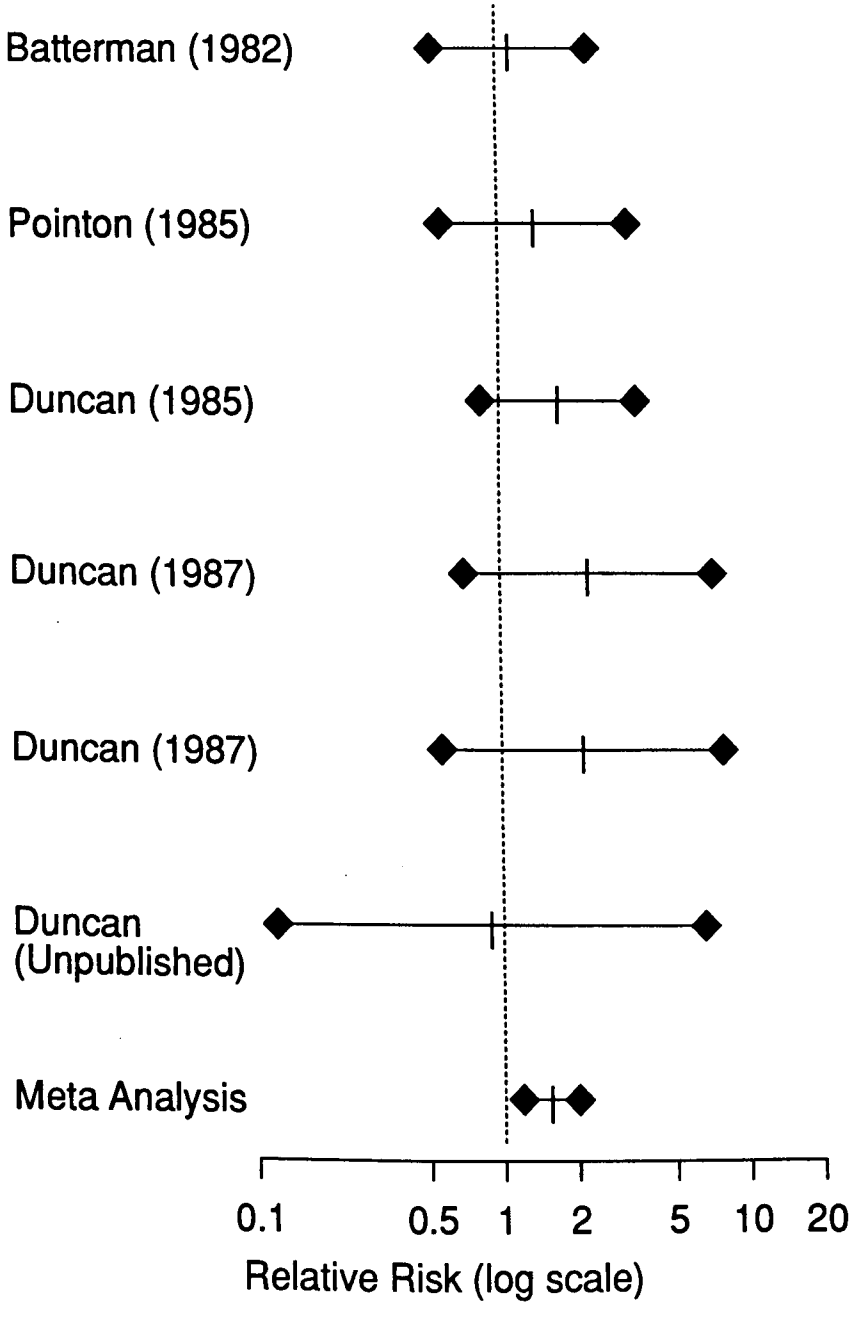


Figure 3.6: Statistical overview of six studies for low energy neutrons compared to photon therapy using relative risk scale.

In contrast to the method for meta analyses suggested by Mantel and Haenszel (1959), we describe briefly the random effects approach of DerSimonian and Laird (1986).

Assume that for the ith trial, the observed treatment effect is y_i, and that this observed quantity is made up of two components, θ_i the true treatment effect in the ith trial, and ϵ_i the sampling error in the ith trial. We assume that ϵ_i has zero mean, and variance σ_i^2.Thus,

$$y_i = \theta_i + \epsilon_i \qquad \epsilon_i \sim [0, \sigma_i^2]$$

The true treatment effect in the ith trial, θ_i, is made up of two components, μ, the population treatment effect, and δ_i, the ith study's deviation from the population treatment effect. δ_i has zero mean and variance σ^2. If σ^2 is zero then the random effects model reduces to that of the fixed effects model described above. Thus,

$$\theta_i = \mu + \delta_i \qquad \delta_i \sim [0, \sigma^2]$$

DerSimonian and Laird (1986) consider four different methods for estimation of the parameters of this model after replacing σ_i^2 by s_i^2, the observed variance in the ith study. Two of these methods yield a non-iterative procedure for estimation of the parameters, whilst the other two methods rely on the assumption of Normality for the distributions for ϵ_i and δ_i. Making this distributional assumption allows the use of either Maximum Likelihood or Restricted Maximum Likelihood. The latter method makes allowance for the fact that σ^2 is also often replaced with an estimate of the between study variance. Both of these methods can be implemented by using the EM algorithm.

3.3.3 Bayesian Meta Analyses

In this section we briefly review various approaches that have been suggested for the application of Bayesian methods in meta analyses. A number of these approaches develop the random effects model suggested by DerSimonian and Laird (1986) and are applications of Bayesian hierarchical models advocated by Raiffa and Schlaifer (1961). Other approaches to meta analysis consider the trials to generate binomial data.

None of the methods reviewed are applied to data, but it is felt that they are an extension of Bayesian methodology in clinical trials and should be included for completeness.

Consider a general Normal hierarchical model. Let $\underline{y}$ to be a vector of observed responses for k trials. These could be risk differences, relative risks or odds risks. Thus,

$$y_i|\theta_i, \sigma_i^2 \sim \mathrm{N}[\theta_i, \sigma_i^2] \tag{3.2}$$

and we assume that σ_i^2 is known or are prepared to put a fully specified distribution on it. At the second stage of the model

$$\theta_i|\mu, \sigma^2 \sim \mathrm{N}[\mu, \sigma^2] \tag{3.3}$$

here μ and σ^2 are unknown hyper-parameters, and the third stage of the model is to specify a prior distribution for μ and σ^2. Supposing for the moment that a Uniform prior density is assumed for μ and we know σ^2, closed form solutions exist for the posterior densities of all the quantities of interest, conditional on σ^2.

Carlin (1992) considers two possibilities for inference. One method requires substitution of an estimate for σ^2 into the closed form solutions for the required posterior densities. This is essentially an empirical Bayes procedure, (Maritz and Lwin, 1989). The alternative would be to take a fully Bayesian approach, and after specifying a prior density for σ^2, integrate the posterior densities for μ and the θ_is with respect to this density in order to obtain the posterior marginal densities for μ and the θ_is. This results in integrals that are not always analytically tractable, and Carlin adopts a Monte Carlo approach to obtain the resulting marginal posterior densities for μ and θ_i. Other methods such as asymptotic approximation (Tierney and Kadane, 1986) or Markov Chain Monte Carlo (Smith and Roberts, 1993) could also be used. Lindley (1971) and Box and Tiao (1967) show that with careful choice of distributions for σ_i^2 and σ^2 posterior distributions can be tractable.

Morris (1992) also considers a Normal hierarchical model for a meta analysis, and cites a deficiency in the method proposed by DerSimonian and Laird (1986). This is that they do not consider the variability of the population variance parameter, σ^2. The only way in which this can be rectified is if we adopt a fully Bayesian approach and obtain the marginal posterior density for σ^2. This density is of interest since if it is near zero nothing is gained by adopting a random effects model.

DuMouchel (1990) considers a more complicated model, of which those considered by Carlin (1992) and Morris (1992) are special cases. DuMouchel assumes $\underline{y} = (y_1, \cdots, y_i, \cdots, y_k)$ represents the responses from k studies, and that $\underline{y}$ has a multivariate normal distribution with mean vector $\underline{\theta}$ and covariance matrix $\tau^2 C$. Algebraically,

$$[\underline{y}|\underline{\theta}, \tau] \sim \mathrm{N}[\underline{\theta}, \tau^2 C] \tag{3.4}$$

where τ^{-2}/p_τ has a chi-square distribution with p_τ degrees of freedom. Comparing (3.4) with (3.2) above we can see that (3.2) is a special case of (3.4) with C diagonal and having elements $\sigma_i^2 \tau^2$. It is assumed that $\underline{\theta}$ also has a multivariate Normal distribution, independently of τ, with mean vector $X\underline{\mu} + \underline{d}$ and covariance matrix $\sigma^2 V$. Algebraically,

$$[\underline{\theta}|\underline{\mu}, \sigma] \sim \mathrm{N}[X\underline{\mu} + \underline{d}, \sigma^2 V] \tag{3.5}$$

where X and $\underline{d}$ are specified inaccordance with prior knowledge. Comparing (3.5) with (3.3) above we see that (3.3) is a special case of (3.5) with X and V identity matrices and $\underline{d}$ zero. Also assuming that $\underline{\mu}$ has a diffuse prior distribution such that,

$$[\underline{\mu}|\sigma] \sim \mathrm{N}[\underline{0}, D] \tag{3.6}$$

where D tends to infinity, and σ^{-2}/p_σ has a chi-squared distribution with p_σ degrees of freedom. As a result of the structure of this model, the posterior densities, or at least the first and second moments, can be evaluated without too much difficulty, i.e one dimensional integrals. The complexity of the model will determine the exact difficulty of the integrals. Further details can be found in DuMouchel (1990).

Both Skene and Wakefield (1990) and Rogatko (1991) consider the case when the responses in the individual trials are assumed not to be Normally distributed, but Binomial. Skene and Wakefield (1990) consider a hierarchical model for the combination of a series of trials when the trials can be considered to have binary outcomes on each of two treatment arms. Therefore if there are k trials the likelihood is the product of $2k$ Binomial distributions. Their parameterisation of the model differs in that for each trial they have two parameters, the log odds of 'failure' on the the standard treatment and the log-odds ratio of the new treatment compared to the standard treatment. Assuming exchangeability of the k trials, the second stage density may be also considered as a product over the k trials each one being conditional on hyper-parameters $\underline{\mu}$ and Σ. $\underline{\mu}$ represents the population means in the standard and new treatment groups, and Σ is the corresponding population covariance matrix. Skene and Wakefield (1990) assume that second stage density for each trial is bivariate Normal conditional on $\underline{\mu}$ and Σ. At the third stage they assume that $\underline{\mu}$ also has a Uniform prior density, but that the prior for Σ has an Inverted-Wishart prior density. Such a parameterisation leads to a reduction in the dimensionality of the problem. Eventually the marginal posterior densities are estimated using Gauss-Hermite quadrature. Unlike the method proposed by Carlin (1992) allows us to specify prior densities separately for the population mean parameters for the standard treatment and the new one.

Rogatko (1991) also considers parameterising his model in terms of the control 'failure' rate and the difference in the control and treatment 'failure' rates. He considers the case when we wish to work with the Binomial distribution, pointing out that the conditions for the Normal distribution to be a 'good' approximation to the Binomial distribution are that the individual trials have 'large' sample sizes. This is obviously not always true, and in many cases a meta analysis is attempted mainly *because* individually each trial has a small sample. Assuming Uniform prior densities for all the parameters Rogatko notes that there are two possible asymptotic approximations for the marginal posterior density for δ. The first approximation yields a Behrens-Fisher distribution for the marginal posterior density of δ and the second method yields a scaled t distribution.

3.3.4 Historical Controls

Finally, we consider an area of clinical trials that has been contentious, that of using historical controls. By historical we mean patients who are not part of the *current* study. Historical controls may either be patients in a previous

randomised trial or they may be in a specialist database, or of course both.

Pocock (1976) considered the use of historical controls as part of clinical trial, while Raghunathan (1991) considered the the analysis of a stratified case-control study when there are historical controls. Both use the controls to specify a prior density for the quantity of interest for the *current* control group. Before considering the use of historical controls in current studies we need to consider the validity of using historical controls at all. Pocock lists six criteria that need to be met in order to avoid the bias that may result from such an exercise. Such historical controls must;

1. have received the same precisely defined standard treatment as the controls in the current study.

2. have been part of a recent clinical study with the same patient eligibility criteria as the current one.

3. have methods of treatment evaluation which are the same as the current study.

4. have the same distribution of patient characteristics as the controls in the current trial.

5. have been in studies conducted at the same organisation with largely the same clinical investigators.

6. have no other indications that different results should be expected from the two sets of controls.

Assuming that these criteria are met Pocock considers using a Normal theory hierarchical model in the clinical trial setting. Assume that for each of the three groups of patients, the treatment group, the current control group and the historical control group there is a random variable, the quantity of interest, and that these are Normally distributed. Pocock assumes that the historical control group can be used as a prior for the randomised control group, and that the posterior density for the treatment effect can be obtained. He assumes that all mean hyperparameters have Uniform prior densities and all variance hyperparameters are replaced with sample variances. Analytic solutions are therefore possible. This modelling approach can also be used when the outcome is binary, and when the sample observations follow an exponential distribution.

Raghunathan (1991) uses a logistic regression model for the analysis of a stratified case-control study, and uses a multivariate Normal prior density for the regression parameters. The components of the prior density for the control group are the obtained from the historical control group. Parameter estimation can either be carried out using numerical integration techniques or by approximating the joint posterior density with a multivariate Normal density.

3.4 Summary

In this chapter we have seen how prior information may be elicited from clinicians, and show this information may be transformed into densities that can be used in a subsequent model. We have also seen that when there are previous studies, which are thought to be relevant to the current one, they can be formally in corporated into a meta analysis, and this can be used to derive a density for quantities of interest. Both of these methods for the formulation of prior densities have been applied to the neutron therapy study described in Chapter 2.

We have also reviewed the application of Bayesian methods to meta analyses. Whilst the application of such methods naturally leads to a density (posterior for the meta analysis, and prior for the current study), unless we are prepared to adopt an empirical Bayes formulation the resulting densities are not always mathematically convenient.

With reference to the elicitation and use of personal prior information an important point to make is the difference between *clinical demands* and *clinical beliefs*. *Clinical demands* relate to the treatment difference that would have to be present before a particular clinician would routinely use the treatment under test in clinical practice. In Chapter 2 we saw that in the neutron therapy study the average *clinical belief* was 46% failure on neutrons compared to 50% on photons, but that the *clinical demand* was 38.5% failure on neutrons compared to 50% failure on photons. Together with the point representing no treatment difference, clinical demands form an *equipoise* interval (Freedman, 1987) in which the treatments are deemed to be equivalent.

In any clinical study there is **always** prior information available, it is only the amount and the precision that vary. Before any major study is undertaken, whether it be a clinical trial or a prospective cohort study, there must be sufficient evidence of a desired effect to warrant the time and expense of such a study. In a clinical trial there is also an ethical consideration, since by letting a patient possibly receive a new treatment, the clinician must be 'confident' that it is no worse than the existing treatment.

Table 3.4 shows the prior densities that have been constructed for the neutron therapy trial described in Chapter 2.

Clinical Beliefs		Meta Analysis	
12 month failure rates		12 month failure rates	
Neutrons	Photons	Neutrons	Photons
$Beta[4.78, 5.68]$	$Beta[23.5, 23.5]$	$Beta[106.4, 98.6]$	$Beta[58.9, 121.8]$
Log relative risk		Log relative risk	
$Normal[-0.116, 0.08]$		$Normal[0.43, 0.02]$	

Table 3.4: Summary of Constructed Prior Densities for Neutron Therapy Study.

Chapter 4

Non-Temporal Models

4.1 Introduction

In this chapter we consider models for binary outcomes, which do not allow for survival times. There are a variety of studies where such models are applicable. For example, in survival studies the outcome is whether the patient is alive or dead at 1,2 or 5 years. In cancer clinical research the advances in anti-emetic drugs is supposed to reduce sickness, and this can be considered as a binary outcome.

In all of these types of studies, there may be prior information about either the rates of the outcome in each group separately, or about the comparison of the rates in the two groups. We have seen in Chapter 3 how such prior information, whether elicited from individuals or obtained from previous studies, may be used to construct prior probability densities for quantities of interest. In this chapter we consider the development of models in which such prior densities may be used when the outcome measure in each group is binary.

In the simplest case when there is a homogeneous group of patients and the event of interest is binary, the likelihood is proportional to a Bernoulli process. If the prior density that we use is a Beta density then as this is conjugate for a binomial likelihood, the posterior density is also a Beta density, such a model has been described by Raiffa and Schlaifer (1961).

In the more common case when there are two groups of patients to be compared, and the event of interest is again binary then it is possible to extend the above 'Beta-Binomial Model'. A number of methods of extension have been advocated, among them Zelen and Parker (1986), Marshall (1988) and Abrams, Ashby and Errington (Submitted). They apply a 'Beta-Binomial' model for each group and inferences about the odds ratio rely on a Normal approximation originally suggested by Lindley (1964, 1965). In Section 4.2.2 the model developed by Abrams, Ashby and Errington (Submitted) is described in detail and applied to the neutron therapy study described in Chapter 2. Section 4.2.3 reviews a number of other approaches to odds models, including Altham (1969), Leonard (1972, 1975), Aitchison and Bacon-Shone (1981) and Nurminen and Mutanen (1987).

Though these models are analytically tractable, and the approximations seem accurate, they do suffer from the deficiency that they cannot easily be extend to the case when there are covariates.

This particular deficiency can be addressed by adopting a logistic model formulation as suggested by Hughes (1991), and also used in medical context by O'Hagan (1990). West and Harrison (1989) have also considered the case of a Bayesian logistic regression model. This model develops the ideas of binary time series models in that it assumes a binary outcome is recorded at a number of time points, and we wish to relate the probability of the event occurring to a number of covariates.

A number of authors have applied the Normal theory models originally suggested by Raifa and Schlaifer (1961), including Spiegelhalter, Freedman and Parmar (In press) and Pocock and Hughes (1990) to clinical trials in which

there are only two groups of patients.

4.2 Odds Models

4.2.1 One group case

If we have a binomial likelihood then a conjugate prior density for the probability of success is the Beta distribution. This model, termed a 'Beta-Binomial' model, was described by Raiffa and Shlaifer (1961), Box and Tiao (1973), Berger (1985) and Smith (1985).

Consider a homogeneous group of n independent patients, and that after a period of time r of these patients experience an event of interest. If we assume that the probability for each patient of the event occurring is θ, then the likelihood for θ is Binomial in form, and is proportional to

$$\theta^{r}(1-\theta)^{n-r}$$

Assume that there is prior information about θ, and that this can be represented by a Beta distribution with parameters α and β, see Chapter 3, Section 3.2.2 for interpretation. The prior density for θ is therefore proportional to

$$\theta^{\alpha}(1-\theta)^{\beta}$$

Using Bayes' theorem we can obtain the posterior density for θ, which is proportional to

$$\theta^{r+\alpha}(1-\theta)^{n-r+\beta}$$

and this is a Beta density, with parameters $r+\alpha$ and $n-r+\beta$.

We may be interested in inferences about the odds of the event in question occurring, i.e $\theta/(1-\theta)$. Lindley (1965) suggests that if θ has a Beta distribution then the logarithm of the odds of θ has an approximate Normal distribution. In turn the odds of θ has an approximate Log-Normal distribution. More specifically if θ has a Beta distribution with parameters α and β then $(\beta+1)\theta/(\alpha+1)(1-\theta)$ has an F distribution with $2(\alpha+1)$ and $2(\beta+1)$ degrees of freedom. Fisher (1958) first suggested that the logarithm of the odds of θ has an approximate Normal distribution with mean $\log_e(\alpha+0.5/\beta+0.5)$ and variance $1/(\alpha+1)+1/(\beta+1)$. This approximation is valid for large and moderate values of α and β, say $\alpha>5$ and $\beta>5$, (Lindley, 1964).

4.2.2 Two group case

In the neutron therapy trial, described in Section 2.2, interest focused on comparison of the 12 month survival rates for the two treatments. Assume that the failure rates are θ_n for neutrons and θ_p for photons. If we assume a Beta prior density for both failure rates, then we can obtain the posterior densities for the two treatment groups separately. As we used conjugate Beta prior densities, the

posterior densities also have Beta densities. Thus we can use the method suggested by Lindley (1965) and described above in Section 4.2.1 to approximate the prior and posterior logarithm of the odds for both treatment groups. Since the logarithm of the odds of failure for both groups is approximately Normally distributed, the logarithm of the ratio of the odds is also approximately Normally distributed. Therefore we can obtain an approximate Log-Normal density for the posterior ratio of the odds of failure in one group compared to that in the other.

More formally, the likelihood for θ_p is proportional to

$$\theta_p{}^{r_p}(1-\theta_p)^{N_p-r_p}$$

where r_p is the number of photon patients dead at 12 months and N_p is the total number of photon patients in the trial. Similarly the likelihood for θ_n can be written in terms of r_n, the number of neutron patients dead at 12 months, and N_n the total number of neutron patients in the trial, and is proportional to

$$\theta_n{}^{r_n}(1-\theta_n)^{N_n-r_n}$$

If we assume *a priori* that θ_p has a Beta distribution with parameters α_p and β_p then the prior density is proportional to

$$\theta_p^{\alpha_p}\,(1-\theta_p)^{\beta_p}$$

where $\alpha_p > -1$ and $\beta_p > -1$. Assuming that θ_n has a Beta distribution, independently of θ_n, with parameters α_n and β_n gives a similar expression,

$$\theta_n^{\alpha_n}\,(1-\theta_n)^{\beta_n}$$

The parameters α_p, β_p, α_n and β_n can be estimated by using the 'Method of Moments' (Maritz and Lwin, 1989). We saw in Section 3.2.2 how this could be done in the neutron therapy example.

The prior information summarised by the prior densities can be combined with the information contained in the likelihood by using Bayes' Theorem. This states that the posterior density is proportional to the product of the likelihood and the prior density. Thus the posterior density for θ_p is proportional to

$$\theta_p^{\alpha_p+r_p}\,(1-\theta_p)^{\beta_p+N_p-r_p}$$

Similarly the posterior density of θ_n is proportional to

$$\theta_n^{\alpha_n+r_n}\,(1-\theta_n)^{\beta_n+N_n-r_n}$$

Applying the Normal approximation of the previous section and described briefly above in Section 4.2.1, to each of the posterior densities separately, yields the posterior density of $\log_e(\theta_p/1-\theta_p)$ as Normal with mean

$$\log_e\left(\frac{\alpha_p+r_p+0.5}{\beta_p+N_p-r_p+0.5}\right)$$

and variance

$$\frac{1}{\alpha_p + r_p + 1} + \frac{1}{\beta_p + N_p - r_p + 1}$$

Similarly, applying the same Normal approximation for the posterior density to $\log_e(\theta_n/1-\theta_n)$ yields a Normal density with mean

$$\log_e\left(\frac{\alpha_n + r_n + 0.5}{\beta_n + N_n - r_n + 0.5}\right)$$

and variance

$$\frac{1}{\alpha_n + r_n + 1} + \frac{1}{\beta_n + N_n - r_n + 1}$$

Therefore the logarithm of the ratio of the odds, $\log_e(\theta_n(1-\theta_p)/\theta_p(1-\theta_n))$ is also Normally distributed with mean

$$\log_e\left(\frac{(\alpha_n + r_n + 0.5)(\beta_p + N_p - r_p + 0.5)}{(\alpha_p + r_p + 0.5)(\beta_n + N_n - r_n + 0.5)}\right) \tag{4.1}$$

and variance

$$\frac{1}{\alpha_p + r_p + 1} + \frac{1}{\beta_p + N_p - r_p + 1} + \frac{1}{\alpha_n + r_n + 1} + \frac{1}{\beta_n + N_n - r_n + 1} \tag{4.2}$$

Often in a Bayesian analysis we wish to consider the case when there is relatively little prior knowledge about the parameters of interest. This is sometimes called a *reference analysis.* In the above analysis, Lindley (1965) suggests that vague prior information can be included by setting the hyper-parameters such that $\alpha_p = \alpha_n = \beta_p = \beta_n = -1$. Although this yields an improper prior density, it does have the property of setting the prior distribution for $\log_e(\theta_p/(1-\theta_p))$ and the prior distribution for $\log_e(\theta_n/(1-\theta_n))$ to be Uniform distributions over the real line. Geisser (1984) also considers the case when we wish to represent relative ignorance about the rate parameter of the binomial distribution. One non-informative prior that Geisser considers is a Uniform prior density on both θ_p and θ_n, also suggested by Bayes (1783). Geisser also reports a vague prior for θ_p as $1/\theta_p(1-\theta_p)$, and similarly $1/\theta_n(1-\theta_n)$ as a prior for θ_n. This is the same as that suggested by Lindley. Jeffreys (1961) suggests using $1/[\theta_p(1-\theta_p)]^{1/2}$ as a non-informative prior for θ_p and correspondingly $1/[\theta_n(1-\theta_n)]^{1/2}$ for θ_n.

In practice there is little difference between the various vague prior densities that have been suggested, especially when r_n, r_p, N_n and N_p are large.

Using the vague prior density suggested by Lindley the logarithm of the posterior ratio of the odds has a Normal distribution with mean

$$\log_e\left(\frac{(r_n - 0.5)(N_p - r_p - 0.5)}{(r_p - 0.5)(N_n - r_n - 0.5)}\right) \tag{4.3}$$

and variance

$$\frac{1}{r_p} + \frac{1}{N_p - r_p} + \frac{1}{r_n} + \frac{1}{N_n - r_n} \tag{4.4}$$

Lindley (1964) suggests ignoring the continuity corrections, '0.5' in (4.3). In this case a reference analysis yields the standard observed logarithm of the ratio of the odds, which will have a Normal distribution with mean

$$\mu = \log_e \left(\frac{r_n(N_p - r_p)}{r_p(N_n - r_n)} \right)$$

and variance the same as (4.2).

We may be interested in inference about the odds ratio instead of the logarithm of the odds ratio. As both the logarithm of the odds ratio and the odds for each treatment separately, are approximately Normally distributed the odds ratio and the odds for each treatment have approximate Log-Normal distributions. More formally if $\log_e \xi$ has a Normal distribution with mean μ and variance σ^2, then ξ has a Log-Normal distribution with mean $\exp(\mu+0.5\sigma^2)$ and variance $\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$, Johnson and Kotz (1970), page 115-117.

4.2.3 Example

Consider the MRC neutron therapy trial described in Section 2.2. We can consider comparing the outcome in the two treatment groups after 12 months follow-up. In order to adjust for the fact there is differential follow-up we can impute the numbers of patients who were alive and dead at 12 months using the actuarial survival proportions.

Clinical prior, using ALL patients.

Table 4.1 shows the imputed numbers of patients using the 12 month actuarial survival proportions 0.617 for photon patients and 0.486 for neutron patients. The frequencies in the body of Table 4.1 are obtained by assuming the marginal totals to be fixed, and using the actuarial survival proportions to obtain the corresponding cell frequencies.

	Photons	Neutrons	Total
Alive	38	45	83
Dead	24	47	71
Total	62	92	154

Table 4.1: Imputed patient status at 12 months based on actuarial survival probabilities for all patients using a censoring date of 21st December 1990.

Table 4.2 shows the results of fitting the 'Beta-Binomial' model described Section 4.2.2, using the prior densities obtained in Section 3.2.2 and shown in Table 3.4. Figure 4.1 shows the corresponding Log-Normal distributions. We can see from Table 4.2 that *a priori* the clinicians believed that the mean odds ratio was near one, i.e indicating no difference between the two treatments. The mean is slightly misleading as a measure of location as the Log-Normal density is skew to the left. We can see from Figure 4.1 that majority of the

density is less than one, and we can calculate this numerically as 59.7%. We can similarly calculate the clinicians' belief that the odds ratio is less than 0.63, their consensus point of clinical demand, is 31.6%.

We can also see from Table 4.2 and Figure 4.1 that if we use the vague prior suggested by Lindley (1965) and Geisser (1984) the corresponding reference posterior density has a mean of 1.749 for the odds ratio, and that very little of this density, i.e 6.6%, is less than one. Therefore, in the light of substantial prior information there appears to be little evidence to support the use of neutrons for pelvic tumours. We can also see that in the light of the data, the clinicians would revise their prior beliefs so that their posterior density had a mean of 1.359, a substantial shift towards the reference posterior density. The clinicians' posterior density also has a reduced standard deviation, compared to both the reference posterior and the clinicians' prior density, as the (amount of) evidence begins to accrue. Figure 4.1 shows this shift in the clinicians' belief well, and we can see that *a posteriori* the clinicians believe that the probability of neutrons being beneficial is 12.7%, i.e an odds ratio less than one, compared to 59.7% *a priori*.

	Mean	SD	$P(OR < 1)$	$P(OR < 0.63)$
Reference Posterior	1.749	0.601	0.066	0.002
Clinical Prior	1.046	0.739	0.597	0.316
Clinical Posterior	1.359	0.335	0.127	0.001

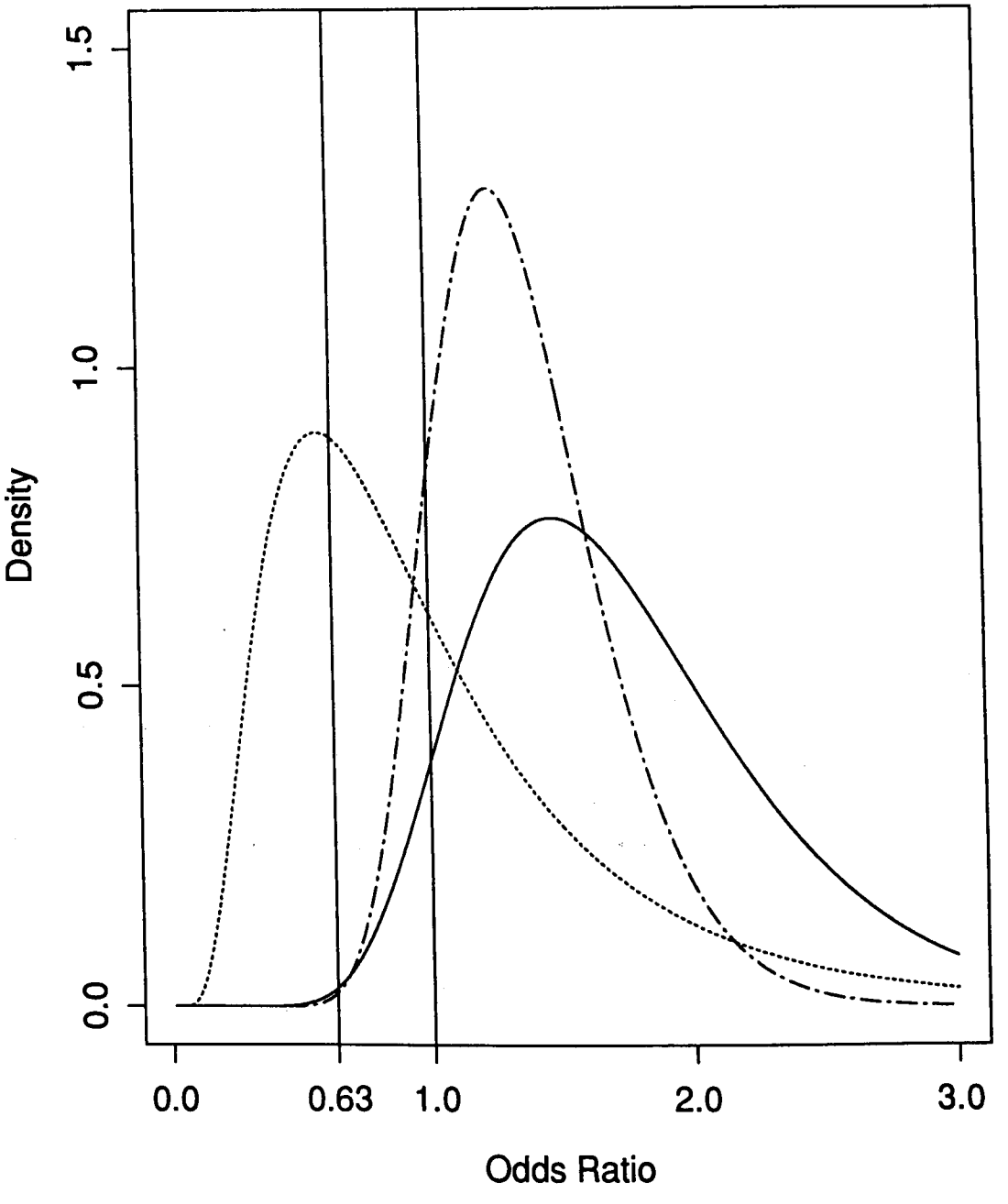
Table 4.2: Odds Ratio (OR) summary using clinical prior for ALL patients, i.e $n = 154$.

Clinical and Meta priors for bladder and rectum patients.

As mentioned in Section 3.2.3 in the derivation of a meta analysis prior for the two treatment groups separately, whilst it may be reasonable to generalise the results of previous studies on bladder and rectum patients to other areas of the pelvic region when we use a comparative measure such as the odds ratio or relative risk, this may not be so when we consider the effects of neutrons and photons separately. Therefore, in using the meta analysis prior densities for the two treatment groups in Table 3.4, we restrict our attention to the bladder and rectum patients in the neutron study. Thus, there are now 119 patients in the dataset rather than 154.

Table 4.3 shows the imputed numbers of patients using the 12 month actuarial survival proportions 0.53 for photon patients and 0.448 for neutron patients based on just the 119 bladder and rectum patients.

Table 4.4 gives summary statistics for the prior and posterior densities when a vague prior, a clinical prior or a meta prior density is used. Figure 4.2 shows the corresponding densities graphically. The most striking point from both Table 4.4 and Figure 4.2 is the discrepancy between the clinical prior density and that based on the previous study results. As above the mean clinicians'



····· clinical prior, —— posterior using reference prior, – · – · – posterior using clinical prior

Figure 4.1: Clinical prior, and clinical and rerefence posteriors for neutron therapy data (ALL patients).

	Photons	Neutrons	Total
Alive	24	33	57
Dead	21	41	62
Total	45	74	119

Table 4.3: Imputed patient status at 12 months based on actuarial survival probabilities for bladder and rectum patients using a censoring date of 21st December 1990.

prior density for the odds ratio was 1.046, whilst based on 119 patients the mean of the reference posterior density, using the vague prior suggested by Lindley and Geisser, is 1.526, but the mean of the prior density based on the previous studies is 2.271, suggesting strong evidence against the use of neutrons for tumours of the pelvic region. In the light of the data both the clinicians' and the meta based prior densities are 'pulled' towards the reference posterior density.

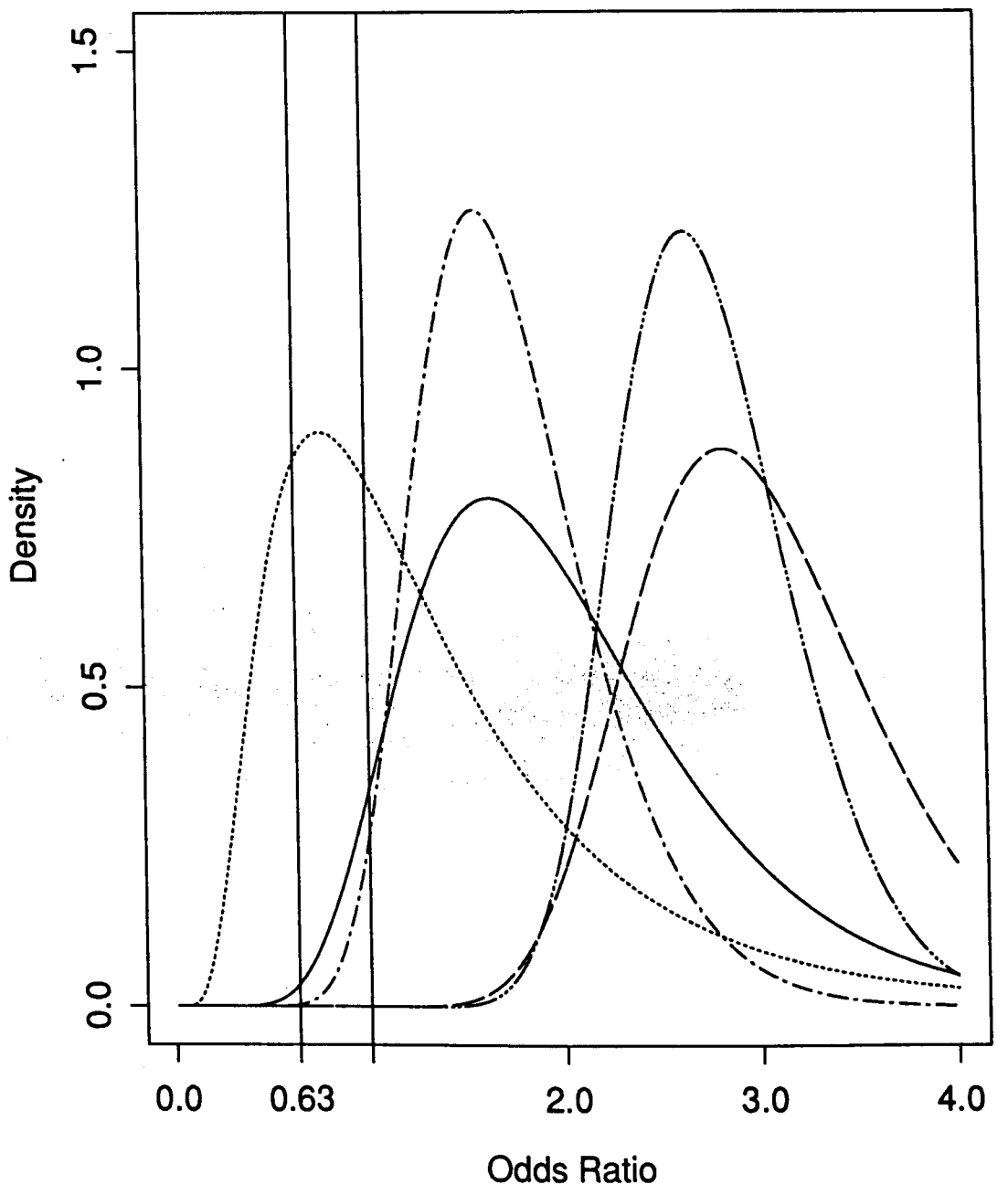
	Mean	SD	$P(OR < 1)$	$P(OR < 0.63)$
Reference Posterior	1.526	0.600	0.177	0.016
Clinical Prior	1.046	0.739	0.597	0.316
Clinical Posterior	1.304	0.348	0.189	0.004
Meta Prior	2.271	0.482	0.000	0.000
Meta Posterior	2.065	0.339	0.000	0.000

Table 4.4: Odds Ratio (OR) summary using meta analysis & clinical prior for rectum & bladder patients only, i.e $n = 119$.

Interpreting Table 4.4 and Figure 4.2, we see that for the ten clinicians, who had a prior consensus belief that neutrons were likely to be beneficial *a posteriori* they should no longer hold these beliefs. Instead they should believe that neutrons were likely to not be of benefit in treating these patients, but that they would still retain some belief in the benefit of neutrons. Individuals who prior to the trial being conducted expressed relative ignorance about the benefit of neutrons, would *a posteriori* believe that neutrons were not likely to beneficial but they would not totally rule out there being a possible benefit of neutrons. Individuals who *a priori* were prepared to accept the results of previous studies as the basis for the current beliefs about neutrons, in the light of the MRC trial, would have only revised their view that neutrons were definitely not beneficial slightly, and would give no credence to the possibility that neutrons may have some benefit.

4.2.4 Other Approaches to Odds Models

As mentioned in the introduction to this chapter a number of authors have developed and used an odds type models similar to that developed in the Sec-



Key : —— Reference posterior, ····· Clinical prior, – · – · – · – Clinical posterior, – – – Meta prior, – – · · · · – – · · · · Meta posterior

Figure 4.2: Log-Normal meta prior, meta posterior, clinical prior, clinical posterior & reference posterior densities for the odds ratio for rectum & bladder patients.

tions 4.2.1 and 4.2.2. In this section we shall give a brief review of these approaches and contrast them with the method described above.

There are not only differences in the development of these models but also in the area of application. Abrams, Ashby and Errington (Submitted) applied a Bayesian odds model approach to clinical trial, but Marshall (1988) and Zelen and Parker (1986) both apply a similar model to case-control studies.

The key difference between these approaches is at what stage the Log-Normal approximation is used, with Abrams *et al* using it separately for each of the two groups, which then means that the distribution of the ratio of the odds has a closed form. Marshall on the other hand only uses it at the final stage when the posterior density of the odds ratio has been obtained which does not have a closed form. However, Marshall does show that the Normal approximation is good even for small values in the 2×2 table by comparing the approximate posterior density with that obtained by numerical integration. Zelen and Parker (1986) also arrive at a Normal approximation for the posterior density for the odds ratio, but like Marshall (1988) they do so only after they have obtained the joint posterior for the odds ratio and a nuisance parameter. As with Marshall the approximate Normal posterior density for the odds ratio is the same as that derived by Abrams, Ashby and Errington (Submitted) apart from continuity corrections, with mean and variance the same as (4.1) and (4.2). Zelen and Parker also verify the fact that this approximation to the posterior density for the odds ratio is good, by comparing it with that obtained via numerical integration.

A different approach to the three described above has been suggested by Leonard (1972, 1975) in which he considered inference when there are m binomial distributions, and there was a two stage prior for each of the distributions. Obviously in the case $m = 2$, this model could be applied in a similar setting as those above. We describe this method briefly. Let θ_1 and θ_2 be the probability of an event occurring in the two groups, and parameterise the model as $\zeta_1 = \log_e(\theta_1/1 - \theta_1)$ and $\zeta_2 = \log_e(\theta_2/1 - \theta_2)$. The first stage prior for ζ_1 and ζ_2 are Normal distributions with mean μ and variance σ^2. The second stage priors are that μ has a Uniform distribution over the real line and $\lambda\nu/\sigma^2$ has a chi-squared distribution with ν degrees of freedom. λ and ν are assumed specified, and λ can be thought of as a 'best' guess at σ^2 and ν is an expression of the confidence in the 'best' guess. Leonard considers two distinct cases, the first when we assume that σ^2 is in fact known, and the second when σ^2 has a chi- squared distribution. In the former case the joint posterior density for $\underline{\zeta}$ is proportional to the product of the likelihood for $\underline{\zeta}$ and a multivariate Normal density. Leonard goes on to suggest that the marginal posterior density for $\underline{\zeta}$ can itself be approximated by a multivariate Normal density in the cases when the number of observed events are not small. In the case when we assume that σ^2 has a chi-squared distribution the posterior density for $\underline{\zeta}$ is proportional to the product of the likelihood for $\underline{\zeta}$ and a multivariate t density. The resulting posterior density can be approximated by a multivariate density but Leonard suggests that this approximation is only very rough and that ideally numerical

techniques should be employed to obtain the density and summary statistics.

A number of other authors have also considered Bayesian inference in a 2×2 table. The most notable are Altham (1969), Aitchison and Bacon-Shone (1981) and Nurminen and Mutanen (1987).

Altham (1969) considers the posterior density of the odds ratio, assuming a Dirichlet prior density for the joint probability of being in a specific cell of a 2×2 table. Altham notes that the exact posterior density of the odds ratio has an unpleasant form. However, she notes that the probability of the odds ratio being less than 1, may be written as a hyper-geometric summation. This fact leads to two Normal approximations for this probability.

Nurminen and Mutanen (1987) consider Bayesian inference in the 2×2 table for all three commonly used measures of association, the risk difference, the risk ratio and the odds ratio. They show that using a general Beta prior distribution the posterior distribution functions of each of measures of association are expressible as finite sums of Beta type quantities. These distribution functions are easily calculated by numerical methods, but Nurminen and Mutanen note that in the case of a Uniform prior density, the posterior density logarithm of the odds ratio may be approximated by a Normal density using the approximation suggested by Lindley (1964) and used by other authors above.

Aitchison and Bacon-Shone (1981) consider specifying separate Beta prior densities for the risk of an event in two groups, in the same way that Abrams, Ashby and Errington (Submitted) have done above. However, Aitchison and Bacon-Shone focus on the risk ratio rather than the odds ratio, i.e if θ_1 and θ_2 are the two risks, then the risk ratio is θ_1/θ_2. They note that re-parameterising in terms of γ, the transformed risk ratio, $\theta_1/\theta_1 + \theta_2$, the resulting posterior density for γ can be expressed in terms of hypergeometric integrals. This density can be easily programmed and has good convergence properties. The posterior density for the original risk ratio can then be obtained via transformation. They also consider the case when the Beta prior densities for the individual risks are set to be Uniform densities.

These methods are straightforward to implement, and the approximations that they rely on have been shown to be good. However, although these methods are valid as an initial analysis, they suffer from the fact that censoring cannot be properly accommodated, and nor can covariates.

4.3 Logistic Regression Models

In this section we consider an extension of the odds type models described in the previous section. Here the odds of an event are not only modelled by the treatment group but also by other covariates that are thought to affect outcome. We first consider static models in which the number of events in a specific time period is known. Second, we consider dynamic models in which the number of events that have occurred up to a number of time points is known.

4.3.1 Static Regression Models

Hughes (1991) describes the use of a logistic model in a clinical trial for the use of primary prophylaxis to reduce bleeding from the oesophageal varices, and the associated risk of mortality. Despite the four previous trials Hughes describes, there is no clear evidence of a reduction in mortality. He considers the use of a logistic model for the risk of dying.

Let Y_i be a binomial random variable with index m_i and parameter θ_i, and y_i be the observed values of Y_i. In Hughes' application y_i is the number that died in the ith group, and θ_i is the risk of dying in the ith group. Then the log-likelihood for $\underline{\theta}$, described in McCullagh and Nelder (1989) page 114 (and Cox (1970)) is

$$\ell(\underline{\theta}|\underline{y}) = \sum_{i=1}^{n} \left[y_i \log\left(\frac{\theta_i}{1-\theta_i}\right) + m_i \log(1-\theta_i) \right] \tag{4.5}$$

Assume that the link is the logit, therefore,

$$g(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \sum_{j=1}^{p} z_{ij}\beta_j \tag{4.6}$$

where z_{ij} is the jth covariate for the ith patient, β_j is the corresponding model parameter and p is the number of covariates. Therefore substituting (4.6) into (4.5) yields

$$\ell(\underline{\beta}|\underline{y}) = \sum_{i=1}^{n}\sum_{j=1}^{p} y_i z_{ij}\beta_j - \sum_{i=1}^{n}\sum_{j=1}^{p} m_i \log\left[1 + e^{z_{ij}\beta_j}\right] \tag{4.7}$$

Similarly the likelihood may be obtained as

$$L(\underline{\beta}|\underline{y}) = \prod_{i=1}^{n} \left\{ e^{\sum_{j=1}^{p} y_i z_{ij}\beta_j} e^{\sum_{j=1}^{p} m_i \log_e\left[1+e^{z_{ij}\beta_j}\right]} \right\} \tag{4.8}$$

Hughes (1991) considers the case when there are only two groups, i.e $p = 2$, so that z_{i1} is one for all patients and z_{i2} is one if the ith patient is in the second group. Therefore (4.6) reduces to

$$\log_e\left(\frac{\theta_i}{1-\theta_i}\right) = z_{i1}\beta_1 + z_{i2}\beta_2$$

and the log odds for group 1 is β_1, and the log odds for group 2 is $\beta_1 + \beta_2$. Therefore β_2 is the logarithm of the odds ratio. The likelihood, (4.8), for the two group case simplifies to

$$L(\beta_1, \beta_2|y_1, y_2, n_1, n_2) = e^{y_1\beta_1 + y_2(\beta_1+\beta_2)} \frac{1}{(1+e^{\beta_1})^{n_1}} \frac{1}{(1+e^{\beta_1+\beta_2})^{n_2}}$$

all that is needed are the total number of patients in each group and the number of patients in each group who died.

In the case when there is relatively little prior information about either of the parameters, the joint prior density can be assumed to have a Uniform distribution over the real plane, and therefore the joint posterior is proportional to the likelihood. Therefore,

$$p(\beta_1, \beta_2 | y_1, y_2, n_1, n_2) \propto e^{y_1\beta_1 + y_2(\beta_1+\beta_2)} \frac{1}{(1+e^{\beta_1})^{n_1}} \frac{1}{(1+e^{\beta_1+\beta_2})^{n_2}} \qquad (4.9)$$

For the purposes of inference, interest focuses on the marginal distribution of β_2, the logarithm of the odds ratio, and this can be obtained by integrating out β_1 from the joint posterior (4.9). In many situations there will be prior information available about the log-odds ratio (or about the actual odds ratio), and this can be represented by a density , $p(\beta_2)$. Retaining a Uniform prior for β_1, the joint posterior will be proportional to the product of $L(\beta_1, \beta_2 | y_1, y_2, n_1, n_2)$ and $p(\beta_2)$. As in the case when there was vague prior knowledge about both of the parameters real interest focuses on the log-odds ratio, β_2, and we therefore obtain the marginal posterior density for β_2 by integrating out β_1.

Hughes (1991) considers three cases in a hypothetical trial for beta-blockers, vague prior information for β_2, informative prior information for β_2 based on previous trial results, and finally an informative prior based on clinical opinion.

The easiest distribution to use for the informative prior case is the Normal distribution, at least from the case of specifying the prior information. However, using a Normal distribution does mean that the integration needed to obtain the marginal posterior density for β_2 is not analytically tractable, and either numerical integration techniques or asymptotic approximation methods need to be used. Remembering that this is the case when there are only two groups of patients, and that there have been no other covariates in the model, in the case when there are covariates we will require to evaluate higher order integrals. We shall see later in this thesis that for more complicated models we will have to resort to either one of these integration methods.

O'Hagan *et al* (1990) also considered the use of Bayesian logistic models in predicting whether patients with keratoconus will require a corneal transplant. They considered the case when the initial corneal radius was known, and wanted to assess the value of this in predicting a future corneal transplant. For the case when there was no informative prior knowledge they considered the use of a grid of values over which the joint posterior was evaluated, and then the marginal densities and moments were obtained by the relevant summation of the grid elements. These calculations were performed on a PC, and for relatively simple models, i.e involving only one or two covariates, the time taken to perform these calculations was acceptable. However, O'Hagan *et al* (1990) note that for more sophisticated problems more powerful numerical integration techniques need to be used, such as Gauss-Hermite quadrature (Naylor and Smith, 1982).

4.3.2 Dynamic Regression Models

West and Mortera (1987) and West and Harrison (1989) develop and apply a Bayesian model for binary time series. These models are part of a more

general class of models for exponential family distributions, termed Dynamic Generalised Linear Models (DGLM), which are themselves extensions of the Dynamic Linear Model (DLM) proposed by Harrison (1989).

Consider the number of patients that have died in a clinical trial by time t, denote this by Y_t, and denote by n_t the number of patients at risk immediately prior to time t. Obviously several observations may be made, and these will form a binary time series. $D_t = Y_1, Y_2, \cdots, Y_{t-1}, Y_t$ which represents the information about the time series up to time t. West and Harrison define a Dynamic Generalised Linear Model (DGLM) by an observation equation and an evolution equation. The evolution equation relates the observed quantities to the linear regression function. Thus the observation equation is

$$p(Y_t|\mu_t, n_t) = \binom{n_t}{Y_t} \mu_t^{Y_t} (1-\mu_t)^{n_t - Y_t} \qquad Y_t = 0, 1, \cdots, n_t \tag{4.10}$$

where μ_t is the probability of a death, and there is a link function $g(\cdot)$ such that $g(\mu_t) = \underline{F}_t^T \underline{\theta}_t$ where $\underline{F}_t$ are covariates and $\underline{\theta}_t$ are the corresponding parameters. The evolution equation is

$$\underline{\theta}_t = \underline{G}_t \underline{\theta}_{t-1} + \omega_t \quad \omega_t \sim [\underline{0}, W_t] \tag{4.11}$$

where G_t is an evolution matrix determining how the parameters vary over time and W_t is a covariance matrix determining how the parameters vary with one another.

Suppose we know the value of $\underline{\theta}_{t-1}$ at time $t-1$, up to the first moments thus

$$(\underline{\theta}_{t-1}|D_{t-1}) \sim [\underline{m}_{t-1}, C_{t-1}]$$

It follows from (4.11) that

$$(\underline{\theta}_t|D_{t-1}) \sim [\underline{a}_t, R_t]$$

where $\underline{a}_t = G_t \underline{m}_{t-1}$ and $R_t = G_t C_{t-1} G_t^T + W_t$. Now what is required is the posterior density for $\underline{\theta}_t$. By assuming a conjugate Beta prior distribution for μ_t , the posterior distribution for μ_t may be obtained. The posterior distribution for $\underline{\theta}_t$ can then be obtained using Linear Bayes Methods (West and Harrison, 1989).

Further details are given in Chapter 14 of West and Harrison (1989) and West, Harrison and Migon (1985) also consider the development and application of DGLMs. In particular they a consider the comparison between GLIM (1978) and *static* DGLMs, by specification of G_t and W_t. Although these models, like the static regression models considered above, allow for the inclusion of covariates, there is no explicit consideration of the actual survival times of the patients. The extension of Dynamic Linear Models to accommodate survival times will be described briefly in Chapter 6.

4.4 Normal Theory Models

4.4.1 General Background

Bayesian inference when the quantity of interest can be assumed to be Normally distributed were described by Raiffa and Schlaifer (1961). They have also been described by Box and Tiao (1973) and Berger (1985).

Consider the case when $X_1, \cdots, X_n$ form a random sample from a Normal distribution with mean μ and variance σ^2. The sample mean $\overline{X}$, has a Normal distribution with mean μ and variance σ^2/n.

The two key assumptions of such a model are; first that the quantity of interest is assumed to be Normally distributed, and second that the variance of this distribution is assumed to be known.

Jeffreys (1961) described the use of a Normal prior density for μ, with mean ν and variance σ^2/m, where ν and m are assumed known. Using Bayes' theorem the posterior distribution for μ is Normal with mean $\frac{\mu n+\nu m}{n+m}$, and variance $\sigma/\sqrt{n+m}$, DeGroot (1986), page 324.

4.4.2 BART

Spiegelhalter, Freedman and Parmar (In press) considers the case when there are two patient groups, and that we have a comparative measure of interest for the two groups. For example in the case of a binary outcome this may the relative risk, or the odds ratio, or in the case when we are considering survival situations it may be the hazard ratio. The underlying statistical model is that described above. We give brief details of the methods involved and the application of such a model to the neutron therapy data outlined in Chapter 2, using both the clinical and meta analysis prior densities developed in Chapter 3.

A software package BART[1] written in Splus (Statistical Sciences Inc., 1990) has been developed for implementing these models.

BART allows there to be a mixture of two prior densities, providing that a weighting factor is also specified. For each prior density the corresponding posterior density is obtained as above. Posterior weights are obtained by updating the likelihood and the corresponding posterior mixture density is obtained. BART also allows the specification of either a single truncated Normal prior distribution or a mixture of two such distributions.

The graphical model outputs are the prior density, the standardised likelihood, the posterior density, and the predictive density based on the prior density. The relevant probabilities of being in certain intervals are calculated for being below the equipoise interval, in the equipoise interval (defined in Section 3.4) and above it.

[1] Available from Dr. D. J. Spiegelhalter, MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge, CB2 2SR.

Example

Applying this model to the neutron therapy example, for both the case when there is prior clinical opinion, and when there is meta prior, the following results were obtained.

Model Inputs

Data	:	$\sigma = 2.04$, $x_m = 0.399$, $m = 109$, $n = 109$ $\delta_I = -0.26$, $\delta_S = 0.0$
Clinical Prior	:	mean=-0.116, precision=51
Meta Prior	:	mean=0.43, precision=209
Mixture Prior	:	mean1=-0.116, precision1=51, mixture weight=0.5 mean2=0.43, precision2=209, mixture weight=0.5

where σ is the standard deviation of an observation, x_m is the observed sample mean, whilst m is the number of observed deaths and n is the number of anticipated future deaths. δ_I, δ_S are upper and lower limits for the interval in which clinical opinion considers the two treatments equivalent. Precision is defined in BART as 'the number of subjects in a imaginary trial that would have yielded these results, assuming that the variation was the same as in the actual trial'.

If the two treatment groups have equal numbers in them, then σ the standard deviation for an observation can be shown to be equal to 2, Tsiatis (1981). However, in the case of the neutron therapy trial, because of the initial bias in randomisation towards neutrons, the two treatment groups did not have equal patient numbers. In order to allow for such an imbalance it can be shown (Ashby, personal communication) that the standard deviation should be changed to 2.04.

Thus, for the clinical prior, assuming a standard deviation of 0.286, together with an imputed trial standard deviation of 2.04, will yield a precision of $(2.04/0.286)^2 = 51$, and similarly for the meta prior the precision will be $(2.04/0.141)^2 = 209$.

	Mean	SD	$P(logHR < 0)$	$P(logHR < -0.26)$
Reference Posterior	0.399	0.195	0.000	0.020
Clinical Prior	-0.116	0.286	0.307	0.658
Clinical Posterior	0.235	0.161	0.001	0.073
Meta Prior	0.430	0.141	0.000	0.001
Meta Posterior	0.419	0.114	0.000	0.000
Mixture Prior	0.157	0.385	0.154	0.330
Mixture Posterior	0.385	0.144	0.000	0.014

Table 4.5: Estimates for the log Hazard Ratio (HR) from BART for the neutron therapy example.

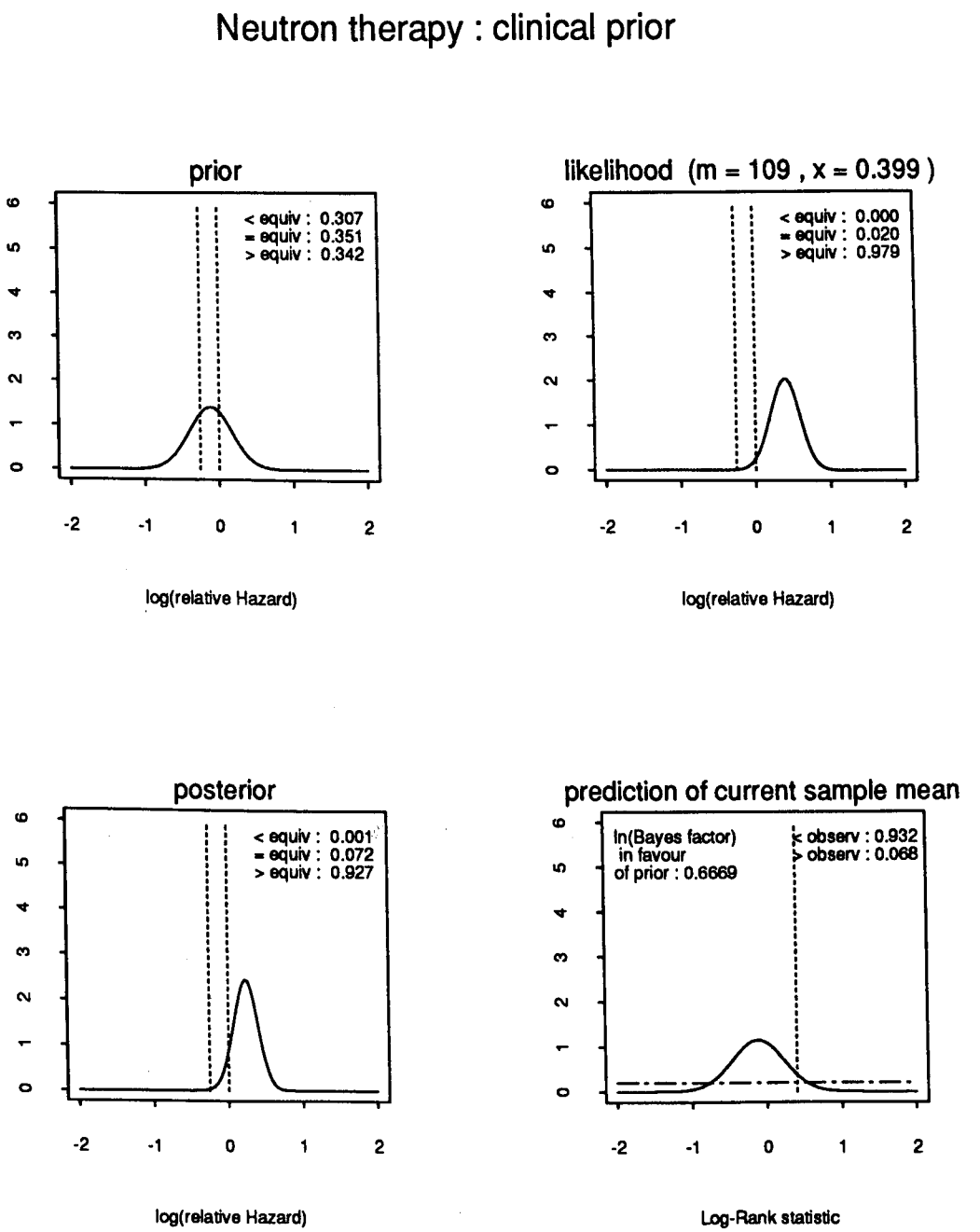


Figure 4.3: Output from BART using a clinical prior for the hazard ratio.

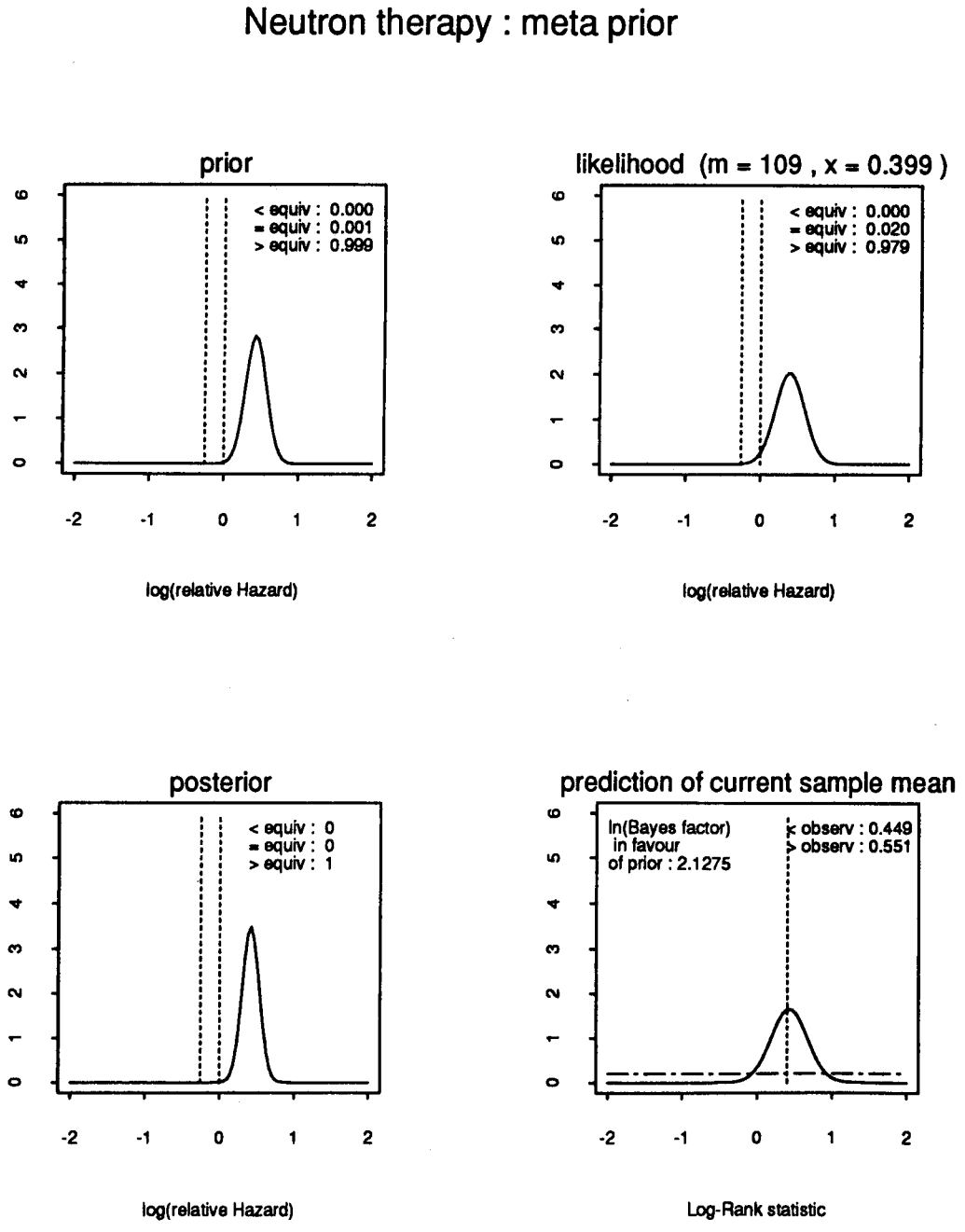


Figure 4.4: Output from BART using a meta prior for the hazard ratio.

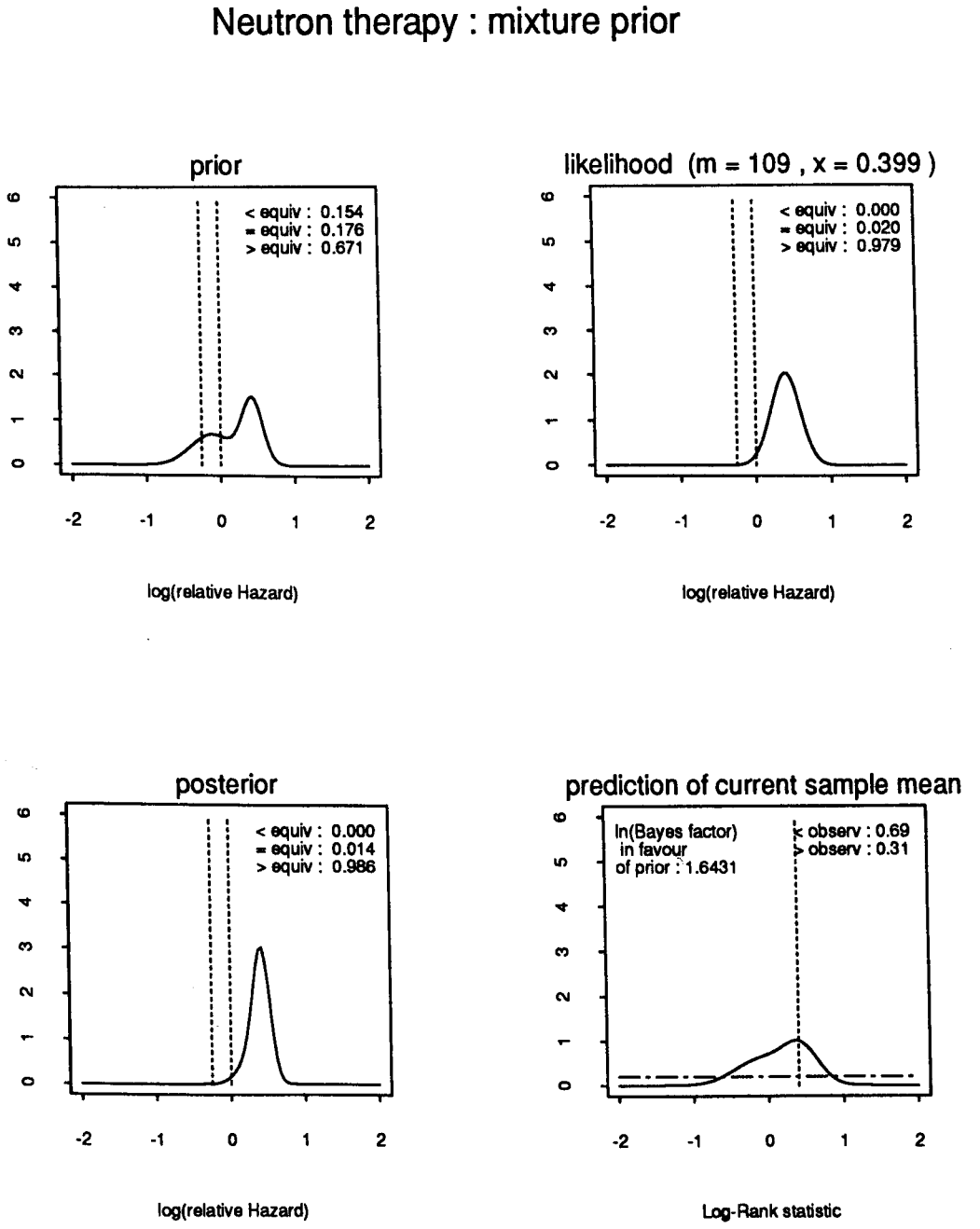


Figure 4.5: Output from BART using a mixture clinical and meta prior for the hazard ratio, with a mixing coefficient of 0.5.

Results

Table 4.5 and Figures 4.3 to 4.5 show the results of using Spiegelhalter's method in the neutron therapy example. The clinical and meta prior densities obtained in Chapter 3 are used, together with a mixture prior, which incorporates both the clinical and meta analysis prior information.

From Figures 4.3 and 4.4 and Table 4.5 we can see that the results for both the clinical and meta based prior are similar to those obtained using the odds ratio method developed by Abrams, Ashby and Errington (Submitted), and shown in Figures 4.1 and 4.2 and Tables 4.2 and 4.4. The main difference being that in BART we use the logarithm of the hazard ratio as the quantity of interest rather than the odds ratio or logarithm of the odds ratio. The clinicians' belief is shifted by evidence from the trial, so that *a posteriori* they have substantially less belief in the efficacy of neutrons for tumours of the pelvic region. Beliefs' based on the meta analysis described in Chapter 3, are merely reinforced by the trial data, as they *a priori* indicated that neutrons were not beneficial in treating cancer of the pelvic region.

Using both models when there is informative prior information there is reduction in the standard deviation of the parameter estimate compared to that in a reference analysis.

4.4.3 Another example

Pocock and Hughes (1990) also use a 'Normal-Normal' model approach in analysing a primary prevention trial into the effect of diet and smoking on coronary heart disease. They specify that a median reduction of 20% in the number of coronary events is expected. Using this figure they specify a prior Normal density for the mean log relative risk, and thus assuming that the observed log relative risk is also Normally distributed, applying Bayes' theorem, as Section 4., yields a Normal posterior density for the mean log relative risk.

4.5 Summary

In this chapter we have shown that there are a number of simple Bayesian models that can be applied to clinical trials. These methods are relatively straight forward to implement and serve as useful initial analysis of clinical trial data.

Parameter estimation in both the odds model approach and the Normal theory models do not require the use of sophisticated estimation techniques, though neither model explicitly allows for differential follow-up nor the inclusion of covariates. The latter deficiency may be addressed by the use of a logistic regression model. Parameter estimation in a logistic model requires the use of either numerical integration or asymptotic approximation techniques, and the problem of censored data remains unresolved.

In the following chapters we address the problems of allowing for survival times, censoring and the inclusion covariates by considering models in which the time to death or censoring is modelled explicitly. Such models will though require the routine use of sophisticated estimation techniques.

Chapter 5

Counting Processes

5.1 Introduction

The aim of this chapter is to introduce counting process notation, and to describe some of the models for event history data that make use of it. Such notation is necessary for the development of models in subsequent chapters.

We start the chapter by introducing a *univariate counting process.* which is a stochastic process that counts the number of events that have occurred as time progresses. Associated with a counting process is an *observational* or *at-risk process*, which counts the number of possible events as time progresses. The behaviour of a counting process is governed by its random *intensity process.* These concepts are extended to the case when there may be more than one *type* of event. Such a situation can be described by a *multivariate counting processes* , together with its associated at-risk process and intensity process.

The theory of counting processes has proved useful in applied problems for two reasons;

1. Aalen (1978) and Johansen (1983) have both pioneered the use of counting process notation in survival analysis, because of the powerful mathematical tools that may be used as a result. Such mathematical tools have enabled the derivation of a number of estimators, and the rigorous proof of their properties.

2. Gill (1984), Andersen (1985, 1988) and Clayton (1988) have advocated the use of counting process notation in describing complicated event-histories, of which survival data is a special case.

In traditional survival models interest often focuses on the hazard rate. Analogously in models for counting processes interest focuses on the intensity process. Thus the main type of model to have been suggested is the multiplicative intensity model. This was originally advocated by Aalen (1978), and includes as special cases many of the more familiar models in survival analysis, such as the Cox proportional hazards model. We describe parameter estimation in the multiplicative intensity model when the model has various parametric forms. Voelkel and Crowley (1984), Keiding and Andersen (1987) and Frydman (1991) all consider the case when the intensity process is estimated non-parametrically. A number of authors including Andersen and Gill (1982) and Jacobsen (1984) have considered the case when the intensity process comprises two separate components, one of which is estimated non-parametrically and the other parametrically, analogous to Cox's proportional hazards model (Cox, 1972). Finally Borgan (1984), Hjort (1986) and Aven (1986) have considered the case when the intensity process is assumed to have a particular parametric form, the latter two considering estimation from a Bayesian perspective. We shall see that the relationship between counting processes and Martingales plays a key role, not only in the motivation of estimators, but also in the derivation of their properties. Finally we comment on the link between parametric survival models and Poisson regression models, as mentioned by Holford (1980), Laird and Oliver

(1981), Lawless (1987) and Clayton (1988). This link is highlighted by adopting a counting process approach.

The chapter concludes with a summary of the key issues that have been addressed.

5.2 Univariate and Multivariate Counting Processes

The following definitions and derivations follow closely those of Andersen and Borgan (1985), Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1992).

5.2.1 Univariate Counting Processes

Consider that for a homogeneous group of patients, i.e they share a common intensity, we have a stochastic process that counts the number of occurrences of specific event up to time t, assuming time to continuous. Denote this *counting process* by $N(t)$. $N(t)$ will be a right continuous step function and will have 'jumps' of size 1 if a single patient has an event at the same time. $N(t)$ is termed a *Univariate counting process.* As well as there being a counting process, $N(t)$, there is also an observable *at-risk process*, denoted $Y(t)$, which counts the number of patients at risk of the specific event just prior to time t. Figures 5.1 and 5.2 shows hypothetical examples of $N(t)$ and $Y(t)$, when there are no tied failure times.

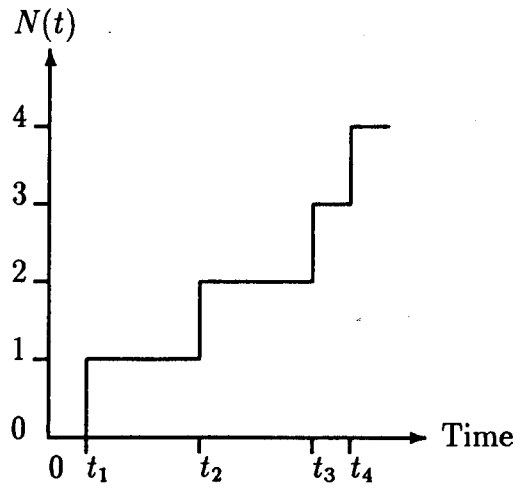


Figure 5.1: A simple counting process, $N(t)$.

A univariate counting process is determined by its *intensity process*, which we shall denote by $\alpha(t)$. Assuming no tied failure times $\alpha(t)\delta t$ can be defined as

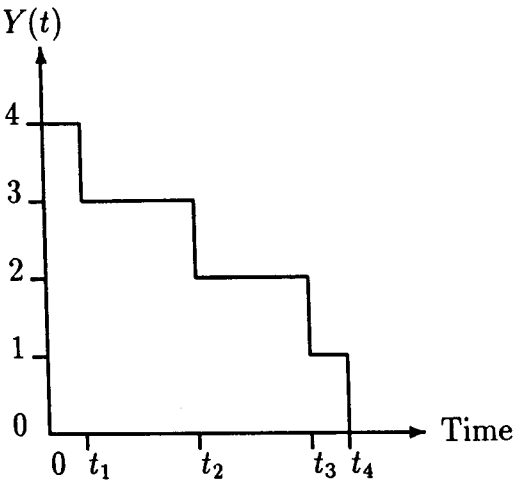


Figure 5.2: A simple risk process $Y(t)$.

the conditional probability that $N(t)$ 'jumps' in a small interval, δt after time t, given everything that had happened until just before t. A short-hand way of describing the information up to a certain time point is $\mathcal{F}_{t-}$, this represents the information just prior to time t, and is called a *filtration*. Thus we talk about a counting process $N(t)$ being adapted to a filtration $\mathcal{F}_{t-}$.

$$\alpha(t)\delta t = P\{[N(t+\delta t) - N(t)] = 1|\mathcal{F}_{t-}\} \tag{5.1}$$

or

$$\alpha(t)\delta t = E(dN(t)|\mathcal{F}_{t-})$$

The intensity process, $\alpha(t)$, is a random quantity since it is conditional on the filtration $\mathcal{F}_{t-}$, which is itself composed of random quantities. The *compensator* of a counting process or the cumulative intensity function , denoted $A(t)$, is defined by

$$A(t) = \int_0^t \alpha(u)du \tag{5.2}$$

5.2.2 Multivariate Counting Processes

We defined above a univariate counting process for a homogeneous group of patients. We now consider the patients individually and define a *multivariate counting process* $N(t) = \{N_i(t), i \in I, t > 0\}$ to be a stochastic process that can be thought of as counting the number of events for patient i up to and including time t, and where $I = \{1, \cdots, n\}$. We will assume that each component of $N(t)$ has 'jumps' of size 1, and that no two processes can 'jump' simultaneously.

In the situations that we will consider some of the patients will not be 'at risk' all the time. Therefore we extend the notion of an observational or at-risk process to allow for this. This process indicates at all time points whether

a particular individual is at risk. Thus $Y(t) = \{Y_i(t), i \in I, t > 0\}$, and as in the univariate case this is a predictable process, given the filtration $\mathcal{F}_{t-}$. Algebraically,

$$Y_i(t) = \begin{cases} 1 & \text{patient } i \text{ is at risk at time } t \\ 0 & \text{patient } i \text{ is not at risk at time } t \end{cases}$$

As in the univariate case $N(t)$ is determined by its random intensity process $\alpha(t)$, which comprises components $\alpha_i(t)$ for each individual. Thus, $\alpha(t) = \{\alpha_i(t), i \in I, t > 0\}$. Figures 5.3 and 5.4 shows a simple example for survival data, in which patient i dies at time t_1, and therefore $Y_i(t)$ is one until t_1 and zero after t_1, and correspondingly, $N_i(t)$ is zero until time t_1 and one after t_1, indicating that an event has occurred. As in the univariate case each $\alpha_i(t)\delta t$ can be thought of as the conditional probability that $N_i(t)$ 'jumps' in a small interval δt about t, given the filtration $\mathcal{F}_{t-}$. Therefore,

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

We may extend the idea of a counting process to the case when there are more than one type of event that can occur. In this case the multivariate counting process $N(t) = \{N_{ik}(t), i \in I, k \in K, t > 0\}$ can be thought of as counting the number of occurrences of event type k for patient i up to and including time t, and where $K = \{1, \cdots, h\}$. In some examples we may also restrict the components, $N_{ik}(t)$, to be counting non-recurrent events. In this case the value of $N_{ik}(t)$ will either be 1 or 0 depending whether patient i has experienced event k. Similarly the at-risk process, $Y(t)$ can be defined $Y(t) = \{Y_{ik}(t), i \in I, k \in K, t > 0\}$. This process indicates when individuals are at risk of a specific type of event. Thus,

$$Y_{ik}(t) = \begin{cases} 1 & \text{patient } i \text{ is at risk of transition } k \text{ at time } t \\ 0 & \text{patient } i \text{ is not at risk of transition } k \text{ at time } t \end{cases}$$

As with the multivariate counting process when there was only one type of event, $N(t)$ is governed by its random intensity process $\alpha(t)$, which comprises components for each individual and each type of event. Thus, $\alpha(t) = \{\alpha_{ik}(t), i \in I, k \in K, t > 0\}$, where $\alpha_{ik}(t)\delta t$ is the conditional probability that individual i will have an event of type k in a small time interval δt about t.

Analogous to the univariate case the progress, as time unfolds, of a multivariate counting process, $N(t)$, is determined by its random intensity process, $\alpha(t)$. For an nh-variate counting process there is an nh-variate intensity process $\alpha(t)$ such that

$$\alpha(t) = \{\alpha_{ik}(t), i \in I, k \in K, t > 0\}$$

where each component, $\alpha_{ik}(t)$, can be defined,

$$\alpha_{ik}(t)\,\delta t = \mathrm{P}\left(N_{ik}(t + \delta t) - N_{ik}(t) = 1|\mathcal{F}_{t-}\right) \tag{5.3}$$

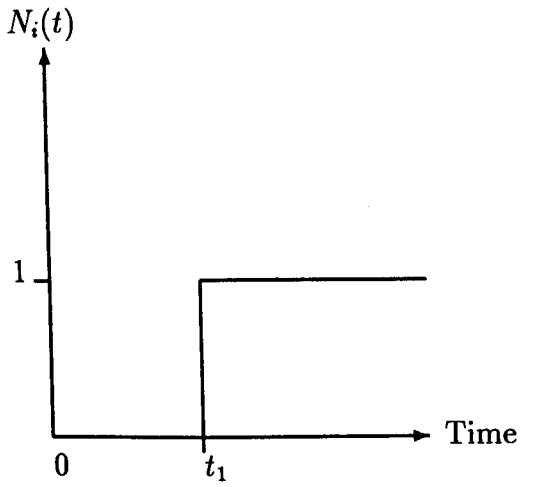


Figure 5.3: A counting process, $N(t)$, for survival data.

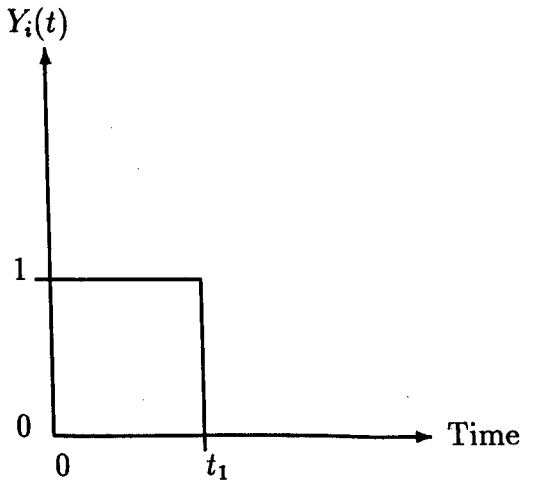


Figure 5.4: A simple risk process, $Y(t)$, for survival data.

where $\mathcal{F}_{t-}$ represents all the available information on the disease history before and just up to time t. Assume

$$dN_{ik}(t) = \begin{cases} 1 & \text{if } N_{ik}(t+\delta t) - N_{ik}(t) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Then we can re-write (5.3) as

$$\alpha_{ik}\,\delta t = \mathrm{P}\left(dN_{ik}(t) = 1 | \mathcal{F}_{t-}\right)$$

The multivariate counting process, $N(t)$, can be extended to accommodate the case when there are a number of time intervals or *epochs*. This is desirable when there are age cohorts or clearly defined time intervals. In this case the multivariate counting process $N(t) = \{N_{ijk}(t), i \in I, j \in J, k \in K, t > 0\}$ can be thought of as counting the occurrences of event type k for patient i in epoch j at time t, and where $J = \{1, \cdots, m\}$. Similarly $Y(t)$ comprises components which indicate whether an individual is at risk of a specific type of event in a particular epoch. Thus,

$$Y_{ijk}(t) = \begin{cases} 1 & \text{patient } i \text{ is at risk of transition } k \text{ at time } t \text{ in epoch } j \\ 0 & \text{patient } i \text{ is not at risk of transition } k \text{ at time } t \text{ in epoch } j \end{cases}$$

As in the previous cases $N(t)$ is determined by an nmh-variate intensity process, $\alpha(t)$ comprising components for each individual, each type of event and for each epoch. Thus, $\alpha(t) = \{\alpha_{ijk}(t), i \in I, j \in J, k \in K, t > 0\}$. Each $\alpha_{ijk}(t)\delta t$ can be thought of as the conditional probability that individual i has an event of type k in epoch j in a small time interval δt about t.

5.3 The Multiplicative Intensity Model

In this section we consider the statistical models that may be applied to counting processes. In Section 5.2 we saw how a counting process, $N(t)$, is governed by its random intensity process, $\alpha(t)$. Therefore we consider modelling the intensity, which is analogous to modelling the hazard rate in standard survival analysis. We may consider non-parametric, parametric or semi-parametric estimation of the intensity process, $\alpha(t)$. The intensity may vary between groups of individuals or it may vary over time according to some parametric family of functions. As we have suggested in earlier sections of this chapter, a class of models that incorporates many of the applied settings discussed above is called the *multiplicative intensity model.* Below we give a more precise definition of the model and consider estimation of the model parameters.

Aalen (1978) developed and advocated the use of the multiplicative intensity model. This model assumes that, conditional upon various criteria, the intensity function, $\alpha(t)$, is a product of the at-risk stochastic process, $Y(t)$, and an unknown function $\lambda(t)$. Following Aalen (1978), Andersen and Borgan (1985) and Fleming and Harrington (1991), we define the multiplicative intensity model as follows.

Definition

Assume that the data can be represented by n triples $(N_i(t), Y_i(t), \underline{Z}_i(t))$, $i = 1, \cdots, n$. $N_i(t)$ are counting processes, $Y_i(t)$ are at-risk processes, defined in Section 5.2.2, and $\underline{Z}_i(t)$ are covariate processes. The only condition is that $\underline{Z}_i(t)$ is fixed given the right-continuous filtration $\{\mathcal{F}_{t-}, t \geq 0\}$ representing the statistical information accruing over time. The n intensity processes $\alpha_i(t)$ are products of an unknown function $\lambda_i(t)$ and the stochastic process $Y_i(t)$. We have seen that $A(t)$, the compensator of $N(t)$, is adapted to the filtration $\{\mathcal{F}_{t-}\}$, and therefore $Y_i(t)$ are also required to be adapted to $\{\mathcal{F}_{t-}\}$, i.e given information up to, but not including, time t, the process $Y_i(t)$ is *predictable*. Algebraically,

$$\alpha_i(t) = \lambda_i[t, \underline{Z}_i(t)]Y_i(t) \tag{5.4}$$

where $\lambda_i[t, \underline{Z}_i(t)]$ is an unknown function.

□

Similarly, if we extend the multiplicative intensity model in (5.4) to the case in which there is more than one type of event, indexed by k, (5.4) becomes

$$\alpha_{ik}(t) = \lambda_{ik}[t, \underline{Z}_{ik}(t)]Y_{ik}(t) \tag{5.5}$$

Note that the dimension of $\underline{Z}_{ik}(t)$ may also vary with k, i.e some covariates may be thought to affect the intensity of some events and not others. If we also consider the case when there are a number of epochs, indexed by j, (5.4) becomes

$$\alpha_{ijk}(t) = \lambda_{ijk}[t, \underline{Z}_{ijk}(t)]Y_{ijk}(t) \tag{5.6}$$

In the following sections we shall show how by defining both $\lambda(t)$ and $Y(t)$, the multiplicative intensity models in (5.4)-(5.6) can accommodate a broad range of different models used in event history analysis. Different censoring mechanisms may be accommodated via the at-risk process $Y(t)$, whilst the functional form of the model can be specified via $\lambda(t)$. Whilst the fully parametric case is discussed in detail, the two cases in which $\lambda(t)$ is non-parametric or semi-parametric are considered only briefly.

5.3.1 Non-Parametric Intensity Model

A number of authors including Voelkel and Crowley (1984), Keiding and Andersen (1987), Fleming and Harrington (1991) and Frydman (1991) have considered the case when we assume that $\alpha(t)$ varies freely over all the possible values that $\alpha(t)$ can take. Fleming and Harrington (1991) show that $N(t) - A(t)$ being a Martingale, yields the Nelson-Aalen estimator as a natural estimator of the compensator $A(t)$, defined in (5.2). This estimate is analogous to the Nelson-Altschuler estimate of the cumulative hazard in standard survival analysis. As a consequence of Martingale calculus proof of many desirable properties of Nelson-Aalen estimator is possible. See Fleming and Harrington (1991) for further details.

5.3.2 Semi-Parametric Intensity Model

One possible functional form for $\lambda(t)$ that has been widely used is the proportional hazards or Cox regression model (Cox, 1972). In this model $\lambda(t) = \lambda_0(t)e^{\underline{\beta}^T \underline{Z}_i(t)}$ where $\lambda_0(t)$, the baseline hazard, is left completely unspecified, and $\underline{\beta}$ is a vector of regression parameters. Thus, (5.4) becomes

$$\alpha_i(t) = \lambda_0(t)e^{\underline{\beta}^T \underline{Z}_i(t)}Y_i(t)$$

Inference is then made via what was termed a partial likelihood, (Cox, 1975). The probability, conditional on surviving up to time t, that an individual dies at time t is the ratio of the intensity for the individual to the sum of the intensities of all individuals at risk at time t. Algebraically,

$$P\{\text{patient } i \text{ dies at time } t | i \text{ survives up to } t\} = \frac{\alpha_i(t)}{\sum_{g \in \mathcal{R}(t)} \alpha_g(t)}$$

where $\mathcal{R}(t)$ is the set of patients at risk immediately prior to time t. This yields a partial likelihood for $\underline{\beta}$,

$$L(\underline{\beta}) = \prod_{i=1}^{n} \left\{ \frac{\alpha_i(t)}{\sum_{g \in \mathcal{R}(t)} \alpha_g(t)} \right\}^{dN_i(t)} \tag{5.7}$$

in the case when $\alpha(t) = \lambda_0(t)e^{\underline{\beta}^T \underline{Z}_i(t)}Y_i(t)$ the baseline intensities cancel in (5.7) and the resulting partial likelihood is

$$L(\underline{\beta}) = \prod_{i=1}^{n} \left\{ \frac{e^{\underline{\beta}^T \underline{Z}_i(t)}Y_i(t)}{\sum_{g \in \mathcal{R}(t)} e^{\underline{\beta}^T \underline{Z}_g(t)}Y_g(t)} \right\}^{dN_i(t)} \tag{5.8}$$

Detailed derivation of the partial likelihood for the Cox regression model can be found in Cox (1972, 1975, 1984). The properties of the estimator $\hat{\underline{\beta}}$ calculated from (5.8) can be derived from the fact that the model is a special case of the multiplicative intensity model, see Andersen and Gill (1982) for further details. In this model the baseline hazard is left unspecified, one possible estimate of the baseline hazard suggested by Breslow (1974) is a generalisation of the Nelson-Aalen estimator. When there are no covariates in the model, this estimate of the cumulative baseline hazard is, $\hat{\Lambda}_0(t)$, is

$$\int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}$$

When there are covariates in the model the estimate becomes

$$\int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)e^{\hat{\underline{\beta}}^T \underline{Z}_i(u)}}$$

where $\hat{\underline{\beta}}$ is derived from (5.8). As with the estimates for $\underline{\beta}$ Martingale calculus can be used to obtain first and second moments for the estimates of the cumulative baseline hazards. See Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1992) for further details.

We may wish to consider the case when there is more than one type of event. In this case the intensity under a Cox proportional hazards model becomes

$$\alpha_{ik}(t) = \lambda_{0k}(t)\, e^{\underline{\beta}_k^T \underline{Z}_{ik}(t)}\, Y_{ik}(t) \tag{5.9}$$

and the partial likelihood (5.8) becomes

$$L(\underline{\beta}_1, \cdots, \underline{\beta}_h) = \prod_{i=1}^{n} \prod_{k=1}^{h} \left\{ \frac{e^{\underline{\beta}_k^T \underline{Z}_{ik}(t) Y_{ik}(t)}}{\sum_{j \in \mathcal{R}(t)} e^{\underline{\beta}_k^T \underline{Z}_{jk}(t) Y_{jk}(t)}} \right\}^{dN_{ik}(t)} \tag{5.10}$$

The key point to notice is that the partial likelihood factorises into a number of components, one for each type of event. This factorisation of the partial likelihood when there is more than one type of event is equivalent defining time dependent strata by occurrences of types of events in standard survival analysis. See Prentice, Williams and Peterson (1981) for further details.

5.3.3 Parametric Intensity Model

When the times at which events occur follow a particular distribution the intensity will have a known parametric form. For example, if times to an event can be assumed to follow an exponential distribution, then the intensity function is constant over time. In this case $Y(t)$ is as before, a censoring indicator, but this time $\lambda(t)$ is also dependent on a vector of parameters $\underline{\theta}$. Therefore (5.4) becomes

$$\alpha_i(t) = \lambda[t|\underline{\theta}, \underline{Z}_i(t)] Y_i(t) \tag{5.11}$$

It may be that the covariates act multiplicatively, but rather than leaving the intensity unspecified it is of some parametric form, i.e

$$\alpha_i(t) = \lambda_0(t|\underline{\theta}) e^{\underline{\beta}^T \underline{Z}_i(t)} Y_i(t) \tag{5.12}$$

Kalbfleisch and Prentice (1980) and Cox (1984) derive the log-likelihood in the case when the intensity is fully parametric. This is as follows; if we consider patient i and whether or not they experience an event at time t. If the event does occur then patient i contributes $f_i(t)$ to the likelihood, i.e the probability of an event occurring at time t. If however the event does not occur and patient i is censored at time t then they contribute $S_i(t)$, i.e the probability that the event occurred for patient i at a time greater than t. If $dN_i(t)$ is an indicator of whether the event occurs for patient i at time t the contribution for the ith patient to the likelihood is

$$f_i(t)^{dN_i(t)}\, S_i(t)^{1-dN_i(t)}$$

Using the relationship between both the density, $f_i(t)$, and the survivor function, $S_i(t)$ and the intensity, $\alpha_i(t)$ that

$$\begin{aligned} f_i(t) &= \alpha_i(t)\, S_i(t) \\ S_i(t) &= e^{-\int_0^t \alpha_i(u)\, du} \end{aligned}$$

the contribution for the ith patient to the likelihood is, in terms of the intensity for the particular event

$$\alpha_i(t)^{dN_i(t)}\, e^{-\int_0^t \alpha_i(u)\, du}$$

Therefore the likelihood for all n patients is

$$L(\underline{\theta}, \underline{\beta}) = \prod_{i=1}^{n} \left\{ \prod_{t \geq 0} \alpha_i(t)^{dN_i(t)} \right\} e^{-\int_0^{t^*} \alpha_i(u)\, du}$$

where $t^* = sup\{t : dN_i(t) = 1, i = 1, \cdots, n\}$. The corresponding log-likelihood is

$$\ell(\underline{\theta}, \underline{\beta}) = \sum_{i=1}^{n} \int_0^{t^*} \log_e[\alpha_i(u)] dN_i(u) - \sum_{i=1}^{n} \int_0^{t^*} \alpha_i(u) du \tag{5.13}$$

where t^* is as above. Substituting (5.12) into (5.13) yields

$$\ell(\underline{\theta}, \underline{\beta}) = \sum_{i=1}^{n} \int_0^{t^*} \log_e[\lambda_0(u|\underline{\theta}) e^{\underline{\beta}^T \underline{Z}_i(u)}] dN_i(u) - \sum_{i=1}^{n} \int_0^{t^*} \lambda_0(u|\underline{\theta}) e^{\underline{\beta}^T \underline{Z}_i(u)} Y_i(u) du \tag{5.14}$$

Therefore the likelihood corresponding to (5.14) is

$$L(\underline{\theta}, \underline{\beta}) = \prod_{i=1}^{n} \prod_{t \geq 0} \frac{\left[\lambda_0(t|\underline{\theta}) e^{\underline{\beta}^T \underline{Z}_i(t)} Y_i(t)\right]^{dN_i(t)}}{\int_0^t \lambda_0(u|\underline{\theta}) e^{\underline{\beta}^T \underline{Z}_i(u)} Y_i(u) du} \tag{5.15}$$

In the rest of this thesis, we shall assume in likelihoods that the product over all time points is assumed, without stating it. We shall write $\int_0^t$ for the integral over all observed event times, rather than $\int_0^{t^*}$ as in (5.14) above.

In the fully parametric intensity case described above there are parallels with regression methods for categorical data. The simplest form of counting process is the Poisson process, in which the intensity is constant over time. Similarly, a piecewise constant intensity gives rise to a time inhomogeneous Poisson process. Lawless (1987) and Clayton (1988) have reviewed the link between survival models that consider time to each event (or between events), and models for counts, such as the Poisson model. Counting process notation highlights this close link between the two approaches.

As with the semi-parametric model above we may wish to allow for more than one type of event. If we consider patient i and whether or not they experience event k at time t. If event k does occur then patient i contributes $f_{ik}(t)$ to the likelihood, i.e the probability of event k occurring at time t. If however event k does not occur and patient i is censored at time t then they contribute $S_{ik}(t)$,

i.e the probability that event k occurred for patient i at a time greater than t. If $dN_{ik}(t)$ is an indicator of whether event k occurs for patient i at time t the contribution for the ith patient to the likelihood for the kth type of event

$$f_{ik}(t)^{dN_{ik}(t)}\, S_{ik}(t)^{1-dN_{ik}(t)}$$

Using the relationship between both the density, $f_{ik}(t)$, and the survivor function, $S_{ik}(t)$ and the intensity, $\alpha_{ik}(t)$ that

$$\begin{aligned} f_{ik}(t) &= \alpha_{ik}(t)\, S_{ik}(t) \\ S_{ik}(t) &= e^{-\int_0^t \alpha_{ik}(u)\, du} \end{aligned}$$

the contribution for the ith patient to the likelihood for the kth transition is

$$\alpha_{ik}(t)^{dN_{ik}(t)}\, e^{-\int_0^t \alpha_{ik}(u)\, du}$$

The overall likelihood now requires the product to be taken over both patients and events, and is

$$L(\underline{\phi}_1, \cdots, \underline{\phi}_h) = \prod_{i=1}^{n} \prod_{k=1}^{h} \left[\alpha_{ik}(t)^{dN_{ik}(t)}\, e^{-\int_0^t \alpha_{ik}(u)\, du}\right]$$

where $\underline{\phi}_k$ represents the model parameters for the kth type of event.

The log-likelihood (5.13) becomes

$$\begin{aligned} \ell(\underline{\theta}, \underline{\beta}) &= \sum_{i=1}^{n} \sum_{k=1}^{h} \log_e[\alpha_{ik}(t)] dN_{ik}(t) \\ &\quad - \sum_{i=1}^{n} \sum_{k=1}^{h} A_{ik}(t) \end{aligned} \tag{5.16}$$

where the intensity, $\alpha_{ik}(t)$ is

$$\alpha_{ik}(t) = \lambda_{0k}(t|\underline{\theta}_k) e^{\underline{\beta}_k^T \underline{Z}_{ik}(t)} Y_{ik}(t) \tag{5.17}$$

Substituting (5.17) into (5.16) yields a log-likelihood of the form

$$\begin{aligned} \ell(\underline{\theta}, \underline{\beta}) &= \sum_{i=1}^{n} \sum_{k=1}^{h} \log_e[\lambda_{0k}(t|\underline{\theta}_k) e^{\underline{\beta}_k^T \underline{Z}_{ik}(t)} Y_{ik}(t)] dN_{ik}(t) \\ &\quad - \sum_{i=1}^{n} \sum_{k=1}^{h} \int_0^t \lambda_{0k}(u|\underline{\theta}_k) e^{\underline{\beta}_k^T \underline{Z}_{ik}(u)} Y_{ik}(u) du \end{aligned} \tag{5.18}$$

and the corresponding likelihood is

$$L(\underline{\theta}, \underline{\beta}) = \prod_{i=1}^{n} \prod_{k=1}^{h} \frac{\left[\lambda_{0k}(t|\underline{\theta}_k) e^{\underline{\beta}_k^T \underline{Z}_{ik}(t)} Y_{ik}(t)\right]^{dN_{ik}(t)}}{\int_0^t \lambda_{0k}(u|\underline{\theta}_k) e^{\underline{\beta}_k^T \underline{Z}_{ik}(u)} Y_{ik}(u) du} \tag{5.19}$$

In both (5.12) and (5.17) there are various parametric forms that $\lambda_0(t|\underline{\theta})$ and $\lambda_{0k}(t|\underline{\theta}_k)$ may take. The simplest form is to let the function $\lambda_0(t)$ be a constant, analogous to a constant hazard in survival analysis. An extension to this model is to let $\lambda_0(t)$ be constant within intervals, with intervals either defined by the user or a function of the data. Breslow (1972) suggests that they should be the distinct death times, whilst Kalbfleisch and Prentice (1973) suggest that the time grid should be chosen independently from the data. Both of these parametric forms for the baseline hazard will be developed in much greater detail in subsequent chapters. An alternative parametric form that is frequently advocated for survival type data is the Weibull distribution, see Cox & Oakes (1984). Others have suggested various parametric forms such as the log-Normal or Gamma, see Cox & Oakes (1984) and Kalbfleisch & Prentice (1980) for further details.

Another approach which has also been advocated is the use of a suitably flexible function, such as a spline, to characterise the baseline hazard. The use of splines in statistics as a means of modelling data has been described at length by deBoor (1978) and Wegman & Wright (1983). The use of splines in survival models has been considered by Wahba (1978), Shaw (1988) and Durrleman and Simon (1989). Shaw (1988) has also consider them from a Bayesian perspective. Spline based models will be given brief consideration in Chapter 8.

5.4 Summary

In this chapter we have introduced counting process notation for describing event history data. For survival data we have seen how counting process notation is exactly analogous to traditional notation. However, counting process notation allows rigorous proof of many of the desirable properties of traditional survival models. It also provides a general framework within which models may be extended to the case in which a number of different types of events may each occur more than once.

In Chapter 6 we will consider Bayesian inference in the fully parametric multiplicative intensity model (5.12) when there is only one type of non-recurrent event. In Chapter 7 we will consider more complicated situations when there are a number of different events, some of which may occur more than once. Such situations can be accommodated by models such as (5.17).

Chapter 6

Two-State Models

6.1 Introduction

In this chapter we develop a simple Bayesian model for the situation in which there are only two possible states that a patient can be in, with non-reversible transitions between these states. The simplest example is a survival model, where the second state is death. During the last twenty years considerable work has been done in Bayesian survival analysis, and Abrams (1989) reviews much of this work.

In developing the models in this chapter we have borne in mind potential extensions to the multi-state case. As a result of this the models developed in this chapter are different in a number of respects to those models that have been developed by other workers in the area. The exclusive use of counting process notation described in Chapter 5 will also be another important part of our development as it is through this that we will see the most obvious ways in which the models may be extended to cope with more complicated scenarios.

The outline of this chapter is as follows; in Section 6.2 we consider the description of two state problems using counting process notation outlined in Chapter 5. In Section 6.3 we consider the general modelling approach when considering fully parametric multiplicative intensity models from a Bayesian perspective. Having considered the development of a model up to a joint posterior, we consider parameter estimation in Section 6.4. Two techniques are described in detail, Laplace approximations (Tierney and Kadane, 1986) and Gauss-Hermite quadrature (Naylor and Smith, 1982). The use of Gibbs sampling (Gelfand and Smith, 1990) is outlined although not required for these models. We also consider how credibility intervals and contours can be obtained. Section 6.5 describes in detail the case when the baseline intensity is constant over time. In this situation an exact solution exists when there are only two patient groups, and this simple model is applied to the neutron therapy data described in Chapter 2. We then consider extending the model to the case when there are several covariates, and have to use one of the parameter estimation procedures mentioned above. This model is applied to the neutron therapy data where there is both clinical prior information and results from previous studies. Sections 6.6 and 6.7 consider the case when the baseline intensity is either piecewise constant or has a Weibull parametric form. In both situations exact solutions do not exist even for two patient groups. Both models are again applied to the neutron therapy data of Chapter 2. Finally in Section 6.8 we consider further extensions of these fully parametric survival models.

6.2 Counting Processes Applied to Survival Data

A standard survival analysis problem is one of the simplest form of counting process. Assuming a homogeneous group of patients for whom we count the number that fail up to time t and the number at risk at time t, those patients

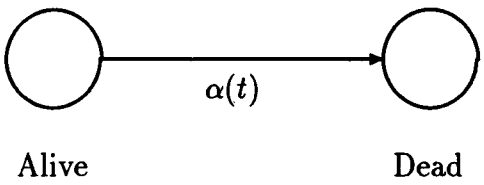


Figure 6.1: 2-State model for survival data.

who are *right censored* contribute only to the number at-risk process. A problem that may also arise is that termed *delayed-entry.* In this situation patients enter the study at some time point other than zero. This may occur when the time scale being used is other than trial time, e.g calendar time or age. In this case the at risk process would not decrease monotonically, but instead may increase at certain time points. Figure 6.1 shows this process diagrammatically, whilst Figures 5.3 and 5.4 show $N(t)$ and $Y(t)$ corresponding to $\alpha(t)$. In the case of death this is an absorbing state, as once a patient has entered they do not leave.

In the case of a simple two-state model for survival, each component of a multivariate counting process, $N_i(t)$, is a 0/1 indicator function; 0 if patient i was still alive at time t, and 1 otherwise. Similarly, the at-risk process, $Y_i(t)$, would also be a 0/1 indicator function; 0 if patient i was not at risk just prior to time t, and 1 otherwise. $N(t)$ and $Y(t)$ for survival data are shown in Figures 5.3 and 5.4 respectively. The intensity process, $\alpha_i(t)$, can be defined such that $\alpha_i(t)\,\delta t$ is the probability that patient i will die in a small interval δt after time t, conditional on everything that has happened up until time t. In this case we can see that the intensity $\alpha_i(t)$, assuming non-informative censoring, will either be the hazard function at time t if patient i has not failed before time t, or if patient i has died prior to time t then the intensity will be zero, therefore,

$$\alpha_i(t) = \lambda_i(t)\, Y_i(t)$$

where $\lambda(t)$ is the hazard function at time t.

Many different censoring mechanisms may be allowed for via the definition of $Y(t)$, but for the rest of this chapter we will assume that censoring is non-informative, i.e the intensity that we have observed is the same as that which would have observed whether or not we had observed the censoring distribution. Any relationship between the intensity and patient covariates can be modelled through $\lambda(t)$.

6.3 General Modelling Approach

In this section we consider a Bayesian version of the multiplicative intensity model in (5.12), in which the baseline intensity has a parametric form and there may be prior information about both the baseline parameters, $\underline{\theta}$, and the regression parameters, $\underline{\beta}$.

Suppose that each patient's survival history can be described by the triple $(N_i(t), Y_i(t), \underline{Z}_i(t))$ defined in Section 5.2.2, with $N_i(t)$ and $Y_i(t)$ defined as in Section 6.2 above for survival data. Following Aalen (1978) a parametric multiplicative model may be assumed such that

$$\alpha_i(t) = \lambda_0(t|\underline{\theta})\, e^{\underline{\beta}^T \underline{Z}_i(t)}\, Y_i(t) \quad i = 1, \cdots, n \tag{6.1}$$

where $\underline{\theta}$ is a vector of baseline intensity parameters and $\underline{Z}_i(t)$ is a vector of possible time-dependent covariates for the ith patient. Assume that prior information about $\underline{\theta}$ and $\underline{\beta}$ may be expressed independently in terms of probability density functions $p(\underline{\theta})$ and $p(\underline{\beta})$. Thus following the construction of the likelihood found in Cox and Oakes (1984), and described briefly in Section 5.3.3,

$$L(\underline{\phi}) = \prod_{i=1}^{n} [\alpha_i(t)]^{dN_i(t)} \left[e^{-\int_0^t \alpha_i(u)du} \right] \tag{6.2}$$

where $\underline{\phi} = (\underline{\theta}, \underline{\beta})$. Therefore the joint posterior, $p(\underline{\phi}|H)$ is proportional to the product of the likelihood, (6.2), and the prior density functions $p(\underline{\theta})$ and $p(\underline{\beta})$. The posterior, $p(\underline{\phi}|H)$, is written *conditional* on H, where H is used to denote all the information available about the parameters in question, both from the data, via the likelihood, and from any other prior information that may be available. Therefore, algebraically,

$$p(\underline{\phi}|H) \propto \prod_{i=1}^{n} [\alpha_i(t)]^{dN_i(t)} \left[e^{-\int_0^t \alpha_i(u)du} \right] p(\underline{\theta}) p(\underline{\beta}) \tag{6.3}$$

The joint posterior density is

$$p(\underline{\phi}|H) = \frac{\prod_{i=1}^{n} [\alpha_i(t)]^{dN_i(t)} \left[e^{-\int_0^t \alpha_i(u)du} \right] p(\underline{\theta}) p(\underline{\beta})}{\int_{\underline{\Theta}} \int_{\underline{\mathcal{B}}} \prod_{i=1}^{n} [\alpha_i(t)]^{dN_i(t)} \left[e^{-\int_0^t \alpha_i(u)du} \right] p(\underline{\theta}) p(\underline{\beta})\, d\underline{\theta} d\underline{\beta}} \tag{6.4}$$

where $\underline{\Theta}$ and $\underline{\mathcal{B}}$ denote the parameter spaces of $\underline{\theta}$ and $\underline{\beta}$ respectively.

Interest often focuses on one particular parameter. In clinical trials this may be a treatment effect. Therefore, we need to obtain the marginal posterior density for that particular parameter from the joint posterior density (6.4). For example suppose interest focuses on a regression parameter, β_1, then we need to integrate the joint posterior (6.4) with respect to all the parameters other than β_1 in order to obtain the marginal posterior density of β_1, i.e

$$p(\beta_1|H) \propto \int_{\underline{\Theta}} \int_{\underline{\mathcal{B}}_{-1}} p(\underline{\phi}|H)\, d\underline{\theta} d\underline{\beta}_{-1} \tag{6.5}$$

where $\underline{\Theta}$ represents the whole parameter space for $\underline{\theta}$, and $\underline{\mathcal{B}}_{-1}$ represents the parameter space for the elements of $\underline{\beta}$ other than β_1.

We may also be interested in the relationship between two particular parameters, and therefore the bivariate posterior density, say $p(\beta_1, \beta_2|H)$. As with the marginal posterior density, we need to integrate (6.4) with respect to remaining parameters in $\underline{\phi}$, i.e

$$p(\beta_1, \beta_2|H) \propto \int_{\underline{\Theta}} \int_{\underline{\mathcal{B}}_{-1,2}} p(\underline{\phi}|H)\, d\underline{\theta} d\underline{\beta}_{-1,2} \tag{6.6}$$

where $\mathcal{B}_{-1,2}$ denotes the parameter space for $\beta_3, \cdots, \beta_p$, p being the number of regression parameters in the model. As with the marginal densities, (6.6) may be standardised, by dividing by the corresponding integrating factor.

We may be interested in not only the parameters themselves but in some function of one or more of them, $h(\underline{\beta})$. We can calculate the posterior mean of $h(\underline{\beta})$ as

$$E[h(\underline{\beta})|H] = \int_{\underline{\Phi}} h(\underline{\beta}) p(\underline{\phi}|H)\, d\underline{\phi} \tag{6.7}$$

We will often be interested in the posterior mean of a particular parameter, say β_1, in which case $h(\underline{\beta}) = \beta_1$. If we are interested in the posterior variance of β_1 we require the mean of β_1^2 and therefore $h(\underline{\beta}) = \beta_1^2$. If we require the posterior correlation of two parameters β_1 and β_2 then $h(\underline{\beta}) = \beta_1 \beta_2$.

Much of the technical side of Bayesian statistics is concerned with evaluating integrals of the form (6.5) to (6.7). There are essentially three possible ways to evaluate or approximate such integrals. The first method is to approximate the joint posterior asymptotically, the second is to use numerical methods (quadrature) and the third method is to use simulation techniques. In the next section we will consider the use of all three of these techniques in evaluating integrals such as (6.5) to (6.7).

6.4 Parameter Estimation and Inference

Advances in computing technology over the last decade now mean that a wider range of methods is available for summarising posterior densities. In cases where analytic solutions do not exist there are three avenues open. The first method relies on asymptotic approximations to integrals. These often involve

assumptions of Normality or at least some of the properties of Normality such as unimodality. One particular method of asymptotic approximation to integrals is due to Laplace (Stigler, 1986), and termed Laplace approximations. This method is described in detail by De Bruijn (1958), and has been popularised in Bayesian integration problems by Tierney and Kadane (1986, 1989). The second method of integral evaluation is quadrature. Here integrals are approximated numerically. Thisted (1988) gives a review of quadrature generally, and more specifically the application of quadrature to Bayesian computations. One quadrature method, Gauss-Hermite quadrature, has been shown by Naylor and Smith (1982) to be a particularly efficient quadrature method for evaluating integrals that arise in Bayesian methodology. The third method of integral evaluation is simulation. Many of the methods that have found recent favour amongst applied statisticians, can be viewed as special case of Markov Chain Monte Carlo simulation. The most prolific, in terms of its usage, is Gibbs Sampling. This method was originally developed for use in image reconstruction, see Geman and Geman (1984), but has more recently been popularised by Gelfand and Smith (1990) and Gelfand, Hills, Racine-Poon and Smith (1990) in more general applied settings.

The structure of the rest of this section will be to outline each of the estimation methods mentioned above, and show how they can be applied to estimate marginal or bivariate densities for parameters of interest or posterior expectations of functions of these parameters. In subsequent sections of this chapter Laplace approximations and Gauss-Hermite quadrature will be used to estimate the parameters of a Bayesian two state multiplicative intensity model.

6.4.1 Laplace Approximations

Tierney & Kadane (1986) describe the use of Laplace's method of approximating integrals of the form $\int_0^\infty e^{nG(\phi)}d\theta$ in the limiting case when $n \to \infty$. Such approximations have been used widely in applied mathematics (Erdelyi, 1956, de Bruijn, 1958, Jeffreys, 1961, Stigler, 1986). This approximation relies upon the assumption that the integrand is *peaked* near its maximum. For a more detailed consideration of this technique see de Bruijn (1958).

By expanding $G(\phi)$ about $\hat{\phi}$, the mode of $G(\phi)$ and approximating $e^{-nG(\phi)}$ in the integrand by a function proportional to a Normal density determined by the second order Taylor series approximation to $G(\phi)$, the following approximation can be obtained

$$\begin{aligned} \int_0^\infty e^{nG(\phi)}\,d\phi &\approx \int \exp\left[nG(\hat{\phi}) - \frac{n(\phi-\hat{\phi})^2}{2\sigma^2}\right] d\phi \qquad (6.8) \\ &= \sqrt{\frac{2\pi\sigma^2}{n}}\, e^{nG(\hat{\phi})}[1+O(n^{-1})] \end{aligned}$$

If the integrals that need to be evaluated, in order to obtain either the marginal densities or general posterior moments, can be written in the necessary form then Laplace's method can be applied.

In order to use this method in the evaluation of integrals like (6.5) we need to consider the multivariate case, i.e replacing ϕ by $\underline{\phi}$. In the case of the parametric multiplicative intensity model (6.1) $\underline{\phi} = (\underline{\theta}, \underline{\beta})$. Let $\underline{\phi}$ have dimension q, and $\underline{\theta}$ have dimension s, with $\underline{\beta}$ therefore having dimension $q - s$. Consider the marginal posterior density for a particular element of $\underline{\phi}$, γ, and denote the other $q - 1$ elements by $\underline{\xi}$. The marginal posterior density for γ is of the form

$$p(\gamma|H) = \frac{\int_{\underline{\Xi}} p(\gamma, \underline{\xi}) e^{\ell(\gamma, \underline{\xi})} \, d\underline{\xi}}{\int_{\underline{\Phi}} p(\underline{\phi}) e^{\ell(\underline{\phi})} \, d\underline{\phi}} \tag{6.9}$$

where $\underline{\Xi}$ and $\underline{\Phi}$ represent the parameter spaces for $\underline{\xi}$ and $\underline{\phi}$ respectively. If the numerator and denominator of (6.9) can be considered separately, and each can be written in the appropriate form, each may be approximated using Laplace's method (6.9). The result of (6.9) then needs to be extended to the multivariate case, and the marginal posterior density can then be approximated.

The multivariate equivalent of approximation (6.9) is given by de Bruijn as

$$\int_{\underline{\Phi}} e^{nG(\underline{\phi})} \, d\underline{\phi} = \left(\frac{2\pi}{n}\right)^{1/2q} \left| \frac{\partial^2 G(\underline{\phi})}{\partial \underline{\phi}^2} \right|_{\underline{\phi}=\hat{\underline{\phi}}}^{-1/2} e^{nG(\hat{\underline{\phi}})} \tag{6.10}$$

where as in the univariate case $\hat{\underline{\phi}}$ is the value of $\underline{\phi}$ which maximises $G(\underline{\phi})$, and $|\cdot|$ represents the determinant of a matrix. We are now able to evaluate the integrals separately in the numerator and denominator of (6.9). Considering the integrand in the numerator, $p(\gamma, \underline{\xi}) e^{\ell(\gamma, \underline{\xi})}$, this can be written as

$$e^{n[\log_e[p(\gamma, \underline{\xi})]/n + \ell(\gamma, \underline{\xi})/n]}$$

and is denoted $e^{nG^*(\underline{\xi})}$. Similarly, the denominator, $p(\underline{\phi}) e^{\ell(\underline{\phi})}$, can be written as

$$e^{n[\log_e[p(\underline{\phi})]/n + \ell(\underline{\phi})/n]}$$

and denoted $e^{nG(\underline{\phi})}$. Note that the $G^*(\underline{\xi})$ is a function of only $q - 1$ parameters, whilst $G(\underline{\phi})$ is a function of q parameters, as we consider evaluating the marginal density for specific values of ξ. We can now apply the multivariate approximation (6.10) to both integrals. Therefore

$$\begin{aligned} \frac{\int_{\underline{\Xi}} p(\gamma, \underline{\xi}) e^{\ell(\gamma, \underline{\xi})} \, d\underline{\xi}}{\int_{\underline{\Phi}} p(\underline{\phi}) e^{\ell(\underline{\phi})} \, d\underline{\phi}} &= \frac{\left(\frac{2\pi}{n}\right)^{1/2(q-1)} \left| \frac{\partial^2 G^*(\underline{\xi})}{\partial \underline{\xi}^2} \right|_{\underline{\xi}=\hat{\underline{\xi}}}^{-1/2} e^{nG^*(\hat{\underline{\xi}})}}{\left(\frac{2\pi}{n}\right)^{1/2q} \left| \frac{\partial^2 G(\underline{\phi})}{\partial \underline{\phi}^2} \right|_{\underline{\phi}=\hat{\underline{\phi}}}^{-1/2} e^{nG(\hat{\underline{\phi}})}} \\ &= \left(\frac{\left| \frac{\partial^2 G^*(\underline{\xi})}{\partial \underline{\xi}^2} \right|_{\underline{\xi}=\hat{\underline{\xi}}}}{\frac{2\pi}{n} \left| \frac{\partial^2 G(\underline{\phi})}{\partial \underline{\phi}^2} \right|_{\underline{\phi}=\hat{\underline{\phi}}}} \right)^{-1/2} e^{n[G^*(\hat{\underline{\xi}}) - G(\hat{\underline{\phi}})]} \end{aligned} \tag{6.11}$$

Tierney and Kadane (1986) investigate the accuracy of the approximation in (6.11), and show that

$$p(\gamma|H) = \tilde{p}(\gamma|H)(1 + O(n^{-1})) \tag{6.12}$$

where $\tilde{p}(\gamma|H)$ is the marginal posterior density obtained using approximation (6.11). The main reason for the error in (6.12) being as large as $O(n^{-1})$ is that the dimensionalities of the two integrals in the numerator and denominator of (6.9) are different. The main consequence is that the estimate for $p(\gamma|H)$ obtained from (6.11), $\tilde{p}(\gamma|H)$, needs to be re-normalised using numerical integration. This can be done simply using a Trapezoidal Rule. This rule approximates the area under a function between successive points as a trapezium. For a linear function this method is exact, but when the function is not linear an error is introduced. This error is determined not only by the curvature of the function between successive points, but also by the number of points that are used. See Thisted (1988) page 264 for further details.

Similarly if the posterior expectation of a function $g(\underline{\phi})$ were required the resulting ratio of integrals would be;

$$E\left[g(\underline{\phi})\right] = \frac{\int_{\underline{\Phi}} g(\underline{\phi})p(\underline{\phi})e^{\ell(\underline{\phi})}\,\partial\underline{\phi}}{\int_{\underline{\Phi}} p(\underline{\phi})e^{\ell(\underline{\phi})}\,\partial\underline{\phi}} \tag{6.13}$$

which can be re-expressed in a suitable form, i.e

$$G^*(\underline{\phi}) = (\log_e[g(\underline{\phi})] + \log_e[p(\underline{\phi})] + \ell(\underline{\phi}))/n$$

Thus, $g(\underline{\phi})$ is required to be wholly positive as the logarithm of $g(\underline{\phi})$ is used.

Tierney and Kadane (1986) only describe the case when $g(\underline{\phi})$ is either a wholly positive or a wholly negative function. In many regression situations a parameter may be near zero. In this case the above method would not be valid, and Tierney *et al* (1989) suggest two possible solutions to this problem;

1. The first method requires us to work with the moment generating functions rather than the required functions directly. This method has the disadvantage that it introduces an unknown constant, the moment generating constant, into the problem, and maximisation then needs to be done analytically.

2. The second method simply requires that a suitably large constant is added to, or subtracted from, the parameters, and estimation proceeds as above, with the additive property of expectation utilised to obtain estimates of the original function.

It is the second approach that is taken in this thesis. This may still be unstable, and even yield negative estimates for the variance of the parameter of interest. In this case there is another possible way to estimate the mean and variance of a parameter of interest, though regularity conditions need to

checked. We may be interested in the mean and variance of γ, having obtained the posterior marginal density of γ, $\tilde{p}(\gamma|H)$, using (6.11) above. We can then estimate the posterior mean of γ by using the approximation,

$$E(\gamma|H) \approx \sum_{\Gamma} \gamma \tilde{p}(\gamma|H)$$

where Γ is the parameter space for γ. Tierney *et al* (1989) have shown that this approximation has an error $O(n^{-1/2})$. In order that we may be able to obtain an estimate of the posterior variance of γ, we require an estimate of the posterior expectation of γ^2. Thus,

$$E(\gamma^2|H) \approx \sum_{\Gamma} \gamma^2 \tilde{p}(\gamma|H)$$

If we are to use this estimate then we need to check that the posterior density $\tilde{p}(\gamma|H)$ decreases at a faster rate than does $1/\gamma^2$. If the functional form of $\tilde{p}(\gamma|H)$ is not known then one way of checking this assumption is to plot $\tilde{p}(\gamma|H)$ against $1/(\gamma - \hat{\gamma})^2$, where $\hat{\gamma}$ is the posterior mode.

Hills and Smith (1992) have noted that the Laplace approximations of Tierney and Kadane outlined in this section are not invariant to non-linear transformations of the model parameters. They also note that the method will perform better when the joint posterior density is close to being a multivariate Normal density. In the case when this is not true, a re-parameterisation of the model should be considered. The aspect of parameterisation with Laplace approximations has also been considered by Achcar and Smith (1990).

6.4.2 Gauss-Hermite Quadrature

Gauss-Hermite quadrature is a form of numerical integration considered for parameter estimation in a Bayesian setting by Naylor and Smith (1982), Smith, Skene, Shaw, Naylor and Dransfield (1985), Smith, Skene, Shaw and Naylor (1987) and Dellaportas and Wright (1991b). If we are interested in the indefinite integral of a function $g(\phi)$, and there exist μ and σ such that,

$$g(\phi) = h(\phi) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\phi-\mu}{\sigma}\right)^2}$$

where $h(\phi)$ is a suitably regular function, e.g a polynomial. Naylor and Smith (1982) show that

$$\begin{aligned} \int_{-\infty}^{\infty} g(\phi)\, d\phi &\approx \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} w_i\, h(\mu + \sqrt{2}\sigma x_i) \qquad (6.14) \\ &= \sum_{i=1}^{n} m_i\, g(z_i) \end{aligned}$$

where

$$\begin{aligned} m_i &= w_i e^{x_i^2}\sqrt{2\sigma} \\ w_i &= \frac{2^{n-1}n!\sqrt{\pi}}{n^2\left(H_{n-1}(x_i)\right)^2} \\ z_i &= \mu + \sqrt{2}\sigma x_i \end{aligned}$$

and x_i is the ith solution of the Hermite polynomial equation $H_n(\phi) = 0$. $H_n(\phi)$ is the coefficient of $\frac{t^r}{r!}$ in $\exp(t\phi - 0.5t^2)$. Naylor and Smith (1982) state that the error term in the approximation in (6.14) will be small if $h(\cdot)$ is approximately a polynomial.

If we consider the multivariate case then the simplest extension of (6.14) is, using the Cartesian product rule,

$$\int_{\Phi} g(\underline{\phi})\,\partial\underline{\phi} = \sum_{i_k} m_{i_k}^{(k)} \cdots \sum_{i_1} m_{i_1}^{(1)}\, g(\phi_{i_1}^{(1)}, \cdots, \phi_{i_k}^{(k)}) \tag{6.15}$$

where $m_{i_j}^{(j)}$ and $z_{i_j}^{(j)}$ are found using (6.15), and μ and σ are replaced with the current mean and variance of ϕ_i.

Naylor and Smith (1982) note that the product form of (6.15) can only be justified if there is no posterior correlation between the elements of $\underline{\phi}$, which in most situations would appear to be unrealistic. To overcome this problem they propose to introduce a transformation of $\underline{\phi}$ to an orthogonal set of parameters and work with these.

Convergence of this iterative procedure can be monitored in a number of ways. Naylor and Shaw (1985) suggest using the following measures. Let $p(\underline{\phi}|H)$ to be the normalising constant defined by

$$p(\underline{\phi}|H) = \int_{\underline{\Phi}} L(\underline{\phi})\, p(\underline{\phi})\,\partial\underline{\phi}$$

Let $p^1(\underline{\phi}|H)$ and $p^0(\underline{\phi}|H)$ be the current and immediately previous estimates of the normalising constant. Then the observed relative change in these estimates (shown as a percentage) is $\Delta p(\underline{\phi}|H)$ is defined

$$\Delta p(\underline{\phi}|H) = \frac{p^1(\underline{\phi}|H) - p^0(\underline{\phi}|H)}{p^0(\underline{\phi}|H)}$$

Consider the jth component of $\underline{\phi}$ and let μ_j^1 and μ_j^0 be current and immediately previous estimates of

$$\mu_j = \int_{\underline{\Phi}} \phi_j\, p(\underline{\phi}|H)\,\partial\underline{\phi}$$

Let σ_j^1 and σ_j^0 be estimates of

$$\sigma_j = \sqrt{\int_{\underline{\Phi}} (\phi_j - \mu_j)^2 p(\underline{\phi}|H)\,\partial\underline{\phi}}$$

We can then define the measures $\Delta\mu_j$ and $\Delta\sigma_j$ as

$$\Delta\mu_j = \frac{\mu_j^1 - \mu_j^0}{\sigma_j^0}$$

and

$$\Delta\sigma_j = \frac{\sigma_j^1 - \sigma_j^0}{\sigma_j^0}$$

respectively. As with the means and standard deviations we can define a measure for the correlations, ρ_{ij}. Let ρ_{ij}^1 and ρ_{ij}^0 be the current and immediately previous estimates of the correlation. The measure $\Delta\rho_{ij}$ is defined by

$$\Delta\rho_{ij} = \rho_{ij}^1 - \rho_{ij}^0$$

Naylor and Shaw (1985) suggest that a simple overall measure of convergence on the current iteration, Δ, can be defined as follows

$$\Delta = \left|\Delta p(\underline{\phi}|H)\right| + \frac{1}{m}\sum_{j=1}^{k}|\Delta\mu_j| + \frac{1}{m}\sum_{j=1}^{k}|\Delta\sigma_j| + \frac{1}{m^2}\sum_{i=2}^{k}\sum_{j=1}^{i-1}|\Delta\rho_{ij}| \qquad (6.16)$$

where m is the number of elements of $\underline{\phi}$ for which moments were calculated and k is the total number of parameters in the problem. Often m and k will coincide.

A software package, called BAYES4, has been developed at Nottingham University, Naylor & Shaw (1985), that will perform Gauss-Hermite quadrature. It is written in FORTRAN and was run on a SUN 4 Sparc station. It only requires the functional form of the log-likelihood and the joint prior density, together with parameter starting values.

Orthogonalising transformation of the original model parameters is often desirable since the performance of Gauss-Hermite quadrature is enhanced if the joint posterior density is approximately a multivariate Normal density. This aspect of the method has been noted by Hills and Smith (1992). Such reparameterisation is a standard feature of the BAYES4 package.

6.4.3 Simulation Techniques

As an alternative to the two methods outlined above, when the joint posterior cannot be well approximated by the product of a Normal distribution and a polynomial, or when it is not uni-modal, a number of simulation techniques have been proposed. The first idea of simulation is to simulate from the joint posterior distribution. In some circumstances this may be difficult, especially if the joint distribution is of a high dimension. A solution to this problem is to decompose the joint distribution into a number of conditional distributions each of which will have a lower dimension and be easier to simulate from than the joint posterior distribution. Two such methods are stochastic substitution sampling (Tanner and Wong, 1987) and Gibbs sampling, both of which are iterative Monte

Carlo procedures. The latter is a special case of Metropolis-Hastings algorithm (Besag and Green, 1993). Both methods are applicable when the joint posterior density can be expressed in terms of the full set of conditional densities, and random samples from the conditional densities may be generated. When the functional form of the joint posterior density is known and at least one of the conditional densities is available a non-iterative importance sampling algorithm has been proposed by Rubin (1987).

Gibbs sampling requires that each of the conditional distributions is sampled from, with the current values of the parameters being updated at each iteration. Eventually the algorithm will converge to the marginal distributions. Gibbs sampling was first suggested by Geman and Geman (1984) for use in image reconstruction, but more recently Gelfand and Smith (1990) have advocated the use of the Gibbs sampler for other statistical problems and Gelfand, Hills, Racine-Poon and Smith (1990) have consider the case of Normal models. Clayton (1991) has considered the case of inference in frailty models using the Gibbs sampler. Gilks, Clayton, Spiegelhalter, Best, McNeil, Sharples and Kirby (1993) describe the use of Gibbs sampling in a range of medical applications.

In some situations random samples have to be generated from distributions for which only the functional form is known. Ripley (1987) suggests an *acceptance-rejection* criterion based on sampling from a ratio of Uniform distributions. The acceptance criterion can sometimes be very strict resulting in a computationally inefficient method. Gilks and Wild (1992) have proposed a more efficient method for the case when the log-likelihood or the posterior is log-concave. This can be shown to be true for a number of models.

Simulation techniques appear to be appealing, especially in cases when the joint posterior distribution is not well behaved. However, a number of problems remain. The ease with which the random samples may be obtained. The ability of simulation techniques to cope with multi modality is not clear. Related to this in the case of the sampling algorithms is the question of convergence. As with Laplace approximations and Gauss-Hermite quadrature Hills and Smith (1992) have shown that 'good' parameterisation of models is necessary for simulation methods to work efficiently.

In this thesis we only use the first two integration techniques, Laplace approximations and Gauss-Hermite quadrature. Initial work on implementing both of these methods showed that they were computationally efficient for relatively well behaved problems. At the same time Gibbs sampling proved to be computationally inefficient due to the nature of the posterior conditional densities, caused by the presense of censoring. It was not until the papers of Kuo and Smith (1992) and Smith and Roberts (1993) that a method for over-coming these difficulties was available. Using these methods the censored observations are treated as additional unknown parameters, and a new set of conditional distributions are derived. Sampling from this new set of conditional distributions is then usually straight forward and does not require sophisticated sampling algorithms. Gibbs sampling is unecessarily complex for the parametric models considered in this thesis, though it does have a role to play in the parameter

estimation in semi-parametric models (Clayton, 1991).

6.4.4 Maximum Relative Likelihood

Before considering models in detail we mention non-Bayesian methods of estimation and in particular maximum relative likelihood (or log-likelihood) (MRL), as we will compare parameter estimates using Bayesian methodology with those using MRL. Assuming that we have a likelihood for the data $L(\underline{\phi})$ and a corresponding log-likelihood, $\ell(\underline{\phi})$, then the maximum relative likelihood, $R(\gamma)$, for a particular parameter, γ, is defined by

$$R(\gamma) = \frac{L(\gamma, \hat{\underline{\xi}})}{L(\hat{\underline{\phi}})}$$

where $\hat{\underline{\phi}}$ is the value of $\underline{\phi}$ which maximises $L(\underline{\phi})$, and $\hat{\underline{\xi}}$ is the value of $\underline{\xi}$, the $q-1$ parameters other than γ, that maximises $L(\underline{\phi})$ when γ is assumed fixed. Similarly the maximum relative log-likelihood, $R^*(\gamma)$, is defined by

$$R^*(\gamma) = \ell(\gamma, \hat{\underline{\xi}}) - \ell(\hat{\underline{\phi}})$$

The set of values of γ for which $R(\gamma) \geq p$ is called the $100p\%$ likelihood region for γ. Similarly, we can use the maximum relative log-likelihood to obtain intervals for γ, i.e $R^*(\gamma) \geq \log_e p$. Kalbfleisch (1985) describes its use in detail for a number of applied settings.

The ideas of maximum relative likelihood and log-likelihood can be extended to the bivariate case. Suppose that the two parameters of interest are γ and δ, with $\underline{\xi}$ now representing the $q-2$ remaining parameters. The maximum relative likelihood is defined by $R(\gamma, \delta) = L(\gamma, \delta, \hat{\underline{\xi}})/L(\hat{\underline{\phi}})$. Similarly the maximum relative log-likelihood is defined by $R^*(\gamma, \delta) = \ell(\gamma, \delta, \hat{\underline{\xi}}) - \ell(\hat{\underline{\phi}})$. As in the univariate case if we are interested in obtaining contours, analogous to intervals, for a set of parameter values then we can use the fact that this function will have a χ^2 distribution on 2 degrees of freedom, and we can calculate the height at which specific contours will need to be drawn, say a 95% or 90% contour.

6.4.5 Approximate Clinical Inference

In the univariate case inference can proceed once either the posterior marginal densities or the posterior moments have been calculated. We may particularly want to obtain the probability of a parameter being either in an interval or being less than a specific value. Both of these cases are relevant to clinical trials. We saw in Section 3.2 that clinicians can be asked to specify both a *clinical belief* and a *clinical demand.* Combined with the actual parameter value that indicates no treatment difference, say 1 in the case of the odds ratio, the *clinical demand* forms an interval in which the treatments are deemed to be equivalent. This concept has been termed *equipoise* in Section 3.4 and by Freedman (1987).

Obviously we are interested in the probability of the treatment difference lying inside or outside this interval. Having obtained the marginal posterior density this probability may be calculated via numerical integration. The simplest method to use is the Trapezoidal Rule, described briefly in Section 6.4.1.

Inference in the bivariate case is not as straight forward as in the univariate case. We can calculate quantities such as the posterior correlation, using Laplace approximations or Gauss-Hermite quadrature. Often we are interested in either the probability that the parameters take values in a certain region, e.g that $\beta_1 < 1$ *and* $\beta_2 < 1$, or we would like to know the region in which there is $100p\%$ probability of the parameters lying.

As with the univariate case, each of these problems may be addressed using numerical integration, i.e a bivariate version of the Trapezoidal Rule. However, in the first case when we are calculating the probability of a parameter lying in such a region this method would require considerable computation in order to achieve sufficient accuracy. For the second problem, calculating the region in which there is $100p\%$ probability of the parameters lying, considerable computation is again required, using a 'search type' algorithm.

A solution to the second type of problem is to use some form of approximation method. One method that could be employed in order to obtain contour regions is to calculate the height that a specific percentage contour is from the maximum of a standard bivariate Normal density, assuming the same correlation as the estimated bivariate density. An approximate contour can then be obtained by either dropping down this height from the maximum of the estimated bivariate posterior density or by using the ratio of the height of the contour to the height at the maximum.

For example, if we are interested in the estimated bivariate posterior density for $\underline{\phi}$, denoted $f_{\underline{\phi}}(\underline{x})$. If we assume that $f_{\underline{\phi}}(\underline{x})$ is centred at the origin for convenience, and that the covariance matrix for the density is Σ. If $f_{\underline{\phi}}(\underline{x})$ was a bivariate Normal density it would have the form

$$\frac{1}{2\pi\,|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\underline{x}^T\Sigma^{-1}\underline{x}}$$

In the case of the bivariate Normal we know that $\underline{x}^T\Sigma^{-1}\underline{x}$ has a χ^2 distribution on 2 degrees of freedom. Therefore for a 95% region

$$\begin{aligned} \int_0^a \frac{1}{2} e^{-u/2} du &= 0.95 \\ -e^{-a/2} + 1 &= 0.95 \\ a &= -2\log_e(0.05) \end{aligned}$$

In this case $a = \underline{x}^T\Sigma^{-1}\underline{x}$. In order that we may obtain a point on the ellipse, we need to obtain the corresponding co-ordinates. For this we need to expand $\underline{x}^T\Sigma^{-1}\underline{x}$, which in the case when $\underline{\phi} = (\phi_1, \phi_2)$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

becomes $\phi_1^2\sigma_2^2 - 2\rho\phi_1\phi_2\sigma_1\sigma_2 + \phi_2^2\sigma_1^2$. Therefore, if we required the 95% contour

$$\phi_1^2\sigma_2^2 - 2\rho\phi_1\phi_2\sigma_1\sigma_2 + \phi_2^2\sigma_1^2 = -2\log_e(0.05) \tag{6.17}$$

Since we have assumed that the density is centred at the origin, we could find the co-ordinates of the point (ϕ_1, ϕ_2) on the 95% ellipse, by setting ϕ_1 equal to zero in the above expression (6.17), and noticing that if $\phi_1 = 0$ this reduces to $\phi_2^2\sigma_1^2 = -2\log_e(0.05)$. We now know that the point $\left(0, \frac{\sqrt{-2\log_e(0.05)}}{\sigma_1}\right)$ lies on the ellipse, and we can obtain the value of the bivariate Normal density at this point, and in turn calculate the difference between the density at this point and the value of the density at the maximum value. Having found this difference, we can then use this difference to obtain the height of the approximate 95% posterior region for $f_{\underline{\phi}}(\underline{x})$, by finding the maximum value of $f_{\underline{\phi}}(\underline{x})$ and dropping down the same height.

6.5 Constant intensity model

In this section we consider the case when the baseline intensity, $\lambda_0(t)$ is constant over time. In the situation when there is only one homogeneous group of patients a conjugate analysis is possible. This 'exact' analysis can be extended to the case when there are two groups, but it is not practical to consider the case when the group of patients is stratified any further. In this situation a more general regression framework is required.

6.5.1 Inference for a single univariate counting process

Abrams (1989), Chapter 3, reviews a number of approaches that have been taken when there is only one group of patients. Both Cox and Oakes (1984) and Martz and Waller (1982) consider the case when the hazard rate is assumed to be constant over time, and a conjugate Gamma prior distribution is used to represent prior information. Martz and Waller (1982) also consider the case when a Uniform distribution is used to represent vague prior information.

Consider a group of n patients indexed by i, for whom we observe either the actual survival time, if they die, or a censoring time if they do not. Therefore the groups survival history may be described by a multivariate counting process $N(t) = \{N_i(t), i \in I, t > 0\}$, which records whether patient i has died or not by time t. Associated with $N(t)$ is an at-risk process $Y(t) = \{Y_i(t), i \in I, t > 0\}$ whose components are one if patient i is at risk of death at time t and zero otherwise. Now $N(t)$ is governed by its random intensity process $\alpha(t) = \{\alpha_i(t), \in I, t > 0\}$. In Section 6.2 we saw that if we assume non-informative censoring and the patients are considered to be independent, then $\alpha_i(t) = \lambda_i(t)Y_i(t)$, where $\lambda_i(t)$ can be interpreted as the hazard for death.

If we now assume that the group of patients can be considered to be homogeneous, and therefore all n patients have the same hazard rate λ, we can construct a univariate counting process. Thus,

$$N(t) = \sum_{i=1}^{n} N_i(t)$$

and

$$\alpha(t) = \sum_{i=1}^{n} \alpha_i(t) = \lambda Y(t)$$

where

$$Y(t) = \sum_{i=1}^{n} Y_i(t)$$

Therefore from Section 5.3.3 we obtain the log-likelihood as

$$\ell(\lambda) = \log_e(\alpha)dN(t) - \lambda \int_0^t Y(u)du \tag{6.18}$$

and the likelihood as

$$L(\lambda) = \lambda^{dN(t)} e^{-\lambda \int_0^t Y(u)du} \tag{6.19}$$

Maximum likelihood estimation can proceed using (6.18) or (6.19). Differentiating (6.18) and setting the derivative to zero,

$$\frac{d\ell(\lambda)}{d\lambda} = \frac{1}{\lambda} dN(t) - \int_0^t Y(u)du = 0$$

thus $\hat{\lambda} = dN(t)/\int_0^t Y(u)du$, i.e the number of deaths per total time spent at risk.

Given the form of the likelihood in (6.19) we can consider putting a conjugate prior on λ as Cox and Oakes (1984) and Martz and Waller (1982) have suggested. Such a conjugate prior has a Gamma distribution, (Cox and Oakes, 1984). Aven (1986) and Clayton (1991) have also noted that the conjugate prior density for the intensity process of a counting process is the Gamma density. Consider the density function for a Gamma distribution, with parameters η and ζ.

$$p(\lambda) = \frac{\eta^{\zeta}\lambda^{\zeta-1}e^{-\eta\lambda}}{\Gamma(\zeta)} \tag{6.20}$$

and where the prior mean of λ, $E(\lambda)$, is ζ/η, and the prior variance, $V(\lambda)$, is ζ/η^2.

Appling Bayes' theorem to the above likelihood (6.19) and prior (6.20) we obtain the posterior for λ, $p(\lambda|H)$, as

$$p(\lambda|H) \propto \lambda^{\zeta+dN(t)-1}e^{-\lambda(\eta+\int_0^t Y(u)\,du)} \tag{6.21}$$

where the posterior mean and variance for λ are $(\zeta + dN(t))/(\eta + \int_0^t Y(u)du)$ and $(\zeta + dN(t))/(\eta + \int_0^t Y(u)du)^2$ respectively. ζ^* and η^* are the parameters of the posterior Gamma distribution, i.e $\zeta^* = \zeta + dN(t)$ and $\eta^* = \eta + \int_0^t Y(u)du$. Notice that if $\eta = \zeta = 0$ then the posterior estimates for the intensity reduce to the maximum likelihood estimates of $dN(t)/\int_0^t Y(u)du$ for the mean and $dN(t)/(\int_0^t Y(u)du)^2$ for the variance.

In this limiting case when $\eta = \zeta = 0$, limiting because the Gamma distribution is only defined when both η and ζ are greater than zero, the prior density is the same as the one we would have obtained had we adopted Jeffreys' prior, see Jeffreys (1961). Jeffreys suggested using minus the expected information matrix as a vague prior density. Using (6.19) we have that

$$\begin{aligned} -E_{\underline{y}|\lambda}\left[\frac{d^2\ell(\lambda)}{d\lambda^2}\right] &= -\frac{1}{\lambda^2} \quad E_{\underline{y}|\lambda}[dN(t)] \\ &= \frac{1}{\lambda^2} \quad \lambda \\ &= \frac{1}{\lambda} \end{aligned}$$

which is the prior distribution using $\eta = \zeta = 0$, and places infinite mass at zero. This exposition assumes that $E_{\underline{y}|\lambda}[dN(t)]$ is the intensity if the time interval we are considering is an interval of length 1, but any time interval can be transformed to a [0,1] interval. Tiao and Box (1967) suggested a method for forming a non-informative prior which is also proportional to $1/\lambda$.

6.5.2 Exact and approximate inference for two univariate counting processes

Now consider the case when each of the n patients can be thought of as falling into one of two groups. The occurrences of death in the two groups can be described by two separate counting processes. The model derivation is in terms of the neutron therapy example described in Section 2.2, where there are two treatment groups, neutrons and photons.

Suppose that we assume the intensity in each of these two groups to be constant but different from one another. Let the intensity for neutron patients be α_n and that for photon patients be α_p. The quantity of interest is the ratio of the two intensities, ψ, i.e $\psi = \alpha_n/\alpha_p$. Under the assumption of non-informative censoring is equivalent to the ratio of the hazard rates in both groups, i.e $\psi = \lambda_n/\lambda_p$. As in the case of a single univariate counting process described in Section 6.5.1, using a conjugate Gamma prior for each of the intensities separately yields corresponding Gamma posterior densities. We can obtain the posterior distribution of ψ as the ratio of two Gamma distributions. This distribution, may either be obtained exactly or it may be approximated using a Normal approximation.

Consider just the photon group,

$$\begin{aligned} \lambda_p &\sim Gamma(\eta_p^*, \zeta_p^*) \\ 2\eta_p^*\lambda_p &\sim Gamma(0.5, \zeta_p^*) \\ 2\eta_p^*\lambda_p &\sim \chi^2_{2\zeta_p^*} \end{aligned}$$

Similarly,

$$2\eta_n^*\lambda_n \sim \chi^2_{2\zeta_n^*}$$

Therefore we have

$$\frac{\eta_n^*}{\eta_p^*}\frac{\lambda_n/2\zeta_n^*}{\lambda_p/2\zeta_p^*} \sim F(2\zeta_n^*, 2\zeta_p^*) \tag{6.22}$$

We are interested in the distribution of the hazard ratio, and there are two possible methods. We can directly obtain the distribution for the ratio or we could use an approximation. First consider the direct method.

If a random variable, X, has a density function, $f(x)$, and we require the density function of a new random variable, Y, defined by the transformation, $Y = r(X)$, then defining $X = s(Y)$, the density for Y, $g(y)$, is given by Cox and Hinkley (1974) as

$$g(y) = f(s(y))\left|\frac{ds(y)}{dy}\right|$$

In this case the transformation from X to Y is only a multiplication by a constant, $\eta_p^*\zeta_n^*/\eta_n^*\zeta_p^*$, and therefore

$$\left|\frac{ds(y)}{dy}\right| = \frac{\eta_n^*\zeta_p^*}{\eta_p^*\zeta_n^*}$$

So if $\eta_n^*\zeta_p^*\lambda_n/\eta_p^*\zeta_n^*\lambda_p$ has an F distribution, with density $f(.)$ and the relevant degrees of freedom, the posterior density of λ_n/λ_p, $g(.)$, is

$$f\left(x\frac{\eta_n^*\zeta_p^*}{\eta_p^*\zeta_n^*}\right)\frac{\eta_n^*\zeta_p^*}{\eta_p^*\zeta_n^*}$$

which can be evaluated for various x.

Lindley (1964, 1965) suggests that in the case when $2\zeta_n^*$ and $2\zeta_p^*$ are large, i.e greater than 20, then an approximation may be used for the F distribution. If such an approximation is valid then this yields a computationally simple procedure for inference. If a random variable, X has an F distribution with v and w degrees of freedom, then $0.5\log(X)$ is approximately Normally distributed with mean $0.5(1/w-1/v)$ and variance $0.5(1/w+1/v)$. Using this approximation we can obtain an approximate distribution for the log of the hazard ratio. Thus,

$$\frac{1}{2}\log_e\left(\frac{\eta_n^*}{\eta_p^*}\frac{\lambda_n/2\zeta_n^*}{\lambda_p/2\zeta_p^*}\right) \dot{\sim} N\left[\frac{1}{4\zeta_p^*}-\frac{1}{4\zeta_n^*},\frac{1}{4\zeta_p^*}+\frac{1}{4\zeta_n^*}\right] \tag{6.23}$$

Therefore after re-arranging we obtain an approximate Normal distribution for the log-hazard ratio,

$$\log_e\left(\frac{\lambda_n}{\lambda_p}\right) \dot{\sim} N\left[\frac{1}{2\zeta_p^*}-\frac{1}{2\zeta_n^*}-\log\left(\frac{\eta_n^*\zeta_p^*}{\eta_p^*\zeta_n^*}\right),\frac{1}{\zeta_p^*}+\frac{1}{\zeta_n^*}\right] \tag{6.24}$$

Therefore the ratio of the intensities, λ_n/λ_p, has an approximate Log-Normal distribution.

We can again consider the case when we assume vague prior information for λ_p and λ_n, i.e $\eta_p = \zeta_p = \eta_n = \zeta_n = 0$. From Section 6.5.1 $\eta_p^* = dN_p(t)$, $\zeta_p^* = \int_0^t Y_p(u)du$, $\eta_n^* = dN_n(t)$ and $\zeta_n^* = \int_0^t Y_n(u)du$. Substituting these values for η_p^*, ζ_p^*, η_n^* and ζ_n^* into (6.24) ψ has a posterior Log-Normal density with approximate mean

$$e^{1/dN_p(t)}\left[\frac{dN_n(t)\int_0^t Y_p(u)du}{dN_p(t)\int_0^t Y_n(u)du}\right] \tag{6.25}$$

and approximate variance

$$\left[\frac{dN_n(t)\int_0^t Y_p(u)du}{dN_p(t)\int_0^t Y_n(u)du}\right]^2\left(e^{3/dN_p(t)-1/dN_n(t)}-e^{2/dN_p(t)-2/dN_n(t)}\right) \tag{6.26}$$

We can see that the mean (6.25) reduces to the ratio of the maximum likelihood estimates in the single group case when $dN_p(t)$ is large, i.e the number of deaths per person days of follow-up.

We may also estimate the hazard ratio, ψ, using maximum likelihood methods. If we assume that $\lambda_p = \delta$ and therefore that $\lambda_n = \delta\psi$, the log-likelihood is given by,

$$\ell(\delta,\psi) = \sum_{i=1}^{n}[d_i\log_e(\delta\psi^{z_i}) - \delta\psi^{z_i}t_i]$$

where d_i is an indicator of death, t_i are survival/censoring times and z_i is an indicator variable representing treatment, i.e 0=photons, 1=neutrons. Partially differentiating once we obtain,

$$\begin{aligned}\frac{\partial \ell}{\partial \delta} &= \frac{\sum_{i=1}^{n} d_i}{\delta} - \psi \sum_{i=1}^{n} z_i t_i - \sum_{i=1}^{n} (1 - z_i) t_i \\ \frac{\partial \ell}{\partial \psi} &= \frac{\sum_{i=1}^{n} d_i z_i}{\psi} - \delta \sum_{i=1}^{n} z_i t_i\end{aligned}$$

and twice

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \delta^2} &= -\frac{\sum_{i=1}^{n} d_i}{\delta^2} \\ \frac{\partial^2 \ell}{\partial \psi^2} &= -\frac{\sum_{i=1}^{n} d_i z_i}{\psi^2} \\ \frac{\partial^2 \ell}{\partial \psi \partial \delta} &= -\sum_{i=1}^{n} z_i t_i\end{aligned}$$

Setting both the above partial first derivatives to zero yields a turning point, which can be shown to be a maximum provided that all the deaths do not occur in only one patient group. Therefore the maximum likelihood estimate, $\hat{\psi}$, of the the hazard ratio, ψ, is

$$\hat{\psi} = \frac{\sum_{i=1}^{n} d_i z_i \times \sum_{i=1}^{n} (1 - z_i) t_i}{\sum_{i=1}^{n} z_i t_i \times \sum_{i=1}^{n} (1 - z_i) d_i}$$

which is simply the ratio of the maximum likelihood estimates for the two hazards. We can obtain the asymptotic covariance matrix for the parameter estimates, and an estimated variance for $\hat{\psi}$ is

$$V(\hat{\psi}) = \frac{\hat{\psi}^2}{\sum_{i=1}^{n} z_i d_i}$$

6.5.3 Example

Construction and use of a clinical prior

In Section 3.2 the elicited beliefs of a non-random sample of clinicians about the efficacy of neutron therapy compared to photon therapy were described. Those beliefs were about the 12 month failure rates, but here we are interested in their beliefs about the hazard rate generally in the two groups. Given the 12 month failure rate it is possible to transform those beliefs into beliefs about the corresponding hazard rate, assuming that it is constant over the trial period.

The density, $f(x)$, of the exponential distribution is $\lambda e^{-\lambda x}$, where λ is the hazard rate. If θ is the 12 month hazard rate then $\int_0^{365} f_\lambda(x)dx = \theta$. Therefore $[-e^{-\lambda x}]_0^{365} = \theta$, which on re-arranging yields an expression for λ of $-\log_e(1 - \theta)/365$. From Section 3.2 we found that the mean 12 month failure rate for

neutrons, θ_n, was 0.46 and assumed that for photons, θ_p, was 0.5. Applying this method yields point estimates of 0.00171 for λ_n and 0.00190 for λ_p.

We would also like to transform the elicited beliefs about that variability of the 12 month failure rates. One possible solution to this problem is to construct an interval of beliefs for θ and transform the lower and upper limits of this interval onto a hazard scale using the method described in the previous paragraph. This interval for the hazard rate could then be used in conjunction with the point estimate to obtain an estimate of the variability in the beliefs about the hazard rate. This solution relies on the assumption that these beliefs are symmetric about the mean. Applying this technique yields variances for λ_n of $9.04E-07$ and for λ_p of $2.07E-07$. Alternatively we could have used the method given by Cox and Hinkley (1974) (page 302) for the asymptotic variance of a one-to-one and differentiable function of a maximum likelihood estimate.

Having obtained the prior estimates for the two intensities and their respective variances we now assume Gamma distributions for these, and use the relationships mentioned in Section 6.5.1 between the mean and variance of a Gamma distribution and its parameters, η and ζ. For the photon group, $E(\lambda_p) = 0.0019 = \zeta_p/\eta_p$ and $V(\lambda_p) = 2.07E-7 = \zeta_p/\eta_p^2$. Solving these two equations for η and ζ subject to the constraint that both η and ζ are greater than zero, yields $\eta_p = 9179$ and $\zeta_p = 17.44$. Similarly the prior estimates for η_n and ζ_n are 1890 and 3.23 respectively.

From the neutron therapy data of Section 2.2, using a censoring date of 21st December 1990, we find that the total number of deaths on the photon arm is 38, and on the neutron arm 71, whilst the total survival time for photon patients is 31453 days, and for neutron patients is 38806 days. Using this data, and applying the methods described in Section 6.5.1 to each group separately we can obtain the posterior estimates for the parameters of the two Gamma distributions, i.e η_p^*, ζ_p^*, η_n^* and ζ_n^*. These are 40632, 55.44, 40696 and 74.23 respectively.

Using these prior and posterior estimates for the parameters of the Gamma distributions, the respective densities for ψ, the ratio of the intensities, could be calculated, using either the exact method or a Normal approximation of Section 6.5.2. Table 6.1 shows various summary statistics for these distributions and they are displayed in Figure 6.2 together with the maximum relative log-likelihood.

We can see from Figure 6.2 that the posterior densities are almost exactly the same for the two methods, but there is a slight difference in the prior densities due to their skew nature. Prior to the trial the group of clinicians had approximately 63-65% belief that neutrons were beneficial, i.e that $\psi < 1$. In the light of the trial results they would believe that there was only approximately a 7% chance that neutrons were beneficial if they were being coherent. Similarly, there was a reduction in the degree of belief that the clinicians had that neutron therapy should be routinely used, i.e $\psi < 0.715$, from approximately 34-35% *a priori* to zero *a posteriori*. From Table 6.1 we can see that the prior mean is shifted from approximately 0.93 to 1.3, towards the maximum likelihood

estimate for the hazard ratio of 1.515, i.e to the other side of 1.

	Exact				Normal approx.			
	Mean	SD	$P_{(1)}$[a]	$P_{(0.72)}$	Mean	SD	$P_{(1)}$	$P_{(0.72)}$
Clinical Prior	0.927	0.544	0.626	0.414	0.953	0.801	0.649	0.432
Clinical Posterior	1.361	0.245	0.049	0.000	1.361	0.248	0.050	0.000
MLE	1.515	0.180	-	-	-	-	-	-
Reference Posterior	1.536	0.311	0.023	0.000	1.524	0.257	0.181	0.000

[a] $P_{(1)}$ and $P_{(0.72)}$ refer to the probability that ψ is less than 1 and 0.72 respectively.

Table 6.1: Maximum likelihood, prior and posterior summary statistics of $\psi = \lambda_n/\lambda_p$, for all patients and using a clinical prior.

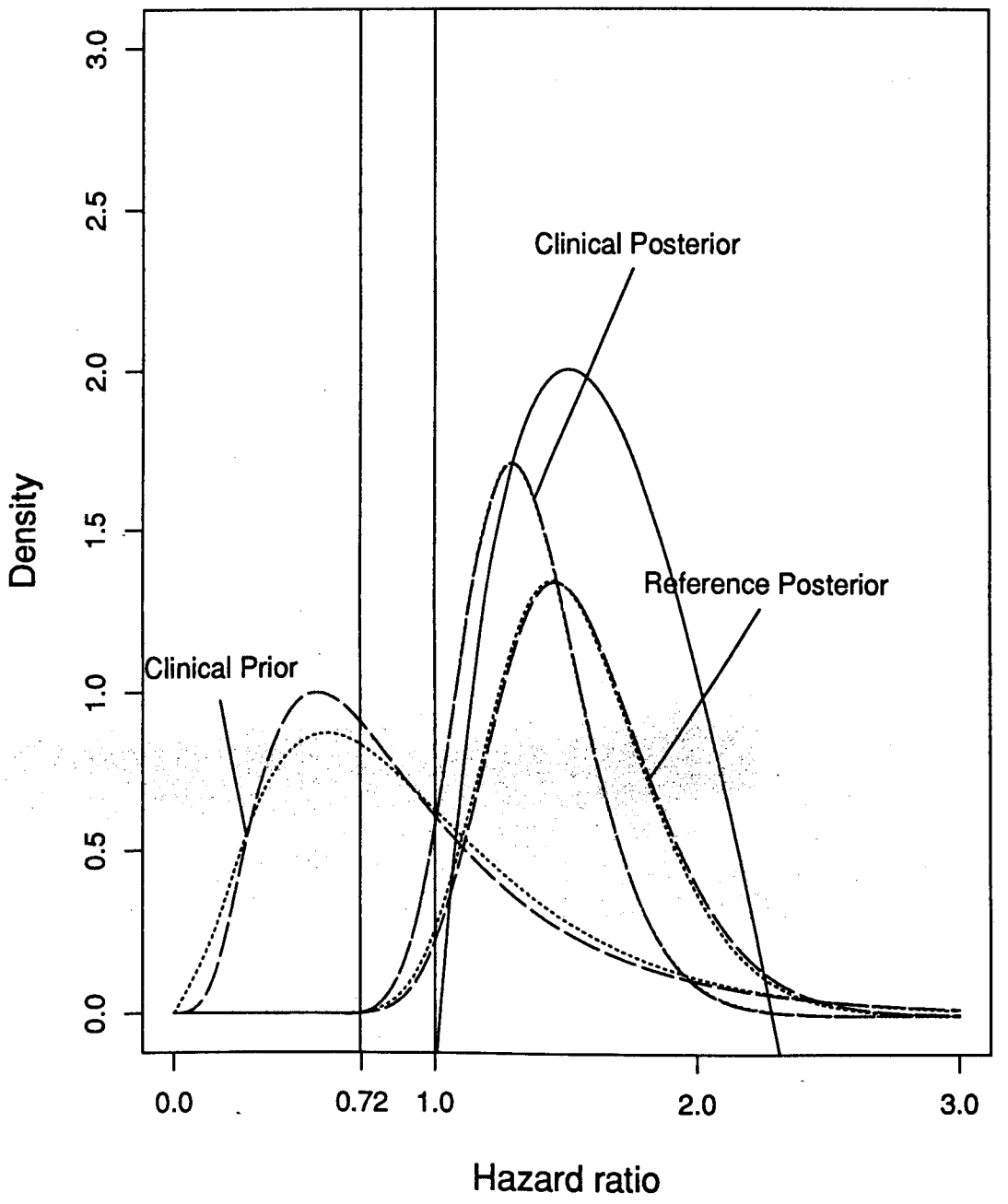
Construction and use of a meta prior

Errington *et al* (1991) report a meta analysis of trials that compared neutron therapy with standard photon therapy for patients with tumours of the pelvic region (rectum and bladder). Table 3.2 shows the numbers of neutron and photon patients who were alive/dead at 12 months in each of the six trials. Figure 3.5 shows the odds ratio and approximate 95% confidence interval for each trial, and the Mantel-Haenszel estimate of the overall odds ratio. As opposed to using a prior based on the information elicited from the ten clinicians as we have done above, we could use the information contained in the meta analysis.

We saw in Section 3.3.2 that overall, out of 6 trials reported, 60 out of 183 (32.8%) photon patients were dead at 12 months, and 107 out of 206 (51.9%) of neutron patients were dead at 12 months. The variances associated with these proportions are 0.0012 for photons and 0.0012 for neutrons. Therefore, $E(\theta_p) = 0.328$, $V(\theta_p) = 0.0012$, $E(\theta_n) = 0.519$ and $V(\theta_n) = 0.0012$.

Using the methods of Section 6.5.2, these 12 month probabilities of death can be transformed into estimates of the hazard, again assuming that the hazard remains constant and that it equals the value at 12 months. Therefore for photons, $E(\lambda_p) = 0.00109$ and $V(\lambda_p) = 2.222E-8$ while for neutrons $E(\lambda_n) = 0.00201$ and $V(\lambda_n) = 4.437E-8$. Assuming conjugate Gamma prior distributions again we can obtain the prior estimates for η and ζ. As before we can use the relationship $\eta = E(\lambda)/V(\lambda)$ and $\zeta = E(\lambda)^2/V(\lambda)$ to obtain estimates for η_p, ζ_p, η_n and ζ_n. These are 49010, 53.37, 45301 and 91.05 respectively.

As the meta analysis only considered rectum and bladder patients we will restrict the data to just these types of patients, $n = 119$. Using a censoring date of 21st December 1990, there are 32 photon deaths and 58 neutron deaths, whilst the total survival times for 19564 days for photon patients and 28900 days for neutron patients. Using the methods of Section 6.5.2 we can obtain the posterior estimates of two Gamma distribution parameters, i.e η_p^*, ζ_p^*, η_n^* and ζ_n^*. These are 68574, 85.37, 74201 and 149.05 respectively.



Key : —— maximum relative log-likelihood, ····· exact density and
– – – – – – – – – Normal approximation

Figure 6.2: Maximum relative log-likelihood, together with exact and approximate prior and posterior densities for hazard ratio (neutrons to photons) for neutron therapy survival data using all patients, $n = 154$.

	Exact				Normal approx.			
	Mean	SD	$P_{(1)}$ [a]	$P_{(0.72)}$	Mean	SD	$P_{(1)}$	$P_{(0.72)}$
MLE	1.227	0.161	-	-	-	-	-	-
Clinical Prior	0.927	0.544	0.626	0.414	0.953	0.801	0.649	0.432
Clinical Posterior	1.180	0.229	0.222	0.006	1.180	0.228	0.221	0.006
Meta Prior	1.877	0.322	0.000	0.000	1.881	0.327	0.000	0.000
Meta Posterior	1.633	0.224	0.000	0.000	1.633	0.223	0.000	0.000
Reference Posterior	1.248	0.278	0.185	0.009	1.266	0.282	0.168	0.006

[a] $P_{(1)}$ and $P_{(0.72)}$ refer to the probability that ψ is less than 1 and 0.72 respectively.

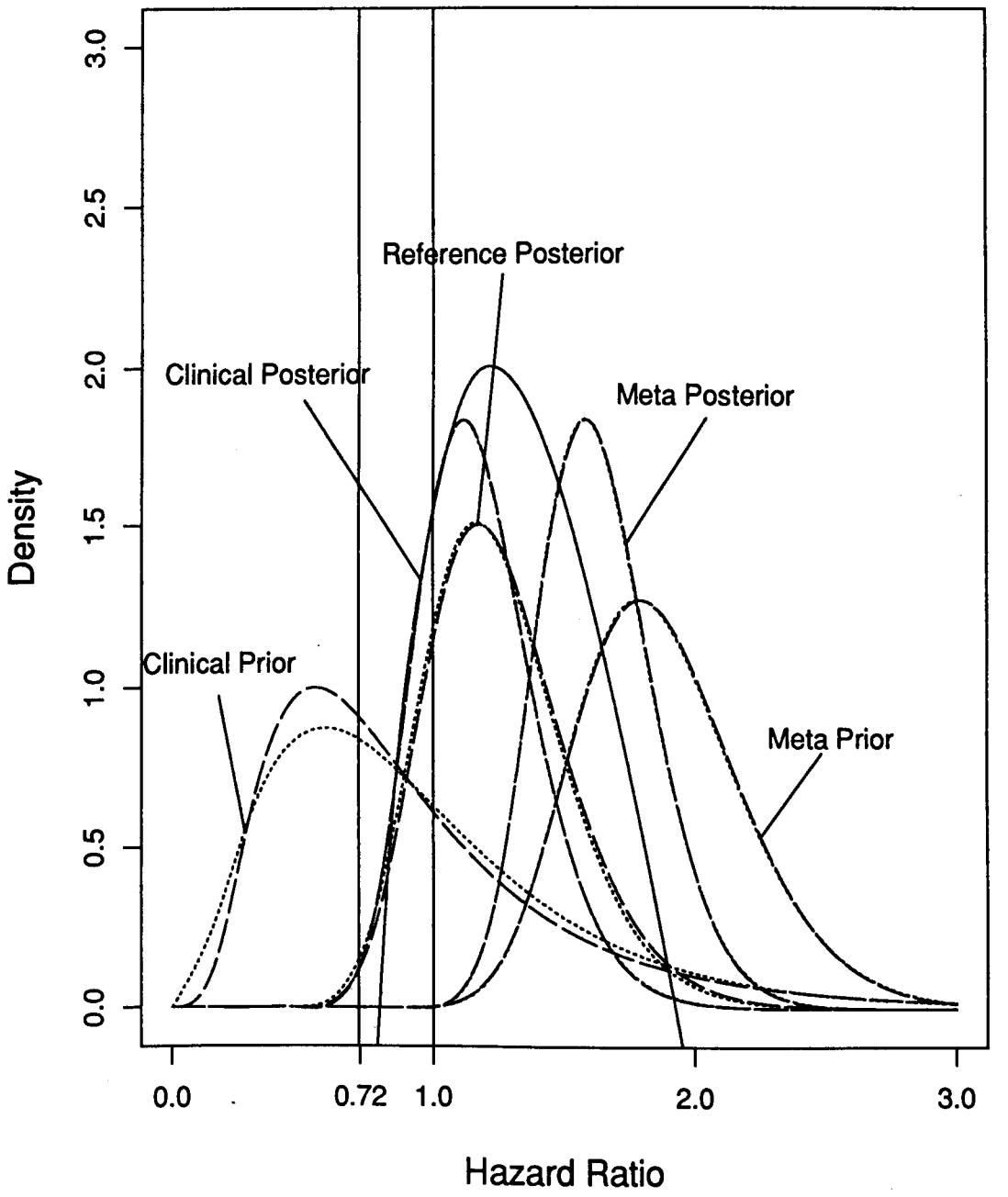
Table 6.2: Maximum likelihood, prior and posterior summary statistics of ψ for rectum and bladder patients, using a clinical prior and an overall meta prior

Table 6.2 shows various summary statistics for the prior to posterior analysis using just the rectum and bladder patients, when both the clinical prior and a prior based on the results of the previous trials described in Errington *et al* (1991) were used. The densities together with the maximum relative log-likelihood are shown in Figure 6.3. We can see that the clinical prior was optimistic about the relative efficacy of neutron therapy, while the prior based on previous trial was not, and in fact suggested that for rectum and bladder patients photon therapy was more effective. When we compare the effects of the data on these priors we see that the clinical prior is shifted quite considerably by the evidence of the trial results, while the meta prior is hardly affected. Not only was there a difference between the prior means, but also between the prior standard deviations. The clinical prior had a standard deviation for ψ of approximately 0.4, whilst the one based on the previous trial results had a standard deviation of 0.29. Note that both the exact solution and the one using a Normal approximation give consistent answers, with the densities in Figure 6.3 being almost indistinguishable.

6.5.4 Regression Model

The problem with the methods of Section 6.5.2 is that they cannot be easily extended to the case when there are more than two patient groups. Therefore in this section we will consider developing a regression model that allows us to include a number of possible covariates. Any model should be flexible and able to accommodate a number of different events, each of which may be recurrent. To achieve this we will make use of the counting process notation introduced in Section 5.2, and utilised in a fully parametric multiplicative intensity model in Section 5.3.3. In Chapter 3 we saw that there is often prior information available about the relative efficacy of treatments, and the regression models that we develop need to be able to use such prior information.

For each patient, indexed by i, there is a counting process, $N_i(t)$, which is either 0 or 1 depending whether the event of interest has occurred. Note that



Key : —— maximum relative log-likelihood, ····· exact density and
– – – – – – – – – Normal approximation

Figure 6.3: Maximum relative log-likelihood, together with exact and approximate prior and posterior densities for hazard ratio (neutrons to photons) for neutron therapy survival data using bladder and rectum patients, $n = 119$.

this is a simplification of the multivariate counting process, where subscript j has been dropped because we only have one time epoch, and subscript k has been dropped because there is only one type of event. Similarly associated with each counting process there is an intensity process $\alpha_i(t)$, and an observational process, $Y_i(t)$, which in this case will be either 0 or 1 depending whether the event has not or has occurred. Both $N_i(t)$ and $Y_i(t)$ in this situation are shown in Figures 5.3 and 5.4.

The simplest form of a multiplicative intensity model is considered when there is one baseline intensity parameter, θ. The multiplicative intensity model of Aalen (1978) that was described in Section 5.3 is of the form

$$\alpha_i(t) = \lambda_0(t|\underline{\theta})\, e^{\underline{\beta}^T \underline{Z}_i}\, Y_i(t) \tag{6.27}$$

and in this particular case then λ_0 is constant and we will parameterise the model such that $\lambda_0 = e^{\theta}$. The reason for doing this is so that θ is a parameter over the real line, but λ_0 remains positive, as it must by the definition of the intensity, $\alpha(t)$. Thus substituting $\lambda_0 = e^{\theta}$ into (6.27) yields

$$\alpha_i(t) = e^{\theta}\, e^{\underline{\beta}^T \underline{Z}_i}\, Y_i(t) \tag{6.28}$$

Initially we will consider the case when there is only one binary covariate in the model. This could indicate treatment regime in the neutron therapy example. Therefore $e^{\underline{\beta}^T \underline{Z}_i}$ is of the form $e^{\beta z_i}$, where z_i is an observed time-independent treatment indicator for the ith patient.

Likelihood

Using an argument analogous to that in Section 5.3.3 we can, under the assumption of non-informative censoring, obtain a likelihood for the two parameters in the model. As we saw in Section 5.3.3, in general the likelihood function is of the form

$$L(\underline{\phi}) = \prod_{i=1}^{n} f(t|\underline{\phi})^{dN_i(t)} S(t|\underline{\phi})^{1-dN_i(t)}$$

where $f(t|\underline{\phi})$ and $S(t|\underline{\phi})$ are the density and survivor function respectively. Usually we would like to work with the log-likelihood function, and express this in terms of transition intensities,

$$\ell(\underline{\phi}) = \sum_{i=1}^{n} \left\{ \log[\alpha_i(t|\underline{\phi})] dN_i(t) - \int_0^t \alpha_i(u|\underline{\phi}) du \right\} \tag{6.29}$$

Given the definition of $\alpha_i(t|\underline{\phi})$ above in (6.28) we can substitute this into (6.29) to obtain an expression for the log-likelihood, for the case when $\underline{\phi} = (\theta, \beta)$.

$$\begin{aligned} \ell(\theta,\beta) &= \sum_{i=1}^{n} \left[\log(\alpha_0(t) e^{\beta z_i} Y_i(t)) dN_i(t) - \alpha_0 e^{\beta z_i} \int_0^t Y_i(u) du \right] \\ &= \sum_{i=1}^{n} \left[\log(e^{\theta} e^{\beta z_i} Y_i(t)) dN_i(t) - e^{\theta} e^{\beta z_i} \int_0^t Y_i(u) du \right] \end{aligned}$$

$$= \theta \sum_{i=1}^{n} [Y_i(t)dN_i(t)] + \beta \sum_{i=1}^{n} [z_i Y_i(t)dN_i(t)]$$
$$-e^{\theta} \left[e^{\beta} \sum_{i=1}^{n} z_i \int_0^t Y_i(u)du + \sum_{i=1}^{n} (1 - z_i) \int_0^t Y_i(u)du \right] \quad (6.30)$$

Similarly the likelihood is,

$$L(\theta, \beta) = e^{\theta \sum_{i=1}^{n} [Y_i(t)dN_i(t)]} e^{\beta \sum_{i=1}^{n} [z_i Y_i(t)dN_i(t)]}$$
$$e^{e^{\theta} \left[e^{\beta} \sum_{i=1}^{n} z_i \int_0^t Y_i(u)du + \sum_{i=1}^{n} (1-z_i) \int_0^t Y_i(u)du \right]} \quad (6.31)$$

Note that $\sum_{i=1}^{n} [Y_i(t)dN_i(t)]$ represents the total number of events in all n patients, in a survival problem this would be the total number of deaths, while $\sum_{i=1}^{n} [z_i Y_i(t)dN_i(t)]$ represents the number of events in the patient group indexed by z_i. $\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du$ represents the total time at risk for patients for whom $z_i = 1$ and $\sum_{i=1}^{n} (1 - z_i) \int_0^t Y_i(u)du$ represents the total time at risk for patients for whom $z_i = 0$.

Prior Distributions

In this section we describe the use of informative and non-informative prior evidence and beliefs. In Chapter 3 we saw that some knowledge is usually available about the treatment parameter, or some function of it. For the baseline parameters we have in some cases assumed arbitrary values for them in eliciting beliefs about the treatment parameter, and so have decided to use a non-informative prior density for the baseline parameter.

Non-Informative Prior Distributions

Jeffreys (1961) suggested that minus the expectation of the information matrix could be used as a vague prior. Let $\underline{\phi} = (\theta, \beta)$, and $\ell(\underline{\phi})$ be the log-likelihood for a two parameter model (6.30). Also let $\underline{x}$ represent *all* the data i.e $x_i = (N_i(t), Y_i(t), \underline{z}_i)$, where $N_i(t)$ is a counting process which is zero if patient i has not died by time t, and one they if they have , $Y_i(t)$ is an at-risk process that is one whilst patient i is alive at time t, but zero when they are dead and $\underline{z}_i$ is a vector of observed time-independent covariates for patient i. The Information Matrix, $I(\underline{\phi})$ is

$$-E_{\underline{x}|\underline{\phi}} \left[\frac{\partial^2 \ell(\underline{\phi})}{\partial \underline{\phi}^2} \right]$$

Consider the partial derivatives of $\ell(\underline{\phi})$, the hessian is

$$\begin{bmatrix} -e^{\theta} \left[e^{\beta} \sum_{i=1}^{n} z_i \int_0^t Y_i(u)du + \sum_{i=1}^{n} (1 - z_i) \int_0^t Y_i(u)du \right] & -e^{\theta} \left[e^{\beta} \sum_{i=1}^{n} z_i \int_0^t Y_i(u)du \right] \\ -e^{\theta} \left[e^{\beta} \sum_{i=1}^{n} z_i \int_0^t Y_i(u)du \right] & -e^{\theta} \left[e^{\beta} \sum_{i=1}^{n} z_i \int_0^t Y_i(u)du \right] \end{bmatrix}$$

Consider now taking the expectation of each element of the hessian with respect to $\underline{x}|\underline{\phi}$. $E_{\underline{x}|\underline{\phi}}\left[-e^{\theta+\beta}\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du\right]$ using the additive properties of expectation reduces to $-e^{\theta+\beta}\sum_{i=1}^{n} E_{\underline{x}|\underline{\phi}}\left[z_i \int_0^t Y_i(u)du\right]$, but as z_i are observed quantities, $E_{\underline{x}|\underline{\phi}}\left[\int_0^t Y_i(u)du\right]$ is simply the reciprocal of the intensity rate i.e $1/e^{\theta+\beta}$ if $z_i = 1$ and zero otherwise. Therefore

$$E_{\underline{x}|\underline{\phi}}\left[-e^{\theta+\beta}\sum t_i z_i\right] = -e^{\theta+\beta}\frac{1}{e^{\theta+\beta}}\sum z_i$$

and so the expectation is a constant i.e $\sum z_i$.

Similarly the expectation of $\sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du$ with respect to $\underline{x}|\underline{\phi}$ can also be shown to be a constant, $\sum_{i=1}^{n}(1-z_i)$.

Therefore using Jeffreys' vague prior is the same as using a Uniform prior regardless of whether we assume that the parameters are *a priori* independent. This result for the exponential model has also be noted by Kass and Slate (1992).

Informative Prior Distributions

In the case of a regression model, we could use a Gamma prior density for θ, but because of the form of the model this does not yield closed form marginal posterior densities or even a closed form joint posterior density, nor would it be applicable for β. Therefore, a more suitable density is the Normal density. From an estimation point of view it has several important features one of which is unimodality. It also has the property that its parameters have clear meaning.

Posteriors

The Bayesian paradigm states that the posterior density is proportional to the likelihood multiplyed by the prior density, i.e

$$P(\underline{\phi}|H) \propto L(\underline{\phi}) \times P(\underline{\phi})$$

The exact posterior density can be obtained by dividing the product of the likelihood and the prior density, by the integral of this product over the whole of the parameter space of $\underline{\phi}$. Thus,

$$p(\underline{\phi}|H) = \frac{L(\underline{\phi})\, p(\underline{\phi})}{\int_{\underline{\Phi}} L(\underline{\phi})\, p(\underline{\phi})\, d\underline{\phi}}$$

Depending upon which prior is used we will obtain different posterior density functions. The 'most' complicated will result when there is a multivariate Normal prior for θ and β. The 'simplest' will be obtained when there are uninformative priors for both θ and β, in this case a 'standardised' likelihood will result.

Using Uniform prior densities for both θ and β yields a joint posterior of the form

$$p(\theta,\beta|H) \propto e^{\theta\sum_{i=1}^{n}[Y_i(t)dN_i(t)]}\, e^{\beta\sum_{i=1}^{n}[z_iY_i(t)dN_i(t)]}\, e^{e^{\theta}\left[e^{\beta}\sum_{i=1}^{n}z_i\int_0^t Y_i(u)du+\sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du\right]} \tag{6.32}$$

Using a Uniform density for θ, but an informative Normal prior density for β with mean η and variance ω^2, both of which are assumed to be known, yields a joint posterior of the form

$$p(\theta,\beta|H) \propto e^{\theta\sum_{i=1}^{n}[Y_i(t)dN_i(t)]}\, e^{\beta\sum_{i=1}^{n}[z_iY_i(t)dN_i(t)]}\, e^{e^{\theta}\left[e^{\beta}\sum_{i=1}^{n}z_i\int_0^t Y_i(u)du+\sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du\right]}\, e^{-1/2\frac{(\beta-\eta)^2}{\omega^2}} \tag{6.33}$$

Finally considering the case when we wish to put an informative prior on both θ and β independently, i.e both Normal densities, yields a posterior density of the form

$$p(\theta,\beta|H) \propto e^{\theta\sum_{i=1}^{n}[Y_i(t)dN_i(t)]}\, e^{\beta\sum_{i=1}^{n}[z_iY_i(t)dN_i(t)]}\, e^{e^{\theta}\left[e^{\beta}\sum_{i=1}^{n}z_i\int_0^t Y_i(u)du+\sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du\right]}\, e^{-1/2\frac{(\beta-\eta)^2}{\omega^2}}\, e^{-1/2\frac{(\theta-\nu)^2}{\zeta^2}} \tag{6.34}$$

where the parameters of the two prior distributions, η, ω^2, ν and ζ^2 are all assumed to be known.

Parameter Estimation

Letting $\underline{\phi}$ be equal to (θ,β), then we are interested in obtaining the marginal posterior densities for both θ and β, and posterior estimates for the mean vector and the covariance matrix for $\underline{\phi}$. We could use any of the three estimation methods described in Section 6.4, but we will concentrate on applying Laplace approximations (Section 6.4.1) and Gauss-Hermite quadrature (Section 6.4.2).

First consider the application of Laplace approximations to the joint posterior in (6.32). In this case $G(\underline{\phi}) = \log_e[p(\theta,\beta|H)]/n$. Thus we need to be able to maximise the logarithm of (6.32) both when θ and β are unknown, and also in the case when one of the parameters is assumed fixed and we wish to maximise with respect to the other parameter. We also require the second derivatives of the logarithm of (6.32). In this simple two parameter example these can be obtained analytically.

$$nG(\underline{\phi}) = \theta\sum_{i=1}^{n}[Y_i(t)dN_i(t)] + \beta\sum_{i=1}^{n}[z_iY_i(t)dN_i(t)] - e^{\theta}\left[e^{\beta}\sum_{i=1}^{n}z_i\int_0^t Y_i(u)du + \sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du\right]$$

In order to maximise $G(\underline{\phi})$ with respect to $\underline{\phi}$ consider the first derivatives

$$\frac{\partial G(\underline{\phi})}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n}[Y_i(t)dN_i(t)] - e^{\theta}\left[e^{\beta}\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du + \sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du\right] \quad (6.35)$$

$$\frac{\partial G(\underline{\phi})}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n}[z_i Y_i(t)dN_i(t)] - e^{\theta}\left[e^{\beta}\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du\right] \quad (6.36)$$

We can see from (6.35) and (6.36) that an explicit solution to the maximisation problem does not exist. Therefore an iterative procedure is required. For this simple two parameter example the necessary calculations were performed in Splus (1990)[1] using the `nlmin` function that is based on a quasi-Newton method. For further details see Dennis and Mei (1979). We shall see in later sections in this chapter, that as the regression model becomes more complicated with many parameters other methods of optimisation need to be employed. This is due partly to the algorithm used in `nlmin` and partly to the time that Splus takes to perform such optimisations.

In order to use the Laplace approximation in (6.11) we also require the hessian of G. In this two parameter problem we can write down the hessian matrix without too much difficulty. This can then be evaluated at the necessary parameter values. In more complicated regression models we may need to consider the use of numerical estimates of the hessian matrix, as differentiation would be time consuming. The second derivatives for the two parameter model are

$$\begin{aligned}
\frac{\partial^2 G(\underline{\phi})}{\partial \theta^2} &= -e^{\theta}\left[e^{\beta}\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du + \sum_{i=1}^{n}(1-z_i)\int_0^t Y_i(u)du\right] \\
\frac{\partial^2 G(\underline{\phi})}{\partial \theta \partial \beta} &= -e^{\theta}\left[e^{\beta}\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du\right] \\
\frac{\partial^2 G(\underline{\phi})}{\partial \beta^2} &= -e^{\theta}\left[e^{\beta}\sum_{i=1}^{n} z_i \int_0^t Y_i(u)du\right] \qquad (6.37)
\end{aligned}$$

The application of Gauss-Hermite quadrature to (6.32), (6.33) and (6.34) is more straight forward. Using BAYES4 which requires only that the functional form of both the log-likelihood and prior density are given, the parameter estimates, their standard deviations and posterior densities were calculated . Convergence was defined as Δ being less than 10^{-3} on two consecutive iterations. Further iterations were are in order to obtain parameter estimates and posterior densities. These were all calculated on a 19×19 grid of points.

Example

Applying model (6.28) to the neutron therapy data in Chapter 2, together with both the clinical and meta priors derived in Sections 3.2 and 3.3 and shown in

[1]Splus is marketed by Statistical Sciences U.K, 52 Sandfield Road, Oxford, OX3 7RJ

Table 3.4, yields the summary statistics in Table 6.3. There are two main points to notice. The first is that Laplace approximations and Gauss-Hermite quadrature give very similar results. This is also borne out by Figures 6.5 and 6.6 which show the marginal posterior densities together with the maximum relative log-likelihood. The second point is that these results reinforce the results obtained from using the non-temporal models of Chapter 4 and the exact analysis of Section 6.5.3. That is, the clinical prior beliefs, that neutrons are likely to be beneficial, should be revised in the light of the data, so that *a posteriori* there would be only slight belief in the beneficial effect of neutrons for pelvic tumours. In contrast to the elicited clinical beliefs, the prior density based on previous trial results would be changed little in the light of the data, and if anything should be revised so that *a posteriori* there was slightly more belief in the relative efficacy of neutron therapy for pelvic tumours, but the vast majority of the mass of the density is to the right of zero indicating that neutrons are not beneficial, compared to photons.

We can also see from Table 6.3 that using either a clinical or meta prior for β leads to a reduction in the standard deviation of the parameter estimate compared to the standard deviation of the maximum likelihood estimate. This is also true for the standard deviation of θ even though we have used a prior Uniform density for this parameter. The covariance estimates appear to change little from those obtained under maximum likelihood. This reflects the fact that we have assumed an independent prior structure, and not hierarchical one. This will be discussed further in the discussion section.

A sensitivity analysis in which the informative prior for the treatment parameter β is varied, whilst a Uniform prior density is assumed for θ, yields the summary statistics in Table 6.4 and the marginal densities for β are shown in Figure 6.7. As would be expected a near doubling of the prior standard deviation results not only in a slightly increased posterior standard deviation, but also in the posterior mean being shifted nearer to the maximum likelihood estimate, Figure 6.7 (a). A doubling of the prior clinical mean, but with the prior standard deviation remaining approximately constant, results in the posterior mean being only just positive, i.e indicating that neutrons are not beneficial, with the posterior standard deviation remaining approximately equal to the case when the clinical prior density is used, Figure 6.7 (b). Figures 6.7 (c), (d) and (e) show the case when the prior mean is fixed at an extreme value, -1.2, indicating a belief in the beneficial effect of neutrons, and the prior standard deviation is varied. In the case when a very small standard deviation is assumed the posterior density is only shifted slightly away from the prior density. Increasing the prior standard deviation from 0.1 to 0.173 results in the posterior density being shifted to the right, as it is influenced to a greater degree by the likelihood. When the prior standard deviation is increased to 0.361 the posterior density is influenced by the likelihood to an even greater extent. There is now a substantial weight of evidence *a posteriori* that neutrons are in fact not beneficial. From Table 6.4 we can see that the posterior probability of neutrons being beneficial fall from 1 when the standard deviation is either 0.1 or 0.173,

to 0.4 when a prior standard deviation of 0.361 is used. Finally we consider the case when a prior with a mean greater than the meta prior mean is used. This indicates that *a priori* there is strong belief in photons being more beneficial for patients with tumours of the pelvic region than neutrons. As we can see from Figure 6.7 (f) the corresponding posterior is only slightly shifted by the weight of evidence in the current trial, with even more sceptical posterior beliefs about the efficacy of neutron therapy for these patients than the actual trial results. The usefulness of a sensitivity analysis is that it is a means by which the strength of evidence for the use of neutron therapy given by the *current* trial results can be assessed. In the sensitivity analysis described here we can see that only when a very informative prior density that conflicted with the *current* trial results was used is there a major discrepancy between the posterior density and the likelihood.

In all of the above analyses we have assumed that there is little prior information about the baseline intensity. However, it should be possible at least to use the results of the meta analysis described in Section 3.3.2 to put some prior distribution on θ, and to use the clinical prior for β. In Section 6.5.3 above we saw that using the results of the meta analysis we could obtain prior estimates for the hazard rate for both neutron and photon patients. For photon patients these were $E(\lambda_p) = 0.00109$ and $V(\lambda_p) = 2.222E-8$. These can be transformed into a prior for θ, i.e $\log_e(\lambda_p)$. If we assume a Normal distribution for this prior distribution of θ it will have a mean of -6.897 and a variance of 0.02. As the meta analysis considered only bladder and rectum patients, we will analyse only data for these patients. Thus the data set is reduced from 154 patients to 119.

Using independent Normal priors yields the estimates in Table 6.5. Figure 6.8 shows the marginal densities for θ and β, whilst Figure 6.9 shows the bivariate densities for θ and β. We can see from Figure 6.8 that the meta prior for θ indicated that photon patients had a smaller hazard than the actual data suggested. Thus the posterior density for θ was approximately half way between the meta prior and the maximum relative log-likelihood. As in the previous example the clinical prior for β indicated that neutrons patients did better than photon patients. Similarly these beliefs should be updated so that *a posteriori* it is believed that neutron patients do not actually fair better than photon patients. A difference in the revision of the prior beliefs about the log hazard ratio in this case compared to when θ was assumed to have a Uniform prior density is that the mean and mode for β is to the right of the maximum likelihood value. At first this appears unusual but we can see from Table 6.5 that the posterior covariance of θ and β is approximately -0.01, which corresponds to a posterior correlation of -0.6. Thus a posterior density for θ with a mode less than that of the maximum relative log-likelihood, will cause the posterior density for β to increased more than it would otherwise have been. This example highlights the fact that careful examination of both the marginal and bivariate posterior densities is required in order that mis-leading inferences are not made.

Figure 6.9 shows that the joint log-likelihood is uni-modal and appears well behaved. Thus both Laplace approximations and Gauss-Hermite quadrature

would be expected to perform reasonably well, (Hills and Smith, 1992).

	θ (Baseline)		β (Neutrons)				
	Mean	SD	Mean	SD	$P_{(0)}$[a]	$P_{(-0.26)}$	$Cov(\theta, \beta)$
MLE	-6.719	0.162	0.415	0.201	-	-	-0.026
Prior	-	∞	-	∞	-	-	0.0
Posterior (L)[b]	-6.732	0.166	0.421	0.201	0.015	0.000	-0.026
Posterior (G-H)[c]	-6.732	0.163	0.421	0.202	0.017	0.000	-0.026
Prior	-	∞	-0.116	0.286	0.654	0.305	0.0
Posterior (L)	-6.619	0.138	0.245	0.162	0.060	0.000	-0.016
Posterior (G-H)	-6.619	0.138	0.245	0.162	0.064	0.001	-0.016
Prior	-	∞	0.430	0.130	0.000	0.000	0.0
Posterior (L)	-6.732	0.123	0.426	0.116	0.000	0.000	-0.009
Posterior (G-H)	-6.732	0.122	0.426	0.116	0.000	0.000	-0.009

[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that β is less than 0 and -0.26 respectively.
[b] L denotes estimation using Laplace approximations.
[c] G-H denotes estimation using Gauss-Hermite quadrature.

Table 6.3: Parameter estimates for constant intensity model using clinical and meta priors, and using 21st December 1990 as censoring date.

	θ (Baseline)		β (Neutrons)				
	Mean	SD	Mean	SD	$P_{(0)}$[a]	$P_{(-0.26)}$	$Cov(\theta,\beta)$
MLE	-6.719	0.162	0.415	0.201	-	-	-0.026
Prior	-	∞	-0.116	0.405	0.613	0.361	0.0
Posterior (L)[b]	-6.663	0.146	0.315	0.178	0.034	0.000	-0.020
Posterior (G-H)[c]	-6.663	0.147	0.315	0.178	0.037	0.000	-0.020
Prior	-	∞	-0.260	0.202	0.901	0.500	0.0
Posterior (L)	-6.526	0.076	0.088	0.140	0.254	0.006	-0.011
Posterior (G-H)	-6.525	0.125	0.088	0.140	0.264	0.006	-0.011
Prior	-	∞	-1.2	0.100	1.0	1.0	0.0
Posterior (L)	-6.089	0.100	-0.863	0.090	1.0	1.0	-0.003
Posterior (G-H)	-6.089	0.101	-0.863	0.090	1.0	1.0	-0.003
Prior	-	∞	-1.2	0.173	1.0	1.0	0.0
Posterior (L)	-6.237	0.107	-0.484	0.129	0.999	0.956	-0.007
Posterior (G-H)	-6.237	0.111	-0.484	0.129	0.999	0.958	-0.007
Prior	-	∞	-1.2	0.361	0.999	0.995	0.0
Posterior (L)	-6.503	0.133	0.047	0.171	0.376	0.032	-0.015
Posterior (G-H)	-6.503	0.136	0.047	0.170	0.392	0.035	-0.016
Prior	-	∞	1.5	0.173	0.000	0.000	0.0
Posterior (L)	-7.196	0.246	1.073	0.139	0.000	0.000	-0.018
Posterior (G-H)	-7.199	0.145	1.073	0.139	0.000	0.000	-0.015

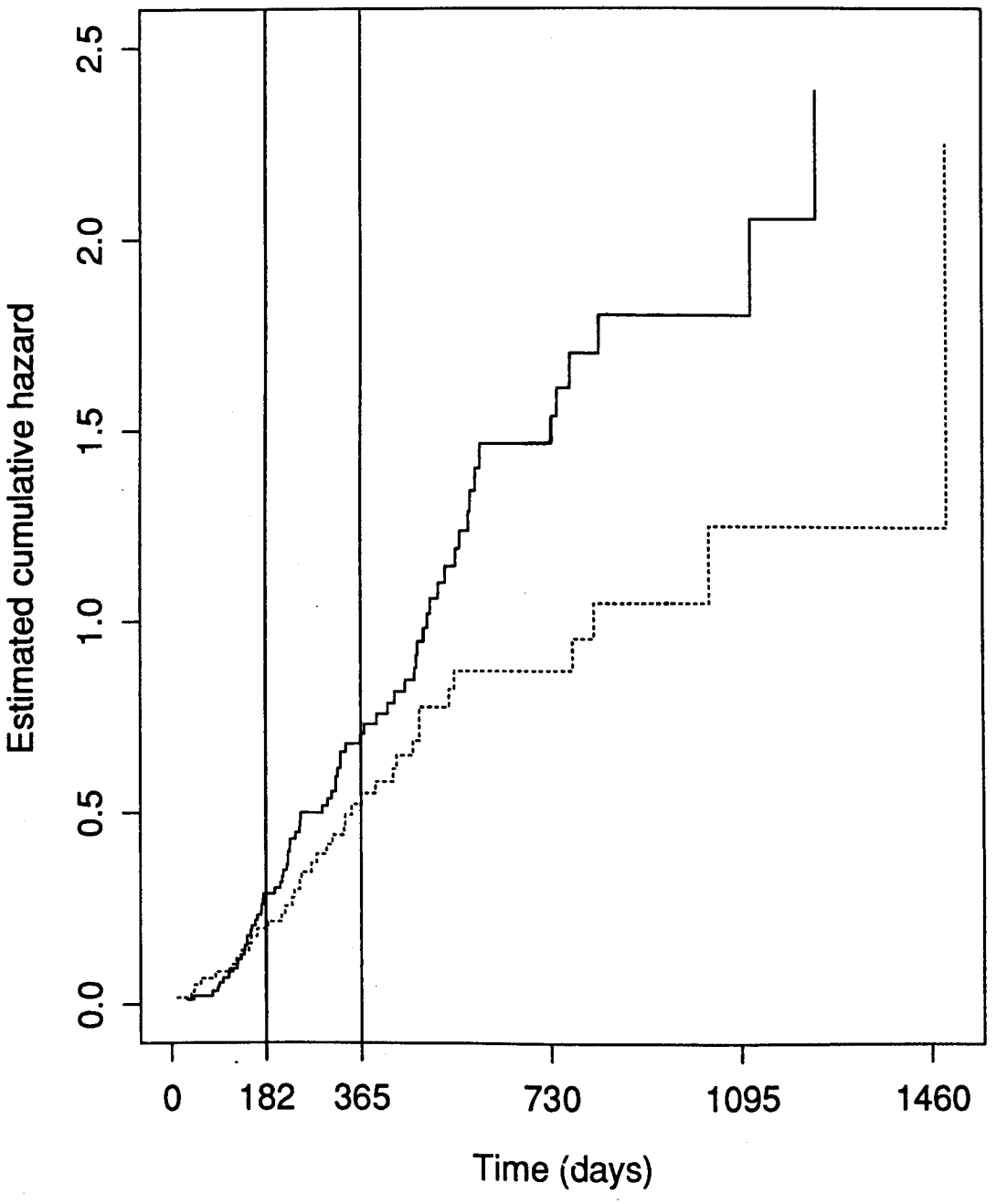
[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that β is less than 0 and -0.26 respectively.
[b] L denotes estimation using Laplace approximations.
[c] G-H denotes estimation using Gauss-Hermite quadrature.

Table 6.4: Parameter estimates for various priors for survival data, using 21st December 1990 as censoring date.

	θ (Baseline)		β (Neutrons)				
	Mean	SD	Mean	SD	$P_{(0)}$[a]	$P_{(-0.26)}$	$Cov(\theta,\beta)$
MLE	-6.416	0.177	0.205	0.220	-	-	-0.031
Prior	-6.897	0.141	-0.116	0.286	0.654	0.305	0.0
Posterior (L)[b]	-6.654	0.108	0.335	0.152	0.013	0.000	-0.009
Posterior (G-H)[c]	-6.654	0.108	0.335	0.152	0.014	0.000	-0.010

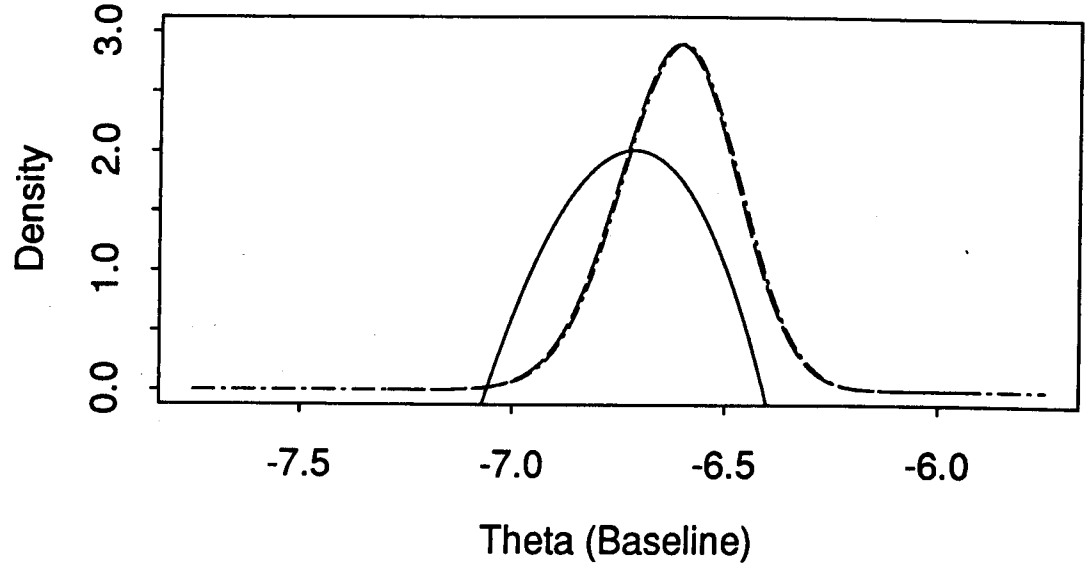
[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that β is less than 0 and -0.26 respectively.
[b] L denotes Laplace approximation.
[c] G-H denotes Gauss-Hermite quadrature.

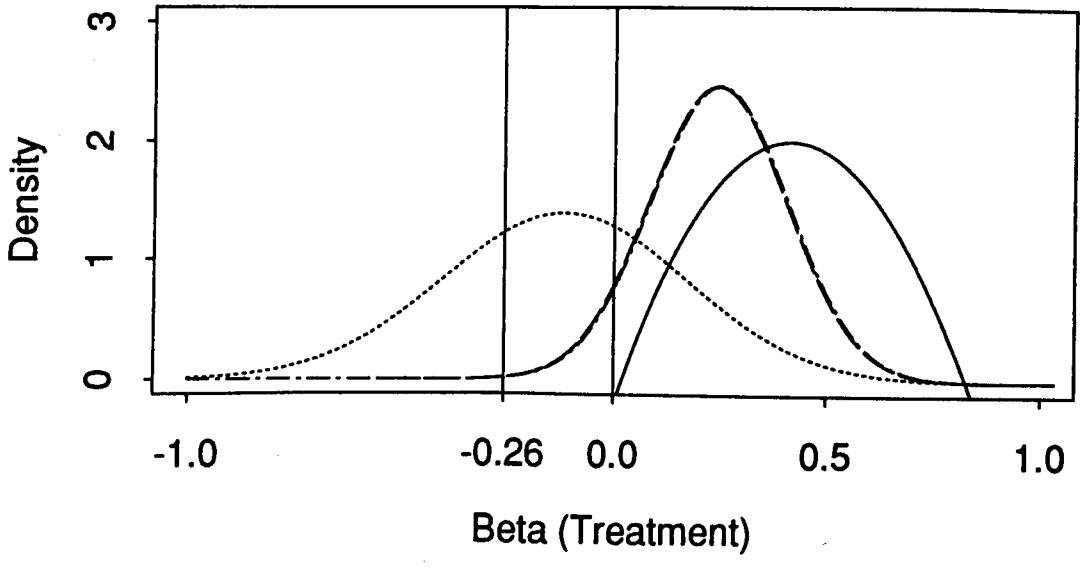
Table 6.5: Rectum & Bladder patients, meta prior for θ and clinical prior for β - using 21st December 1990 as a censoring date, i.e $n = 119$.



Key : —— neutrons and ····· photons

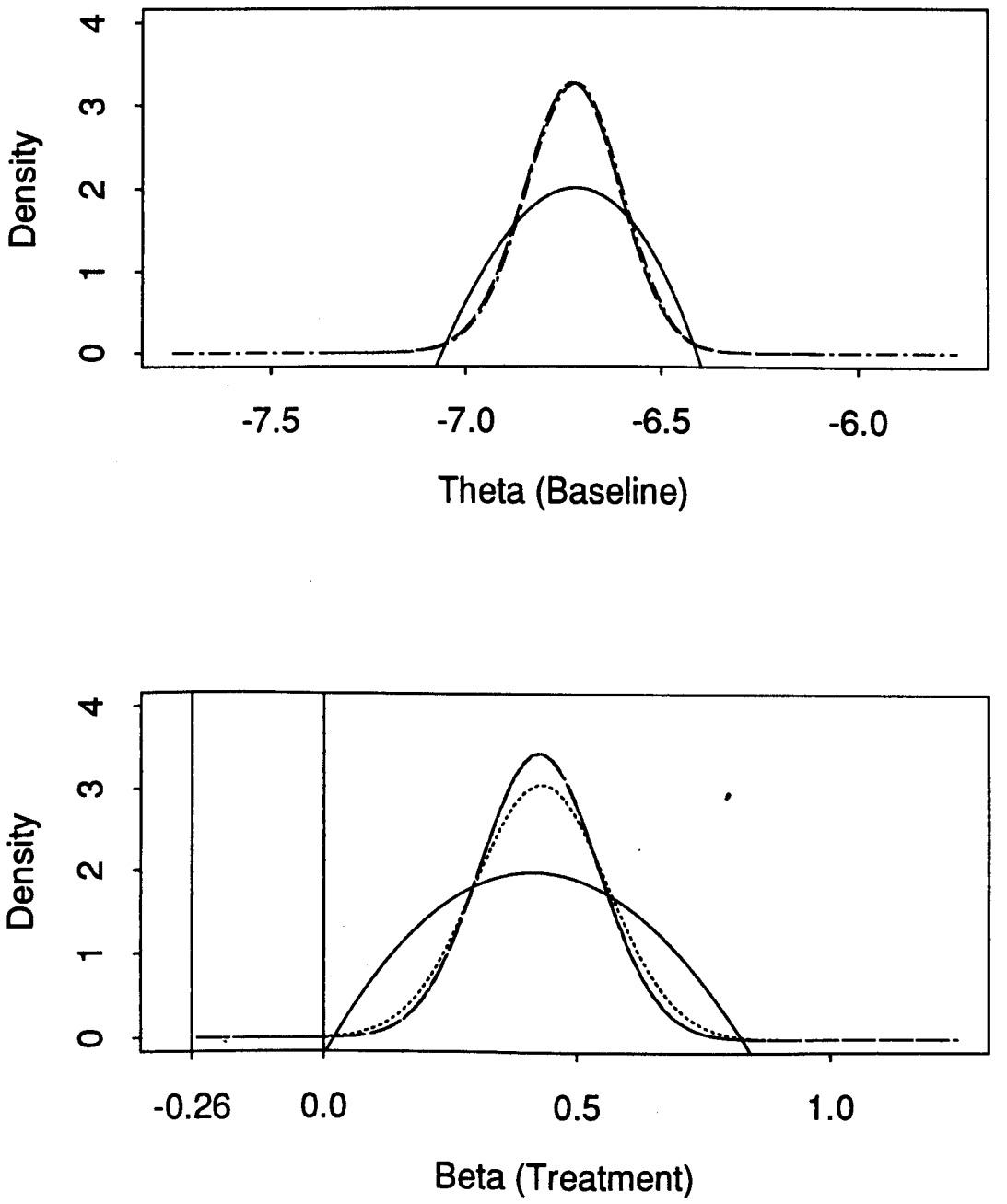
Figure 6.4: Cumulative hazard based on Kaplan-Meier estimates for the survivor function for neutrons and photons, using survival data (n=154).





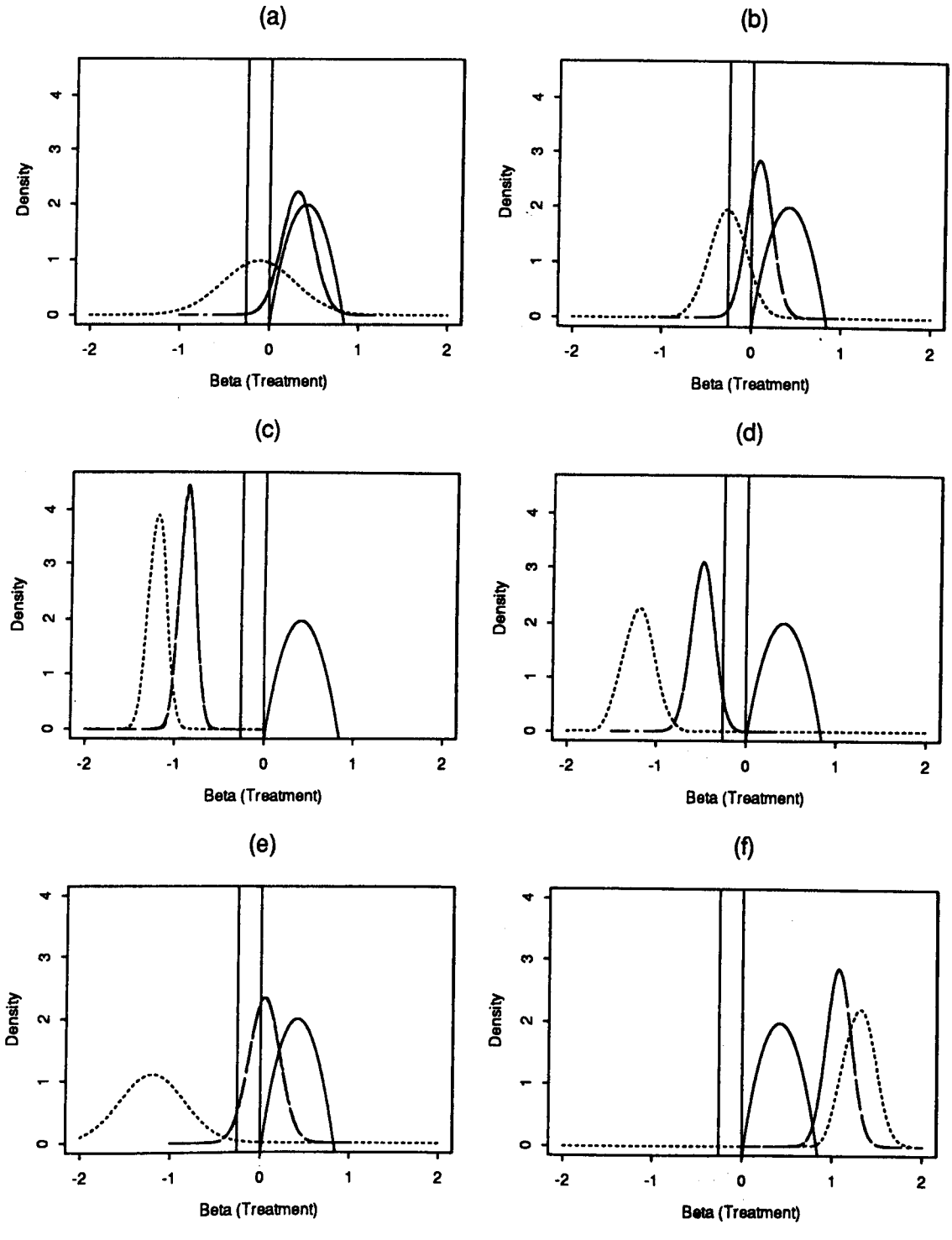
Key : — maximum relative log-likelihood, ····· clinical prior, – · – · – · – posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.5: Maximum relative log-likelihood and marginal densities for survival model assuming constant baseline intensity, vague prior for θ and clinical prior for β.



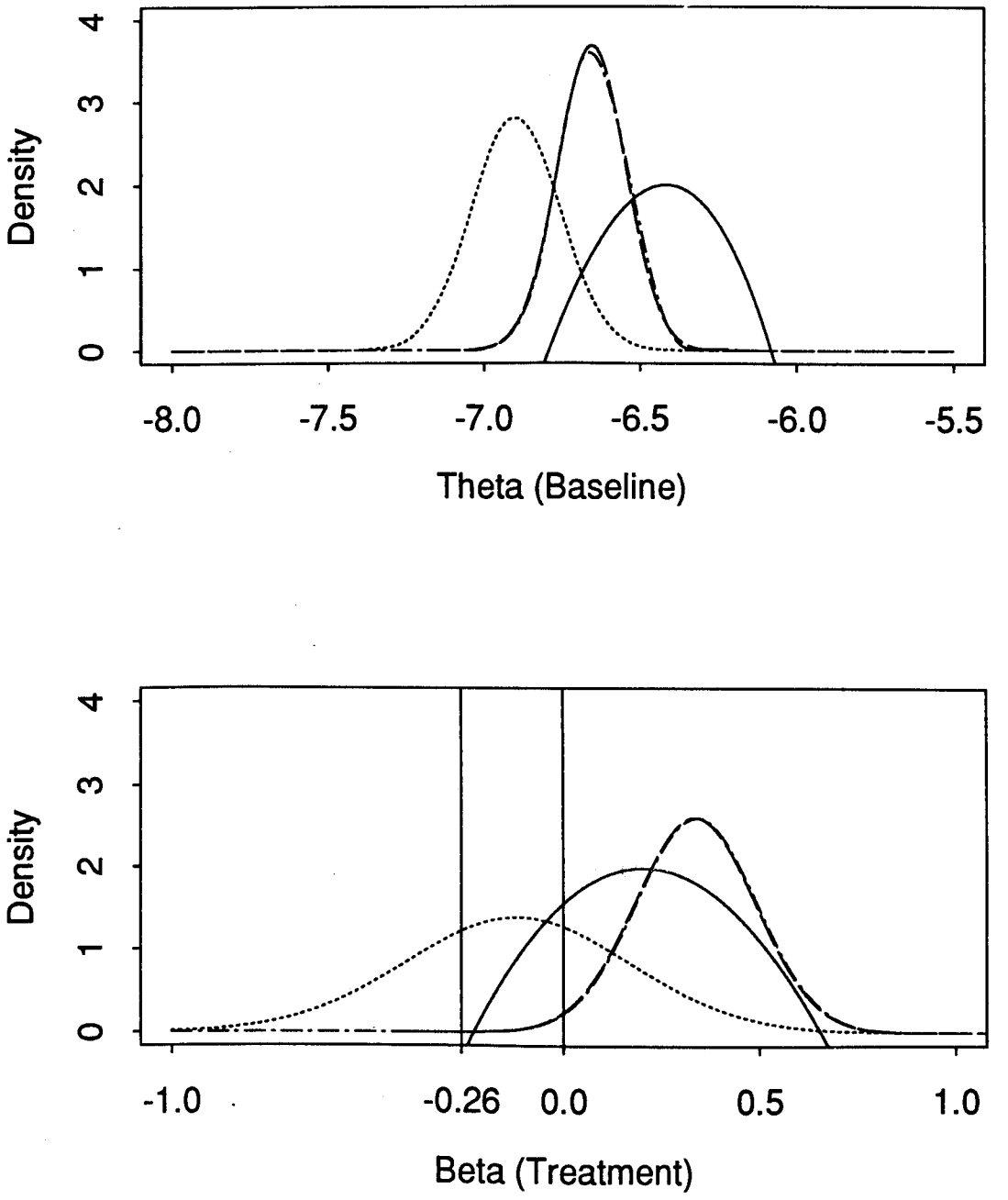
Key : — maximum relative log-likelihood, ····· meta prior, —·—·— posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.6: Maximum relative log-likelihood and marginal for survival model assuming constant baseline intensity, vague prior for θ and meta prior for β.



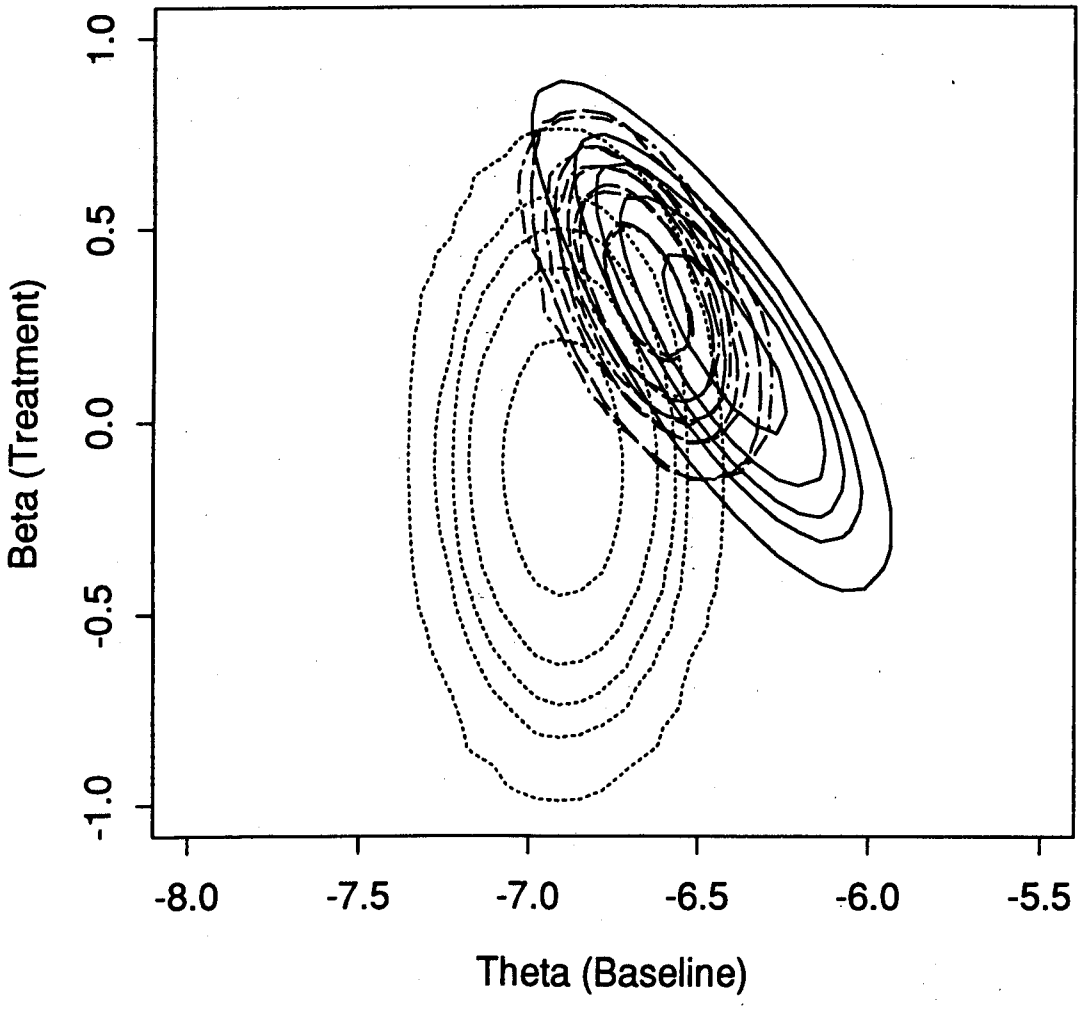
Key : —— maximum relative log-likelihood, ····· prior, — · — · — · — posterior (Laplace) and — — — — — — posterior (Gauss-Hermite); left and right vertical lines are at -0.26 and 0.0 respectively.

Figure 6.7: Maximum relative log-likelihood and marginal densities for survival model assuming constant baseline intensity, vague prior for θ and various priors for β.



Key : —— maximum relative log-likelihood, ····· prior, – · – · – · – posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.8: Maximum relative log-likelihood and marginal posterior densities for survival model assuming constant baseline intensity, meta prior for θ and clinical prior for β.



Key : —— maximum relative log-likelihood, ····· prior, – · – · – · – posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.9: Maximum relative joint log-likelihood and joint densities for survival model assuming constant baseline intensity, meta prior for θ and clinical prior for β, (contours are 50%, 80%, 90%, 95% and 99% credibility regions).

6.6 Piecewise Constant Intensity

In this section we extend the model (6.1) to the case when the baseline intensity is allowed to vary, in a constrained manner, with time. The simplest way that this can be achieved is to split the time scale into a finite number of epochs, and to allow the baseline intensity to be constant within each epoch.

6.6.1 General Model

One possible extension of the parametric form for the baseline intensity in (6.1) is to consider it to be piecewise constant. This means that we specify a set of time points, $\underline{\tau}$, for which the baseline intensity remains constant between consecutive points. We can therefore consider a multivariate counting process, $N(t) = \{N_{ij}(t), i \in I, j \in J, t > 0\}$, each component of which counts the number of events, in the case of death zero or one, that have occurred up to time t in the jth epoch. $N(t)$ is governed by a random intensity process $\alpha(t) = \{\alpha_{ij}(t), i \in I, j \in J, t > 0\}$. Associated with $N(t)$ is an at-risk process $Y(t) = \{Y_{ij}(t), i \in I, j \in J, t > 0\}$. Therefore (6.39) becomes

$$\alpha_{ij}(t) = \lambda(t|\theta_j)\, e^{\underline{\beta}_j^T \underline{z}_{ij}(t)}\, Y_{ij}(t) \quad \tau_j < t \leq \tau_{j+1} \quad j = 1, \cdots, m \tag{6.38}$$

and parameterising as in the constant case (6.38) becomes

$$\alpha_{ij}(t) = e^{\theta_j}\, e^{\underline{\beta}_j^T \underline{z}_{ij}(t)}\, Y_{ij}(t) \quad \tau_j < t \leq \tau_{j+1} \quad j = 1, \cdots, m \tag{6.39}$$

where typically $\tau_1 = 0$.

The log-likelihood can be written in the general case

$$\ell(\underline{\theta}, \underline{\beta}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left\{ \log_e[\alpha_{ij}(t)] dN_{ij}(t) - \int_0^t \alpha_{ij}(u)\, du \right\} \tag{6.40}$$

Substituting (6.39) into (6.29) yields a log-likelihood of the form

$$\ell(\underline{\theta}, \underline{\beta}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left\{ \log_e[e^{\theta_j}\, e^{\underline{\beta}^T \underline{z}_i(t)}\, Y_{ij}(t)] dN_{ij}(t) - \int_0^t e^{\theta_j}\, e^{\underline{\beta}^T \underline{z}_i(u)}\, Y_{ij}(u)\, du \right\} \tag{6.41}$$

6.6.2 Simple Model

We can consider the case when there are only two time intervals, and one regression parameter. As in the constant intensity model above in Section 6.5 this could be thought of as the treatment effect in the neutron therapy example.

Model (6.39) reduces to the simpler form

$$\alpha_{ij}(t) = e^{\theta_j}\, e^{\beta z_i}\, Y_{ij}(t) \quad \tau_j < t \leq \tau_{j+1} \quad j = 1, 2 \tag{6.42}$$

and similarly the log-likelihood (6.41) reduces to

$$\begin{aligned}\ell(\theta_1,\theta_2,\beta) &= \sum_{i=1}^{n}\sum_{j=1}^{2}\left\{\log_e[e^{\theta_j}\,e^{\beta z_i}\,Y_{ij}(t)]dN_{ij}(t) - \int_0^t e^{\theta_j}\,e^{\beta z_i}\,Y_{ij}(u)du\right\} \\ &= \sum_{ij}\theta_j dN_{ij}(t) + \beta\sum_{ij} z_i dN_{ij}(t) - \sum_{ij}\int_0^t e^{\theta_j}\,e^{\beta z_i}\,Y_{ij}(u)\,du\end{aligned}$$

As in the two parameter case the term $\sum_{ij}\log_e[Y_{ij}(t)]\,dN_{ij}(t)$ disappears in the log-likelihood since when $Y_{ij}(t) = 1$, $\log_e[Y_{ij}(t)]$ is zero, and when $Y_{ij}(t) = 0$, $dN_{ij}(t)$ is by definition zero. Similarly the likelihood is of the form

$$L(\theta_1,\theta_2,\beta) = e^{\sum_{ij}\theta_j dN_{ij}(t)}\,e^{\beta\sum_{ij} z_i dN_{ij}(t)}\,e^{\sum_{ij}\int_0^t e^{\theta_j}\,e^{\beta z_i}\,Y_{ij}(u)\,du} \tag{6.43}$$

In the case when we assume Uniform prior densities for all three parameters, θ_1, θ_2 and β the joint posterior will be proportional to the likelihood (6.43). However, we may wish to use an informative prior density, especially for the regression parameter, β. As we discussed earlier, a convenient distributional form for such a prior density would be the Normal distribution. Therefore in this case the posterior joint density, $p(\underline{\phi}|H)$ would take the form

$$p(\underline{\phi}|H) \propto e^{\sum_{ij}\theta_j dN_{ij}(t)}\,e^{\beta\sum_{ij} z_i dN_{ij}(t)}\,e^{\sum_{ij}\int_0^t e^{\theta_j}\,e^{\beta z_i}\,Y_{ij}(u)\,du}\,e^{-\frac{1}{2\sigma^2}(\beta-\mu)^2} \tag{6.44}$$

where μ and σ^2 are the prior mean and variance respectively, and are both assumed to be known.

As in Section 6.5, estimation of $\underline{\phi}$ could proceed using either Laplace approximations of Section 6.4.1 or Gauss-Hermite quadrature of Section 6.4.2. As with the two parameter model in Section 6.5, in this simple case we can write down the second derivatives. Optimisation requires an iterative search method, and as with the model in Section 6.5 we found that using `nlmin` in Splus (1990) was sufficient.

In order to use the Laplace approximations described in Section 6.4.1 and used above in the two parameter model of Section 6.5 we require the first and second derivatives of $G(\underline{\phi})$, the log posterior. In this three parameter problem the derivatives are straight forward to write down, and programming them is also straight forward. The first derivatives for a model in which all the prior densities are Uniform distributions are

$$\begin{aligned}\frac{\partial G(\underline{\phi})}{\partial\theta_1} &= \sum_i dN_{i1}(t) - \sum_i\int_0^t e^{\theta_1}\,e^{\beta z_i}\,Y_{i1}(u)\,du \\ \frac{\partial G(\underline{\phi})}{\partial\theta_2} &= \sum_i dN_{i2}(t) - \sum_i\int_0^t e^{\theta_2}\,e^{\beta z_i}\,Y_{i2}(u)\,du \\ \frac{\partial G(\underline{\phi})}{\partial\beta} &= \sum_{ij} z_i dN_{ij}(t) - \sum_{ij}\int_0^t z_i\,e^{\theta_j}\,e^{\beta z_j}\,Y_{ij}(u)\,du\end{aligned}$$

and the second derivatives of $G(\underline{\phi})$ are

$$\begin{aligned}
\frac{\partial^2 G(\underline{\phi})}{\partial \theta_1^2} &= -\sum_i \int_0^t e^{\theta_1} e^{\beta z_i} Y_{i1}(u)\, du \\
\frac{\partial^2 G(\underline{\phi})}{\partial \theta_1 \theta_2} &= 0 \\
\frac{\partial^2 G(\underline{\phi})}{\partial \theta_1 \beta} &= -\sum_i \int_0^t z_i\, e^{\theta_1} e^{\beta z_i} Y_{i1}(u)\, du \\
\frac{\partial^2 G(\underline{\phi})}{\partial \theta_2^2} &= -\sum_i \int_0^t e^{\theta_2} e^{\beta z_i} Y_{i2}(u)\, du \\
\frac{\partial^2 G(\underline{\phi})}{\partial \theta_2 \beta} &= -\sum_i \int_0^t z_i\, e^{\theta_2} e^{\beta z_i} Y_{i2}(u)\, du \\
\frac{\partial^2 G(\underline{\phi})}{\partial \beta^2} &= -\sum_{ij} \int_0^t z_i^2\, e^{\theta_j} e^{\beta z_j} Y_{ij}(u)\, du
\end{aligned}$$

6.6.3 Example

We consider applying the three parameter model described above to the neutron therapy data described in Section 2.2. In this case we assume $\underline{\tau} = (0, 365, \infty)$, the time grid, which from Figure 6.4 would seem reasonable as there is a slight reduction in the gradient of the cumulative hazards for both groups above 365 days.

Tables 6.6 and 6.7 show the parameter estimates, while Figures 6.10 and 6.11 show the marginal densities for this model using the clinical and meta prior densities for β. Table 6.8 shows the parameter estimates obtained using the model of Gamerman and West described briefly later in this chapter. In this case the evolution matrix H_j was set to be the identity matrix, so that there was no auto-regressive structure.

We can see from Tables 6.6 and 6.8 and Figures 6.10 and 6.11 that assuming a piecewise constant baseline hazard makes little difference to the overall results of the analysis. Using the clinical prior, the data, via the likelihood shifts the prior to the right, indicating that *a posteriori* there is little evidence that neutrons are beneficial for patients with pelvic tumours. Similarly using a meta prior, there is little difference between the prior and posterior densities for the treatment effect, as the prior density is located at the same point as the likelihood. Assuming the parameter estimates to be approximately Normally distributed, 95% credibility intervals for the difference $\theta_1 - \theta_2$ all include zero indicating that there is little evidence of a difference in baseline intensities below and above 365 days. The results obtained using SURVIVAL, Table 6.8, are in broad agreement with the model developed in Section 6.6.2, in that there is evidence that the baseline intensity is constant between the two intervals, and the clinical beliefs of Section 3.2.2 should be revised substantially in the light of the current data.

Table 6.7 shows covariances for the three parameter model. We can see that when vague prior densities are used for all three parameters posterior covariances are almost identical to those obtained using maximum likelihood, for both Laplace approximations and Gauss-Hermite quadrature. When either a clinical prior density or a meta prior density for the treatment effect was used there was a slight reduction, in absolute terms, in the posterior covariance compared to that under maximum likelihood, as would be expected.

	θ_1 (Baseline)		θ_2 (Baseline)		β (Neutrons)			
	Mean	SD	Mean	SD	Mean	SD	$P_{(0)}$[a]	$P_{(-0.26)}$
MLE	-6.687	0.181	-6.764	0.202	0.409	0.202	-	-
Prior	-	∞	-	∞	-	∞	-	-
Posterior (L)[b]	-6.702	0.182	-6.786	0.204	0.415	0.202	0.019	0.000
Posterior (G-H)[c]	-6.703	0.182	-6.786	0.203	0.415	0.202	0.019	0.000
Prior	-	∞	-	∞	-0.116	0.286	0.654	0.305
Posterior (L)	-6.585	0.158	-6.683	0.190	0.240	0.162	0.068	0.001
Posterior (G-H)	-6.585	0.158	-6.683	0.186	0.240	0.162	0.069	0.001
Prior	-	∞	-	∞	0.430	0.141	0.001	0.000
Posterior (L)	-6.706	0.141	-6.788	0.177	0.424	0.116	0.000	0.000
Posterior (G-H)	-6.706	0.143	-6.788	0.177	0.424	0.116	0.000	0.000

[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that β is less than 0 and -0.26 respectively.
[b] L denotes estimation using Laplace approximations.
[c] G-H denotes estimation using Gauss-Hermite quadrature.

Table 6.6: Parameter estimates for piecewise 2-state model with $\underline{\tau} = (0, 365, \infty)$ days, vague prior for θ_1, θ_2 and various priors for β, using survival data with 21st December 1990 as censoring date.

	$Cov(\theta_1, \theta_2)$	$Cov(\theta_1, \beta)$	$Cov(\theta_2, \beta)$
MLE	0.017	-0.028	-0.024
Vague Prior	-	-	-
Posterior (L)[a]	0.017	-0.027	-0.024
Posterior (G-H)[b]	0.017	-0.028	-0.025
Clinical Prior	-	-	-
Posterior (L)	0.010	-0.017	-0.014
Posterior (G-H)	0.009	-0.017	-0.015
Meta Prior	-	-	-
Posterior (L)	0.005	-0.009	-0.008
Posterior (G-H)	0.006	-0.009	-0.008

[a]L denotes estimation using Laplace approximations.
[b]G-H denotes estimation using Gauss-Hermite quadrature.

Table 6.7: Covariances for piecewise 2-state model with $\underline{\tau} = (0, 365, \infty)$ days for neutron data with censoring date 21st December 1990

	θ_j (Baseline)		β_{1j} (Neutrons)	
	Mean	SD	Mean	SD
Prior ($t = 0$)	0	10	0	10
$j = 1$ ($0 < t \leq 365$)	-6.594	-	0.269	-
$j = 2$ ($t > 365$)	-6.718	0.162	0.415	0.201
Prior ($t = 0$)	0	10	-0.116	0.286
$j = 1$ ($0 < t \leq 365$)	-6.513	-	0.135	-
$j = 2$ ($t > 365$)	-6.637	0.144	0.277	0.164
Prior ($t = 0$)	0	10	0.43	0.141
$j = 1$ ($0 < t \leq 365$)	-6.594	-	0.266	-
$j = 2$ ($t > 365$)	-6.660	0.122	0.322	0.115

Table 6.8: Parameter estimates for piecewise 2-state model with $\underline{\tau} = (0, 365, \infty)$ days, vague prior for baseline intensity and various priors for β, using SURVIVAL.

6.6.4 More Sophisticated Models

We wish now to consider a more complicated model, so that we may allow for more than two time periods and more than one covariate factor. Such a model is not only of interest in its own right, but also as a means of testing the assumptions of simpler models of Sections 6.5.1 and 6.6.2. The model that we will use in this section has three baseline parameters, separating the time scale into three intervals, and we consider the case when there are three explanatory factors. For the neutron therapy data set, this model would allow us to consider not only treatment but also the three sites cervix, bladder and rectum.

We first consider the likelihood for such a model, and then the estimation of the parameters by use of Laplace approximations (Section 6.4.1) and Gauss-Hermite quadrature (Section 6.4.2). For both the maximum likelihood estimation and that using Laplace approximations the first and second derivatives are required. In this more complicated setting we also make use of Numerical Algorithms Group (NAG) routines for the maximisation of the log-likelihood and posterior densities.

Model

Assuming proportional intensities, and maintaining the notation of the rest of the chapter, the baseline parameters are θ_1, θ_2 and θ_3, whilst the explanatory factors are β_1, β_2 and β_3. As before we let $\underline{\phi} = (\theta_1, \theta_2, \theta_3, \beta_1, \beta_2, \beta_3)$. Therefore,

$$\alpha_{ij}(t) = e^{\theta_j}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_{ij}(t) \quad \tau_j < t \leq \tau_{j+1} \quad j = 1,2,3 \tag{6.45}$$

Re-parameterising so that the baseline intensity in each time epoch is incremental, i.e $\theta_1^* = \theta_1$, $\theta_2^* = \theta_1 + \theta_2$ and $\theta_3^* = \theta_1 + \theta_2 + \theta_3$, yields

$$\alpha_{ij}(t|\underline{z}_i) = e^{\theta_j^*}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_{ij}(t) \quad \tau_j < t \leq \tau_{j+1} \quad j = 1,2,3 \tag{6.46}$$

where $\underline{z}_i = (z_{1i}, z_{2i}, z_{3i})^T$ representing a vector of covariates. This parameterisation allows us to consider the assumption of a time varying baseline intensity by looking at the marginal densities for θ_2 and θ_3.

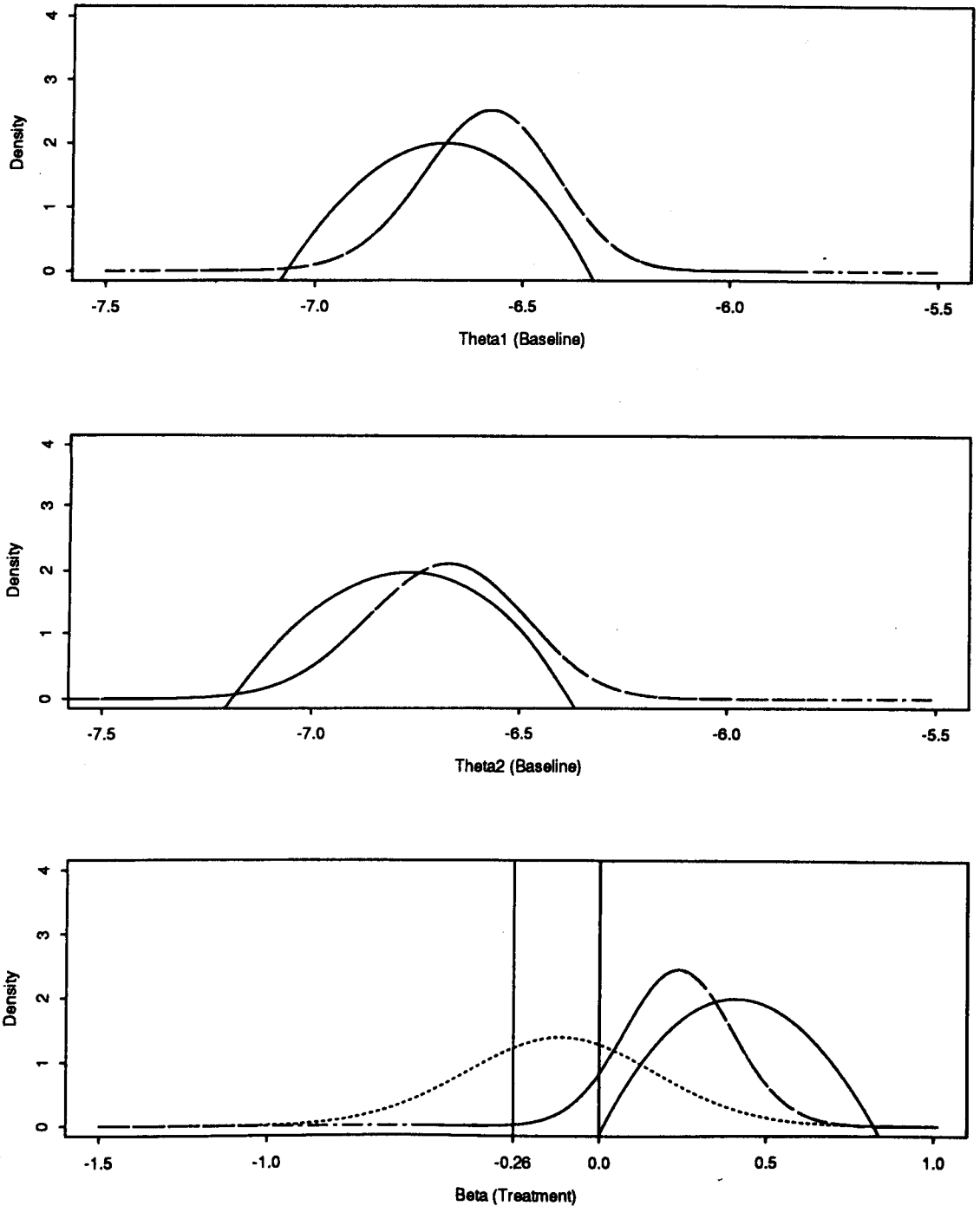
Log-Likelihood

The log-likelihood is given by substituting (6.46) into (6.40) which yields

$$\begin{aligned} \ell(\underline{\phi}) &= \sum_{i=1}^{n} \left\{ \log_e[e^{\theta_j^*}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_{ij}(t)]\, dN_{ij}(t) - \int_0^t e^{\theta_j^*}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_{ij}(u)\, du \right\} \\ &= \sum_{ij} \theta_j^*\, dN_{ij}(t) + \sum_{ij} \underline{\beta}^T \underline{z}_i\, dN_{ij}(t) - \sum_{ij} \int_0^t e^{\theta_j^*}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_{ij}(u)\, du \end{aligned} \tag{6.47}$$

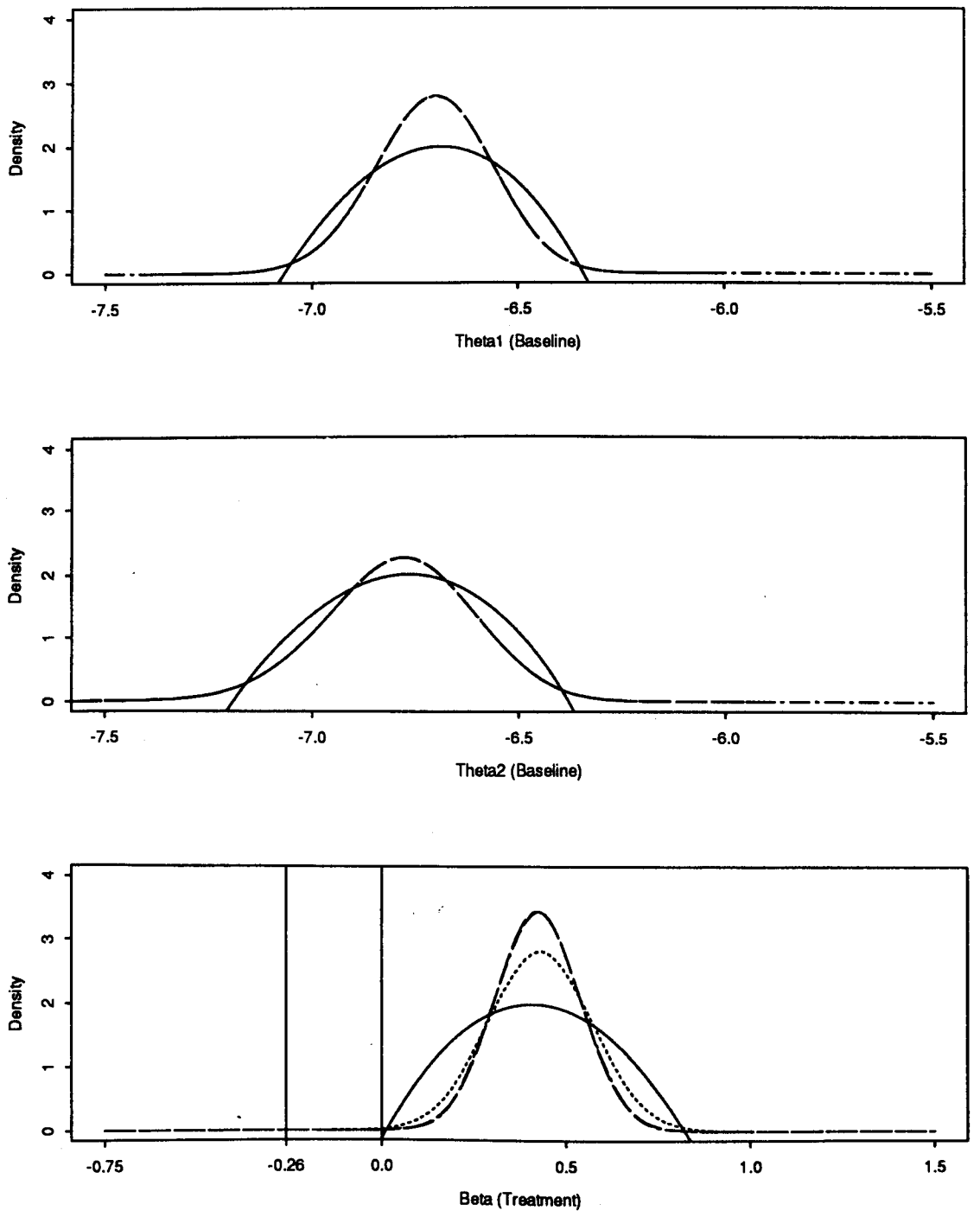
and the corresponding likelihood has the form

$$L(\underline{\phi}) = e^{\sum_{ij} \theta_j^*\, dN_{ij}(t)}\, e^{\sum_{ij} \underline{\beta}^T \underline{z}_i\, dN_{ij}(t)}\, e^{-\sum_{ij} \int_0^t e^{\theta_j^*}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_{ij}(u)\, du} \tag{6.48}$$



Key : — maximum relative log-likelihood, ····· clinical prior, – · – · – · – posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.10: Maximum relative log-likelihood and marginal densities for survival model assuming piecewise constant intensity, vague priors for θ_1, θ_2 and clinical prior for β.



Key : —— maximum relative log-likelihood, ····· clinical prior, – · – · – · – posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.11: Maximum relative log-likelihood and marginal densities for survival model assuming piecewise constant baseline intensity, vague priors for θ_1, θ_2 and meta prior for β.

Posterior Distributions

The joint posterior in the case when each of the models parameters has a Uniform prior density is proportional to (6.48). In the case when one of the regression parameters is a treatment effect there will often be prior information, and this can be expressed in terms of a Normal distribution as has been discussed above. In this case the joint posterior, $p(\underline{\phi}|H)$ is of the form

$$p(\underline{\phi}|H) \propto e^{\sum_{ij} \theta_j^* \, dN_{ij}(t)} e^{\sum_{ij} \underline{\beta}^T \underline{z}_i \, dN_{ij}(t)} e^{-\sum_{ij} \int_0^t e^{\theta_j^*} e^{\underline{\beta}^T \underline{z}_i} Y_{ij}(u) \, du} e^{-\frac{1}{2\sigma^2}(\beta_1 - \mu)^2} \quad (6.49)$$

where β_1 is the treatment effect parameter and the Normal distribution has mean and variance μ and σ^2 respectively, and are assumed known.

Estimation of Model Parameters

As in the case of the three parameter piecewise constant intensity model described above in Section 6.6.1 we can use both Laplace approximations and Gauss-Hermite quadrature to estimate $\underline{\phi}$. Using Laplace approximations we are required not only to maximise the log-posterior, which in this case is a function of six parameters, but also to evaluate the hessian at this point. For this model the `nlmin` function in Splus (1990) proved to be too slow. An alternative maximisation routine such as routine E04LBF in the Numerical Algorithms Group (NAG) library which performs both constrained and unconstrained optimisation of a function of several variables was used instead. This routine uses a modified Newton algorithm and requires the the first and second derivatives of the function to optimised, see NAG manual volume 4 for further details and references. The analytic form of the hessian was found using the MAPLE (1989) computer algebra package to do the differentiation of the log-posterior. This routine proved particularly useful since it allowed constrained optimisation of the log-posterior, which was useful in the case of obtaining the marginal posterior densities, since it meant that the problem did not have to be re-programmed for each model parameter separately.

Few problems were encountered in applying Gauss-Hermite quadrature to this particular model using BAYES4. The convergence criterion was the same as that used in the previous models. Time taken for convergence was greater than in the three parameter model, but was still acceptable, i.e less than 5 minutes.

6.6.5 Example

As in Section 6.5.3 we will consider the neutron therapy data described in Section 2.2. In this example we will restrict attention to only those patients in the neutron therapy study who had tumours of the rectum, bladder or cervix. In this case $n = 147$. We will assume that $\underline{\tau} = (0, 182, 365, \infty)$, and z_{1i}, z_{2i} and z_{3i} represent neutron treatment, cervix and rectum respectively. Using NAG routine E04LBF to maximise the log-likelihood and posterior densities in order

to apply Laplace approximations, and using the same convergence criteria as with the previous model for Gauss-Hermite quadrature, the following results where obtained.

Table 6.9 shows parameter estimates for model (6.46) using Uniform, clinical and meta priors for β_1 and Uniform priors for all the other parameters. Figures 6.12 and 6.13 show the marginal densities using either a clinical prior density for the treatment effect or a meta prior density.

We can see from Table 6.9 and Figures 6.12 and 6.13 that as with the three parameter model applied to the neutron therapy data the elicited clinical prior beliefs should be updated so that *a posteriori* they suggest that neutrons are not in fact beneficial for pelvic tumours. The meta prior density is modified only slightly in the light of the trial results.

Table 6.10 shows the results of using SURVIVAL[2] the computer package that implements the models of Gamerman and West (1987b), briefly mentioned in Section 4.3.2. We can see that the same trend is apparent using these models as it was using our models. That is clinical beliefs should be updated in the light of the trial results so to be much more in agreement with the data. Prior beliefs based on the results of previous studies would change little in the light of the current results.

We can also consider testing whether $\theta_1 = \theta_2 = \theta_3$, i.e whether there appears to be any evidence for a time varying baseline intensity. From Figures 6.12 and 6.13 we can see that the marginal posterior densities for both θ_2 and θ_3 are centred at points other than zero indicating evidence for varying baseline intensities. However, we can also see that the effect of the posterior densities for θ_2 and θ_3 is to cancel one another out. So, the baseline intensity increases between τ_2 and τ_3, but then reverts to its level of e^{θ_1} for times greater than τ_3.

[2]SURVIVAL can be obtained from Professor Mike West, Institute of Statistics and Decision Sciences, Duke University, North Carolina, USA.

	θ_1	θ_2 (Baseline)	θ_3	β_1 (Neutrons)	β_2 (Cervix)	β_3 (Rectum)
MLE	-6.712	0.469	-0.323	0.336	-0.661	0.045
	(0.098)	(0.209)	(0.237)	(0.204)	(0.284)	(0.212)
Prior	-	-	-	-	-	-
	(∞)	(∞)	(∞)	(∞)	(∞)	(∞)
Posterior (L)[a]	-6.740	0.470	-0.323	0.342	-0.683	0.043
	(0.244)	(0.240)	(0.238)	(0.207)	(0.299)	(0.212)
Posterior (G-H)[b]	-6.740	0.470	-0.323	0.342	-0.683	0.043
	(0.244)	(0.240)	(0.239)	(0.207)	(0.301)	(0.213)
Prior	-	-	-	-0.116	-	-
	(∞)	(∞)	(∞)	(0.286)	(∞)	(∞)
Posterior (L)	-6.635	0.465	-0.330	0.184	-0.714	0.054
	(0.226)	(0.239)	(0.238)	(0.166)	(0.298)	(0.212)
Posterior (G-H)	-6.635	0.465	-0.330	0.184	-0.714	0.054
	(0.225)	(0.240)	(0.239)	(0.166)	(0.301)	(0.213)
Prior	-	-	-	0.430	-	-
	(∞)	(∞)	(∞)	(0.141)	(∞)	(∞)
Posterior (L)	-6.778	0.472	-0.320	0.401	-0.671	0.039
	(0.215)	(0.240)	(0.238)	(0.117)	(0.296)	(0.212)
Posterior (G-H)	-6.778	0.472	-0.320	0.401	-0.671	0.039
	(0.214)	(0.240)	(0.239)	(0.117)	(0.300)	(0.212)

[a] L denotes estimation using Laplace approximations.
[b] G-H denotes estimation using Gauss-Hermite quadrature.

Table 6.9: Parameter estimates, means (standard deviations) for piecewise 2-state model with $\underline{\tau} = (0, 182, 365, \infty)$ days, and various priors for $\underline{\phi}$, using survival data with 21st December 1990 as censoring date.

	θ_j (Baseline)		β_{1j} (neutrons)		β_{2j} (cervix)		β_{3j} (rectum)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Prior ($t = 0$)	0	10	0	10	0	10	0	10
$j = 1$ ($0 < t \leq 182$)	-6.558	-	0.190	-	-1.980	-	0.107	-
$j = 2$ ($182 < t \leq 365$)	-6.307	-	0.246	-	-0.836	-	-0.259	-
$j = 3$ ($t > 365$)	-6.526	0.190	0.334	0.204	-0.625	0.295	0.053	0.212
Prior ($t = 0$)	0	10	-0.116	0.286	0	10	0	10
$j = 1$ ($0 < t \leq 182$)	-6.445	-	0.010	-	-2.048	-	0.114	-
$j = 2$ ($182 < t \leq 365$)	-6.210	-	0.084	-	-0.884	-	-0.247	-
$j = 3$ ($t > 365$)	-6.438	0.175	0.188	0.166	-0.673	0.295	0.061	0.212
Prior ($t = 0$)	0	10	0.43	0.141	0	10	0	10
$j = 1$ ($0 < t \leq 182$)	-6.608	-	0.283	-	-2.052	-	0.085	-
$j = 2$ ($182 < t \leq 365$)	-6.325	-	0.284	-	-0.878	-	-0.271	-
$j = 3$ ($t > 365$)	-6.508	0.159	0.312	0.116	-0.655	0.295	0.047	0.212

Table 6.10: Parameter estimates for piecewise 2-state model with $\underline{\tau} = (0, 182, 365, \infty)$ days, and various priors for $\underline{\phi}$, using SURVIVAL.

6.6.6 Other Approaches

A number of authors have considered Bayesian inference for models in which the baseline hazard is piecewise constant. Abrams (1989, Chapters 5 and 6) reviews many of these.

Cornfield and Detre (1977) considered the case when failure was a time-dependent Poisson process, such that the probability of failure in an interval of width h was $Y(t)\lambda(t)h + O(h)$, where $Y(t)$ denotes the number of patients at risk at time t and $\lambda(t)$ is the hazard function. They consider the case when there are m intervals, indexed by j, and the hazard function be constant within each interval, i.e λ_j, and that the number at risk in the jth interval is Y_j, and that the number of failures in the jth interval is x_j. The likelihood for $\underline{\lambda} = (\lambda_1, \cdots, \lambda_j, \cdots, \lambda_m)$ can be written as

$$L(\underline{\lambda}) = \prod_{j=1}^{m} e^{-\lambda_j Y_j h} \frac{(\lambda_j Y_j h)^{x_j}}{x_j!} + O(h) \tag{6.50}$$

Cornfield and Detre suggest independent Gamma prior distribution for $\underline{\lambda}$. Thus $\lambda_j h$ has a gamma distribution with parameters ν and ζ. Therefore the joint prior is of the form

$$p(\underline{\lambda} h|\nu, \zeta) = \prod_{j=1}^{m} \frac{\zeta^{\nu h}}{(\nu h - 1)!} (\nu h)^{-1} e^{-\zeta \lambda_j h} \tag{6.51}$$

Using Bayes' Theorem to combine (6.50) and (6.51) yields a joint posterior density that is of the form

$$p(\underline{\lambda}|H) \propto \prod_{j=1}^{m} \lambda_j^{x_j + \nu h - 1} e^{-\lambda_j h(Y_j + \zeta)} + O(h) \tag{6.52}$$

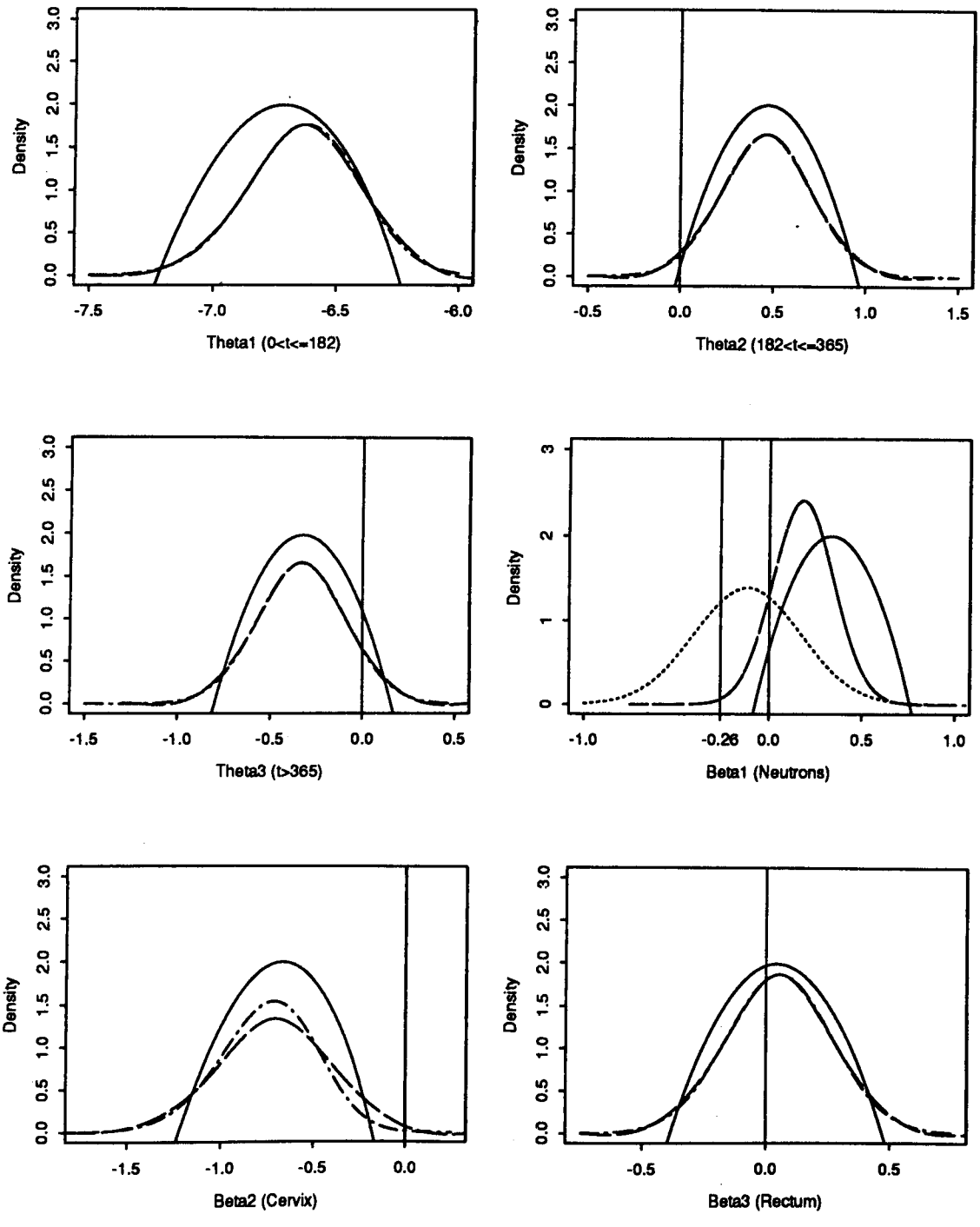
Kalbfleisch and MacKay (1978) point out that the approximation in (6.50) will not necessarily remain true throughout the prior to posterior analysis, and that the joint posterior (6.52) does not have an error of $O(h)$. Kalbfleisch and MacKay suggest a correction term for (6.52).

Cornfield and Detre also consider the case when there are two groups of patients, and that the hazard rates in the two groups are related so that $\lambda_2(t) = \phi\lambda_1(t)$. Assuming that the hazard $\lambda_1(t)$ is of the same form as before and that we also have a prior density of the same form, and independently we assume that we can specify a prior density, $g(\underline{\phi})$, for ϕ, the joint posterior is of the form

$$p(\underline{\lambda}, \phi) \propto g(\phi) \prod_{j=1}^{m} \phi^{x_{2j}} \lambda_{1j}^{x_{1j} + x_{2j} + \nu h - 1} e^{-\lambda_{1j} h(Y_{2j}\phi + Y_{1j} + \zeta)} \tag{6.53}$$

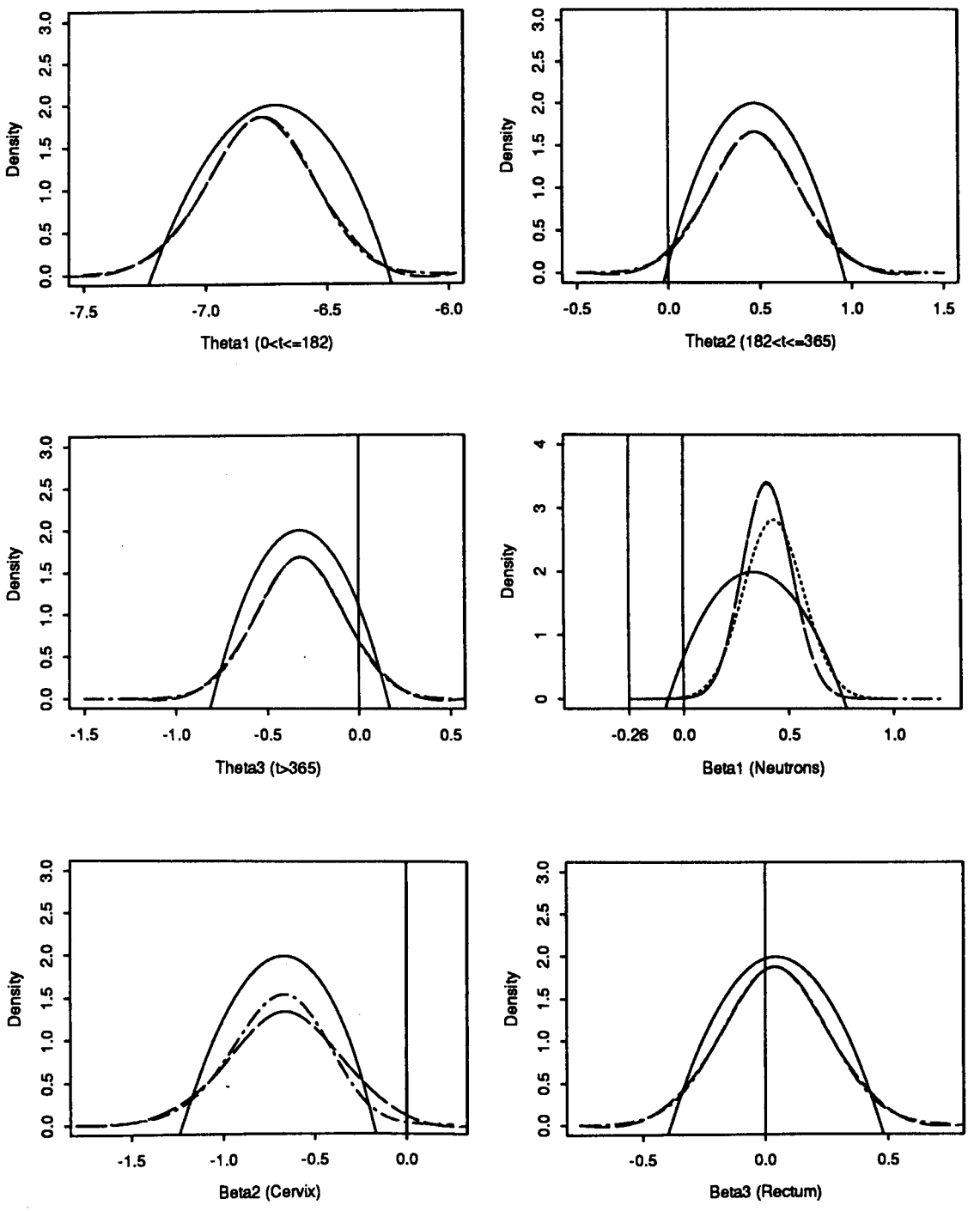
Integrating $\underline{\lambda}$ out of (6.53) yields a marginal posterior density for ϕ of the form

$$p(\phi) \propto g(\phi) \prod_{j=1}^{m} \frac{\phi^{x_{2j}}}{(x_{1j} + x_{2j})(Y_{2j}\phi + Y_{1j} + \zeta)} + \nu h + O(h) \tag{6.54}$$



Key : — maximum relative log-likelihood, ····· clinical prior, – · – · – · – posterior (Laplace) and – – – – – – posterior (Gauss-Hermite)

Figure 6.12: Maximum relative log-likelihoods and marginal densities for survival model assuming piecewise exponential baseline intensity with vague priors for $\underline{\theta}$, β_2 and β_3, and clinical prior for β_1.



Key : —— maximum relative log-likelihood, ····· clinical prior, — · — · — · — posterior (Laplace) and — — — — — — posterior (Gauss-Hermite)

Figure 6.13: Maximum relative log-likelihoods and marginal densities for survival model assuming piecewise exponential baseline intensity with vague priors for $\underline{\theta}$, β_2 and β_3, and meta prior for β_1.

Kalbfleisch (1978) has also proposed a Bayesian survival model in which the baseline hazard is considered to comprise a finite number of intervals in which the hazard is constant. A similar model has also been proposed by Burridge (1980). In this model the cumulative hazard is assumed to have an independent increments Gamma process prior distribution. The mean of $\Lambda_0(t)$, the cumulative hazard, is $\Lambda^*(t)$ a known function, and the variance is $\Lambda^*(t)/c$. The increments, $d\Lambda_0(t)$, have a Gamma distribution with shape and scale parameters $cd\Lambda^*(t)$ and c respectively. This prior form has the advantage that, conditional on $\underline{\beta}$, it is conjugate for censored survival data, as we saw in Section 6.4.1, and also more generally for counting processes, (Aven, 1986). Therefore, conditional on $\underline{\beta}$, the posterior gamma distribution for $d\Lambda_0(t)$ has shape and scale parameters $cd\Lambda^*(t)+dN(t)$ and $c+\sum_{i=1}^{n} Y_i(t)\, e^{\underline{\beta}^T \underline{Z}_i(t)}$ respectively, where $dN(t)$ denotes the total number of events.

Kalbfleisch (1978) mentions that often $\Lambda^*(t)$ is really a nuisance parameter, and interest focuses on the regression parameters, $\underline{\beta}$. Integrating the joint posterior distribution with respect to $\Lambda_0(t)$ yields a marginal posterior for $\underline{\beta}$ that is proportional to

$$\prod_{t\geq 0} \frac{\prod_i \left[Y_i(t)\, e^{\underline{\beta}^T \underline{Z}_i(t)}\right]^{dN_i(t)}}{\left[c+\sum_i Y_i(t)\, e^{\underline{\beta}^T \underline{Z}_i(t)}\right]^{cd\Lambda^*(t)+dN(t)}} \tag{6.55}$$

With the improper vague prior for $\Lambda_0(t)$ corresponding to $c=0$, the posterior distribution for $\underline{\beta}$, (6.55), is then proportional to Cox's partial likelihood. Clayton (1991) notes when $\Lambda_0(t)$ is considered parametric difficulties arise and no simple solution exists. However, throughout this analysis a Uniform prior on $\underline{\beta}$ is assumed, which seems unrealistic, since there is very often information about the relative effects of treatment or other patient characteristics, as we saw in Sections 3.2 and 3.3.

Gamerman and West (1987a, 1987b, 1991) have also considered the case when the time scale is made up of a finite number of intervals. The main difference between their approach and that of Cornfield and Detre (1977), Kalbfleisch (1978) and the one outlined above, is that they assume the regression parameters to vary over time, i.e be dynamic, and that there is an auto-regressive structure for the priors. This model is an extension of the Dynamic Generalised Linear Model (DGLM) introduced in Section 4.3.2 in which survival times were ignored. The model has the form

$$\begin{aligned} \lambda_i(t) &= e^{\underline{\beta}_j^T \underline{Z}_i} \qquad \tau_j < t \leq \tau_{j+1} \\ \underline{\beta}_j &= H_j(b_j)\underline{\beta}_{j-1} + \omega_j \qquad (6.56) \\ \omega_j &\sim [0, W_j] \qquad (6.57) \end{aligned}$$

where $\underline{\beta}$ are unknown parameters that need to be estimated. H_j is the system evolution matrix and is typically dependent on b_j, the length of the jth interval. ω_j are errors with zero mean and covariance matrix W_j. In many situations H_j

can be the identity matrix resulting in the parameters following a simple random walk. W_j describes the degree of uncertainty with which the parameters evolve from time t_{j-1} to t_j. W_j is often a function of the interval length, b_j, and in the limit as W_j tends to zero the static model is obtained, whereas a larger value of W_j allows greater freedom for the βs to change over time.

At the $(j-1)$th time point the posterior distribution for β_{j-1} can be partially specified in terms of a mean m_{j-1} and a covariance matrix C_{j-1}. Thus

$$[\beta_{j-1}|D_{j-1}] \sim [m_{j-1}, C_{j-1}] \tag{6.58}$$

where D_{j-1} is the information set containing all the information up to time T_{j-1}. The prior distribution for β_j is

$$[\beta_j|D_{j-1}] \sim [a_j, P_j] \tag{6.59}$$

where

$$\begin{aligned} a_j &= H_j(b_j)m_{j-1} \\ P_j &= H_j(b_j)C_{j-1}H_j^T(b_j) + W_j \end{aligned}$$

The likelihood factorises into a term for each interval and Gamerman (1991) obtains the jth component of the overall likelihood as

$$\prod_{i=1}^{r_j} [\lambda_{ij}]^{\delta_{ij}} e^{-\lambda_{ij}(t_{ij}-t_{j-1})} \tag{6.60}$$

where r_j is the number of patients alive at the beginning of the jth interval, δ_{ij} is an indicator of death for the ith patient in the jth interval, and t_{ij} is the survival time of the ith patient in the jth interval.

Estimation of the model parameters, i.e the βs, can proceed once an initial input is given, i.e $[\beta_1|D_0]$ and the errors, ω_j, have been specified, by cycling through equations (6.56), (6.58), (6.59) and (6.60). This estimation procedure will involve a high dimensional integration. Gamerman (1991) suggests using linear Bayes methods. This relies on the fact that, for the likelihood (6.60), a Gamma distribution is a conjugate prior for λ_j. Linear Bayesian methods can then be used to obtain the corresponding means and variances for the β_js. Further details can found in West, Harrison and Migon (1985), West and Harrison (1989) and Gamerman (1991).

A number of authors have described the link between parametric survival models and log-linear models, most notably Lawless (1987) and Clayton (1988). In the case of a piecewise constant intensity model this corresponds to a multi-way table defined by the time grid, $\underline{\tau}$, and regression factors $\underline{z}$. The methods that we have outlined for piecewise constant intensity model are equivalent to putting independent Normal or Uniform prior densities on each of the margins of this table defined by $\underline{z}$.

6.7 Weibull Intensity Model

In this section we consider the multiplicative intensity model suggested by Aalen (1978) (5.12) but with $\lambda(t|\underline{\theta})$ having a parametric form which is a power transformation of the time scale. In survival analysis such a parametric form corresponds to the survival times following a Weibull distribution. Therefore we say that the intensity has a Weibull parametric form.

The Weibull distribution has been used extensively in reliability and industrial life-testing, see Davidson (1988) and Crowder *et al* (1991) for reviews, but has not been used extensively in medical applications with the exception of Peto and Lee (1973), Prentice (1975) and Aitkin and Clayton (1980). Other authors have considered it to be too restrictive and unrealistic. However, as it generalises the constant intensity model (6.28) of Section 6.5 it adds an extra level of sophistication to the models so far considered.

A number of authors have considered a Bayesian approach to survival data when the survival times can be assumed to follow a Weibull distribution, Abrams (1989) Chapter 4 reviews many of these. These approaches can be classified into two groups. The first approach is to work directly with the Weibull distribution. Unlike the exponential distribution there is no conjugate prior, although if we assume that one of the distribution's parameters is fixed, the index parameter, then the problem reduces to that of a constant intensity discussed in Section 6.5. However, if we are prepared to use either asymptotic approximation methods or numerical methods to estimate parameters, the Weibull distribution poses no particular problems. This approach has been considered by Bhattacharya (1967), Canavos and Tsokos (1973), Martz and Waller (1982), Singpurwalla and Song (1987) and Dellaportas and Wright (1991a). The other approach is to transform the survival times by taking the logarithms of them, and treating these transformed times as the dependent variable in a linear regression model. The errors in such a model have extreme value distributions. This approach has been considered by Achcar *et al* (1985, 1987) and Sweeting (1981, 1982, 1987).

6.7.1 General Model

The Weibull distribution is a generalisation of the exponential distribution and allows a power dependence of the hazard on time. The two parameter version of the distribution leads to a hazard function of the form

$$\lambda(t) = \rho\,\alpha\,(\rho\,t)^{\alpha-1} \tag{6.61}$$

where ρ is a scale parameter and α is an index parameter, such that $\rho > 0$ and $\alpha > 0$. The hazard, $\lambda(t)$, is monotone decreasing for $\alpha < 1$, increasing for $\alpha > 1$ and reduces to the exponential for $\alpha = 1$. As ρ is required to be non-negative we may re-parameterise the model by replacing ρ^{α} by e^{θ_2}. We now need to estimate θ_2 rather than ρ, and e^{θ_2} satisfies the condition that it is always greater than zero. The hazard function now becomes

$$\lambda(t) = \alpha\,t^{\alpha-1}\,e^{\theta_2}$$

As α is also required to be greater than zero we need to replace α by a term such as e^{θ_1}, and estimate θ_1 instead. Thus, the hazard now is of the form

$$\lambda(t) = e^{\theta_1}\, t^{e^{\theta_1}-1}\, e^{\theta_2}$$

One feature of the Weibull distribution is that if we assume that the failure times of a group of patients are distributed with a Weibull distribution then the accelerated life model and the proportional hazards model coincide. Here we will develop a model from the proportional hazards view. The model is $\lambda(t) = \lambda_0(t)e^{\underline{\beta}'\underline{Z}}$, where $\lambda_0(t)$ has the form of a Weibull distribution, and under our parameterisation is $e^{\theta_1}t^{e^{\theta_1}-1}e^{\theta_2}$. Therefore,

$$\lambda(t) = e^{\theta_1}t^{e^{\theta_1}-1}e^{\theta_2}e^{\underline{\beta}'\underline{Z}} \tag{6.62}$$

If we consider a counting process, $N(t)$, such that each patient can have at most one event, for example death, and that associated with $N(t)$ is an observational process, $Y(t)$, which indicates whether a particular patient is at risk at a specific time, t, then the model becomes

$$\alpha_i(t) = e^{\theta_1}\, t^{e^{\theta_1}-1}\, e^{\theta_2} e^{\underline{\beta}^T \underline{z}_i}\, Y_i(t) \tag{6.63}$$

Substituting (6.63) into (5.14) yields a log-likelihood of the following form

$$\begin{aligned}\ell(\underline{\theta},\underline{\beta}) \;=\; & \theta_1 \sum_{i=1}^{n} dN_i(t) + (e^{\theta_1}-1)\sum_{i=1}^{n} dN_i(t)\log(t) + \theta_2 \sum_{i=1}^{n} dN_i(t) + \sum_{i=1}^{n} \underline{\beta}^T \underline{z}_i dN_i(t) \\ & - \sum_{i=1}^{n} \int_0^t e^{\theta_1}\, u^{e^{\theta_1}-1}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_i(u)\, du \end{aligned} \tag{6.64}$$

Notice now that a value of zero for θ_1 indicates that the exponential distribution would be adequate for the specific problem. The corresponding likelihood function is

$$\begin{aligned} L(\underline{\theta},\underline{\beta}) \;=\; & e^{\theta_1 \sum_{i=1}^n dN_i(t)}\, e^{(e^{\theta_1}-1)\sum_{i=1}^n dN_i(t)\log(t_i)}\, e^{\theta_2 \sum_{i=1}^n dN_i(t)} \\ & e^{\sum_{i=1}^n \underline{\beta}^T \underline{z}_i dN_i(t)}\, e^{-\sum_{i=1}^n \int_0^t e^{\theta_1}\, u^{e^{\theta_1}-1}\, e^{\underline{\beta}^T \underline{z}_i}\, Y_i(u)\, du} \end{aligned} \tag{6.65}$$

In the case when we consider vague prior information about all the parameters, the joint posterior density, $p(\underline{\phi}|H)$ is proportional to (6.65). In the case when there is prior information about one or more of the model parameters the joint posterior will be proportional to the product of the prior densities and (6.65). We can see that in the case when we assume that both θ_1 and θ_2 are unknown, no conjugate analysis is possible, and parameter estimation requires one of the estimation methods described in Section 6.3.

6.7.2 A Simple Model

As an example consider the case when there is only one regression parameter in the model. This could represent a treatment effect, such as that in the neutron therapy trial. Model (6.63) becomes

$$\alpha_i(t) = e^{\theta_1}\, t^{e^{\theta_1}-1}\, e^{\theta_2} e^{\beta z_i}\, Y_i(t) \tag{6.66}$$

The log-likelihood, (6.64) simplifies to

$$\begin{aligned}\ell(\underline{\theta},\beta) &= \theta_1 \sum_{i=1}^{n} dN_i(t) + (e^{\theta_1}-1)\sum_{i=1}^{n} dN_i(t)\log(t_i) + \theta_2 \sum_{i=1}^{n} dN_i(t) \\ &\quad + \beta \sum_{i=1}^{n} z_i dN_i(t) - \sum_{i=1}^{n} \int_0^t e^{\theta_1}\, u^{e^{\theta_1}-1}\, e^{\beta z_i}\, Y_i(u)\, du \qquad (6.67)\end{aligned}$$

and the likelihood (6.65) similarly simplifies to

$$\begin{aligned}L(\underline{\theta},\beta) &= e^{\theta_1 \sum_{i=1}^{n} dN_i(t)}\, e^{(e^{\theta_1}-1)\sum_{i=1}^{n} dN_i(t)\log(t_i)}\, e^{\theta_2 \sum_{i=1}^{n} dN_i(t)} \\ &\quad e^{\beta \sum_{i=1}^{n} z_i dN_i(t)}\, e^{-\sum_{i=1}^{n} \int_0^t e^{\theta_1}\, u^{e^{\theta_1}-1}\, e^{\beta z_i}\, Y_i(u)\, du} \qquad (6.68)\end{aligned}$$

As above when we assume Uniform prior densities for $\underline{\phi}$ the joint posterior is proportional to (6.65). In the case when there is prior information about the treatment effect, this is most conveniently summarised by a Normal distribution and the joint posterior is of the form

$$\begin{aligned}p(\underline{\phi}|H) \propto\ & e^{\theta_1 \sum_{i=1}^{n} dN_i(t)}\, e^{(e^{\theta_1}-1)\sum_{i=1}^{n} dN_i(t)\log(t_i)}\, e^{\theta_2 \sum_{i=1}^{n} dN_i(t)}\, e^{\beta \sum_{i=1}^{n} z_i dN_i(t)} \\ & e^{-\sum_{i=1}^{n} \int_0^t e^{\theta_1}\, u^{e^{\theta_1}-1}\, e^{\beta z_i}\, Y_i(u)\, du}\, e^{-\frac{1}{2\sigma^2}(\beta-\mu)^2} \qquad (6.69)\end{aligned}$$

where μ and σ^2 are the prior mean and variance of β.

In order that we may make use of both Laplace approximations described above we require the first and second derivatives of the log-posterior, and functions of it. In the case when we assume that all the parameters have Uniform prior densities the first order partial derivatives are

$$\begin{aligned}\frac{\partial \ell}{\partial \theta_1} &= \sum_{i=1}^{n} dN_i(t) + e^{\theta_1}\sum_{i=1}^{n} dN_i(t)\log_e(t) - e^{\theta_2}\sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\theta_1}\log(t_i)(e^{\beta}z_i + (1-z_i))Y_i(t_i) \\ \frac{\partial \ell}{\partial \theta_2} &= \sum_{i=1}^{n} dN_i(t_i)Y_i(t_i) - e^{\theta_2}\sum_{i=1}^{n} t_i^{e^{\theta_1}}(e^{\beta}z_i + (1-z_i))Y_i(t_i) \\ \frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^{n} dN_i(t_i)z_i Y_i(t_i) - e^{\theta_2}\sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\beta} z_i Y_i(t_i)\end{aligned}$$

The second order partial derivatives are

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \theta_1^2} &= e^{\theta_1} \sum_{i=1}^{n} dN_i(t_i) Y_i(t_i) + e^{\theta_1} \sum_{i=1}^{n} dN_i(t_i) \log(t_i) Y_i(t_i) - \\
&\quad e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\theta_1} \log(t_i)(e^{\beta} z_i + (1 - z_i)) Y_i(t_i) \\
&\quad - e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\theta_1^2} \log(t_i)^2 (e^{\beta} z_i + (1 - z_i)) Y_i(t_i) \\
\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} &= -e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\theta_1} \log(t_i)(e^{\beta} z_i + (1 - z_i)) Y_i(t_i) \\
\frac{\partial^2 \ell}{\partial \theta_1 \partial \beta} &= -e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\theta_1} \log(t_i) e^{\beta} z_i Y_i(t_i) \\
\frac{\partial^2 \ell}{\partial \theta_2^2} &= -e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} (e^{\beta} z_i + (1 - z_i)) Y_i(t_i) \\
\frac{\partial^2 \ell}{\partial \theta_2 \partial \beta} &= -e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\beta} z_i Y_i(t_i) \\
\frac{\partial^2 \ell}{\partial \beta^2} &= -e^{\theta_2} \sum_{i=1}^{n} t_i^{e^{\theta_1}} e^{\beta} z_i Y_i(t_i)
\end{aligned}$$

We found that the `nlmin` function in Splus (1990) did optimise (6.69), and (6.65) in the case of maximum likelihood estimates. This model converged quickly using BAYES4 (1985) using the same convergence criteria as with the simple constant intensity model (6.28) of Section 6.5.

The model in (6.63) and the log-likelihood in (6.64) have been parameterised in a way that is convenient for parameter estimation using the `nlmin` function in Splus (1990), which can perform only unconstrained optimisation. Such a parameterisation, though practical, may not necessarily be the most efficient. Hills and Smith (1992) have shown how 'good' parameterisation is helpful for Bayesian estimation in a number of models.

Cox and Reid (1987) discuss the implications of parameter orthogonality in detail. They show that when parameters are orthogonal, maximum likelihood estimates change only slightly when other parameters change. Though they discussed this from a classical point of view it has implications for Bayesian estimation methods. Using both Laplace approximations and Gauss-Hermite quadrature the nearer to multivariate Normality the likelihood is the more likely our estimates are to remain 'stable', regardless of nuisance parameters.

Cox and Reid (1987) also suggest a method for inducing parameter orthogonality, much in the vein of Jeffreys (1961). In particular they consider the case of the Weibull distribution when the index of the distribution is the parameter of interest. Unfortunately the re-parameterisation that they suggest would require a constrained optimisation in order to maximise the log-posterior. More importantly a key assumption made by Cox and Reid (1987) was that the likelihood was proportional to the density function, and whilst this is true when

we observe all the times of death, when there are censored observations the likelihood no is longer proportional to the density, instead it becomes a mixture of the density and the survivor function as we have seen above.

6.7.3 Example

As an illustration of this two-state model, we will apply it to the neutron therapy data described in detail in Section 2.2. As a comparison for the results obtained using the Bayesian model described above in Section 6.6.2, the maximum likelihood estimates for the model parameters were also calculated. These were obtained using the `nlmin` function in Splus (1990) to maximise the log-likelihood (6.67).

We consider the case when there is vague prior information about the shape parameter, θ_1, and the other baseline parameter θ_2, and we therefore assume that they each have a Uniform prior density. In the case of β we consider three different situations; when there is relatively little prior information about the hazard ratio, when clinical opinion about the relative treatment merits has been elicited, and finally when there is information available from previous studies. In the latter two cases we assume that the prior density for β is Normally distributed and that all the parameters are *a priori* independent. Sections 3.2 and 3.3 give more details about the elicitation of the clinical prior and the use of previous trial results. The clinical prior has a mean of -0.116 and a variance of 0.082, while the prior based on the previous trial results has a mean of 0.43 and a variance of 0.02.

Applying this model with the priors outlined above to the Neutron study data we obtain the parameter estimates in Table 6.11. A value for θ_1 of zero would indicate that the exponential distribution is adequate. From Table 6.11 we can see that in the case when we use independent Uniform priors for all three parameters the posterior estimates of the mean and standard deviation are very similar to those obtained via maximum likelihood as we would expect. For the clinical prior the estimates of the parameter means for θ_1 and θ_2 change slightly but the standard deviations remain approximately the same. Since we used Uniform priors for both of these parameters we would not expect them to change greatly, as this would only be through their correlation with β, the treatment parameter, on which a relatively strong prior was placed. When we consider what has happened to the posterior estimate for β we see that the mean has been moved closer to zero, and that there has been a considerable reduction in the standard deviation. There was also a considerable reduction in the probability that β was less than 0, and similarly less than -0.26, the clinically relevant difference. Finally, in the case of the prior obtained from previous trial results, there has been a slight shift in the β point estimate towards the maximum likelihood estimate, and again a considerable reduction in the standard deviation. This is to be expected in the case of this prior even more so than with the clinical prior as the standard deviation was much smaller, i.e 0.141 compared to 0.286. As with both the other priors the point estimates

and the estimates of the standard deviations of θ_1 and θ_2 are approximately the same as the maximum likelihood, as expected.

Table 6.12 shows the covariances for the three different prior formulations. The posterior covariances, especially for (θ_1, θ_2) and (θ_1, β) are approximately the same as those obtained using maximum likelihood. This is to be expected since we have assumed that the priors for each of the parameters are independent, and therefore any correlation is due to the effect of the likelihood. We can see from the maximum likelihood covariances that there is little evidence of strong relationship between either θ_1 or θ_2 with β. The only noticeable difference occurs when the prior for β is informative. In this case the posterior covariances are smaller in absolute terms that those obtained via maximum likelihood.

Figures 6.15 and 6.17 show individual marginal prior and posterior densities for each of the parameters. We can see from Table 6.12 and Figures 6.15 and 6.17 that there is very close agreement between the results obtained using Laplace approximations and those obtained via Gauss-Hermite quadrature. Both estimation methods worked well reflecting the fact that the model was relatively well behaved. Figure 6.16 displays the bivariate densities for the case when we assume Uniform prior densities for θ_1 and θ_2 and a prior density based on elicited clinical beliefs for β, the treatment effect. From Figure 6.16 we can see that although there is evidence of strong correlation between θ_1 and θ_2, the joint posterior densities for θ_1 and β, and θ_2 and β appear to be approximately bivariate Normal densities. This indicates that parameter estimates for β should be 'stable', since poor estimation of θ_1 and θ_2 would have little influence on β, (Cox and Reid, 1987, Hills and Smith, 1992).

Table 6.11 and Figures 6.15 and 6.17 show that there is considerable evidence to suggest that θ_1 is not zero. This indicates that an intensity model in which the baseline intensity has a Weibull parametric form is preferable to one in which it is constant. This can also be seen in Figure 6.14 which shows the log-log survivor functions for the two treatment groups plotted against log time. These lines can be seen to have approximately constant gradients indicating the appropriateness of the Weibull assumption (Kalbfleisch and Prentice, 1980).

6.7.4 Other Approaches

In this section we consider briefly a number of other approaches that have been taken to the problem of survival models when the survival times are assumed to come from a Weibull distribution. We first consider those models that like the one we have described above in which the data are not transformed to a linear model.

A simple case is when $\theta_1 = 0$ and a constant hazard is obtained and the methods of the above section on constant intensity models can be applied. In the case when both ρ and α in (6.61) are considered unknown Martz and Waller (1982) addressed the case when α was assumed to have a Uniform prior density and ρ had either a Uniform density or an Inverted Gamma density. In both situ-

ations numerical integration techniques or approximation methods are required to obtain the corresponding posterior densities.

Singpurwalla and Song (1987) have considered estimation of the survivor function (or reliability function in industrial terminology) when an *expert* expresses their belief about the median life to an *analyst* who then specifies the prior distributions based on the *expert's* beliefs. They did not consider the case when covariate measurements may be available. They do however use the Laplace approximations suggested and described by Tierney and Kadane (1986) and note that these perform particularly well for the Weibull distribution.

Sweeting (1981, 1982, 1987) considers the case of a 'location-scale' regression model in which the response is the logarithm of the survival times, and the corresponding errors have an Extreme Value distribution. Algebraically,

$$\log_e[t_i] = \underline{\beta}^T \underline{z}_i + \sigma\epsilon_i \qquad i = 1, \cdots, n \tag{6.70}$$

where t_i are the survival times, $\underline{z}_i$ is the covariate vector, and ϵ_i are error terms with an extreme value distribution and σ is a scale parameter, such that $\sigma > 0$. Sweeting (1987) generalises the results for a general location-scale regression model described in Sweeting (1984) to the case when there are censored observations. These approximations mean that the scale parameter, σ, has an approximate χ^2 distribution, and the regression parameters, $\underline{\beta}$, have approximate multivariate t distributions.

Achcar, Brookmeyer and Hunter (1985) and Achcar and Bolfarine and Pericchi (1987) also consider the case when survival times follow a Weibull distribution, but can be transformed so that they follow an extreme value distribution. Laplace approximations (Tierney and Kadane, 1986) are then used to estimate the the distribution parameters. They do not consider a regression structure as Sweeting (1981) does.

An important point to notice about all the other Bayesian approaches utilising the Weibull distribution is that none can easily be extended to the case when there are time-dependent covariates or strata.

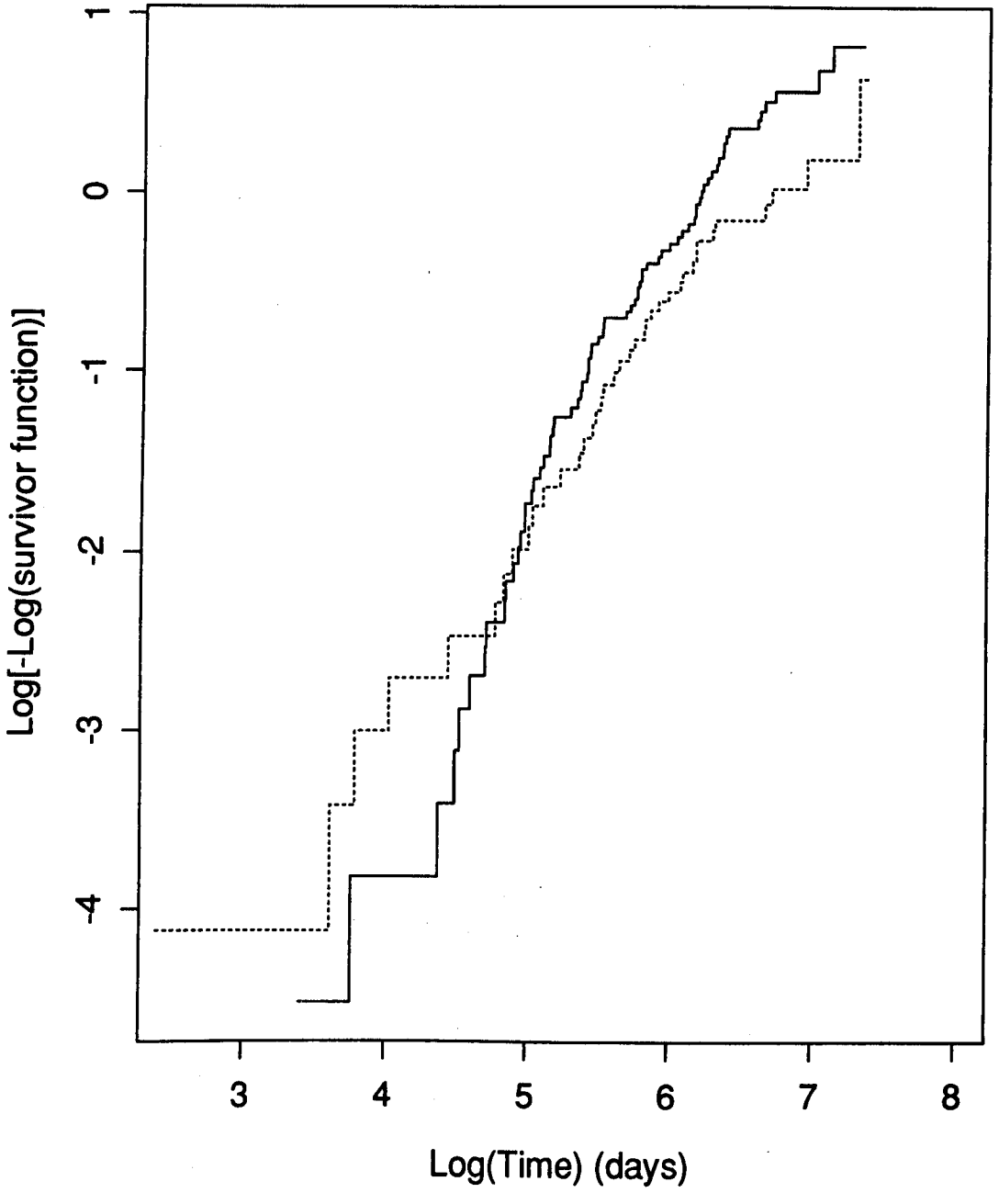
	θ_1 (Shape)		θ_2 (Baseline)		β (Neutrons)			
	Mean	SD	Mean	SD	Mean	SD	$P_{(0)}$ [a]	$P_{(-0.26)}$
MLE	0.140	0.077	-7.695	0.603	0.443	0.202	-	-
Prior	-	∞	-	∞	-	∞	-	-
Posterior (L)[b]	0.135	0.077	-7.698	0.601	0.449	0.202	0.011	0.000
Posterior (G-H)[c]	0.135	0.077	-7.698	0.601	0.449	0.202	0.011	0.000
Prior	-	∞	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)	0.129	0.078	-7.536	0.590	0.263	0.162	0.052	0.001
Posterior (G-H)	0.129	0.077	-7.536	0.589	0.263	0.162	0.052	0.001
Prior	-	∞	-	∞	0.430	0.141	0.001	0.000
Posterior (L)	0.134	0.075	-7.683	0.585	0.435	0.116	0.000	0.000
Posterior (G-H)	0.134	0.076	-7.683	0.585	0.435	0.117	0.000	0.000

[a] $P_{(0)}$ and $P_{(-0.26)}$ represent the probability that $\beta < 0$ and $\beta < -0.26$ respectively
[b] L denotes estimation using Laplace approximations.
[c] G-H denotes estimation using Gauss-Hermite quadrature.

Table 6.11: Survival model, with Weibull baseline hazard and various priors

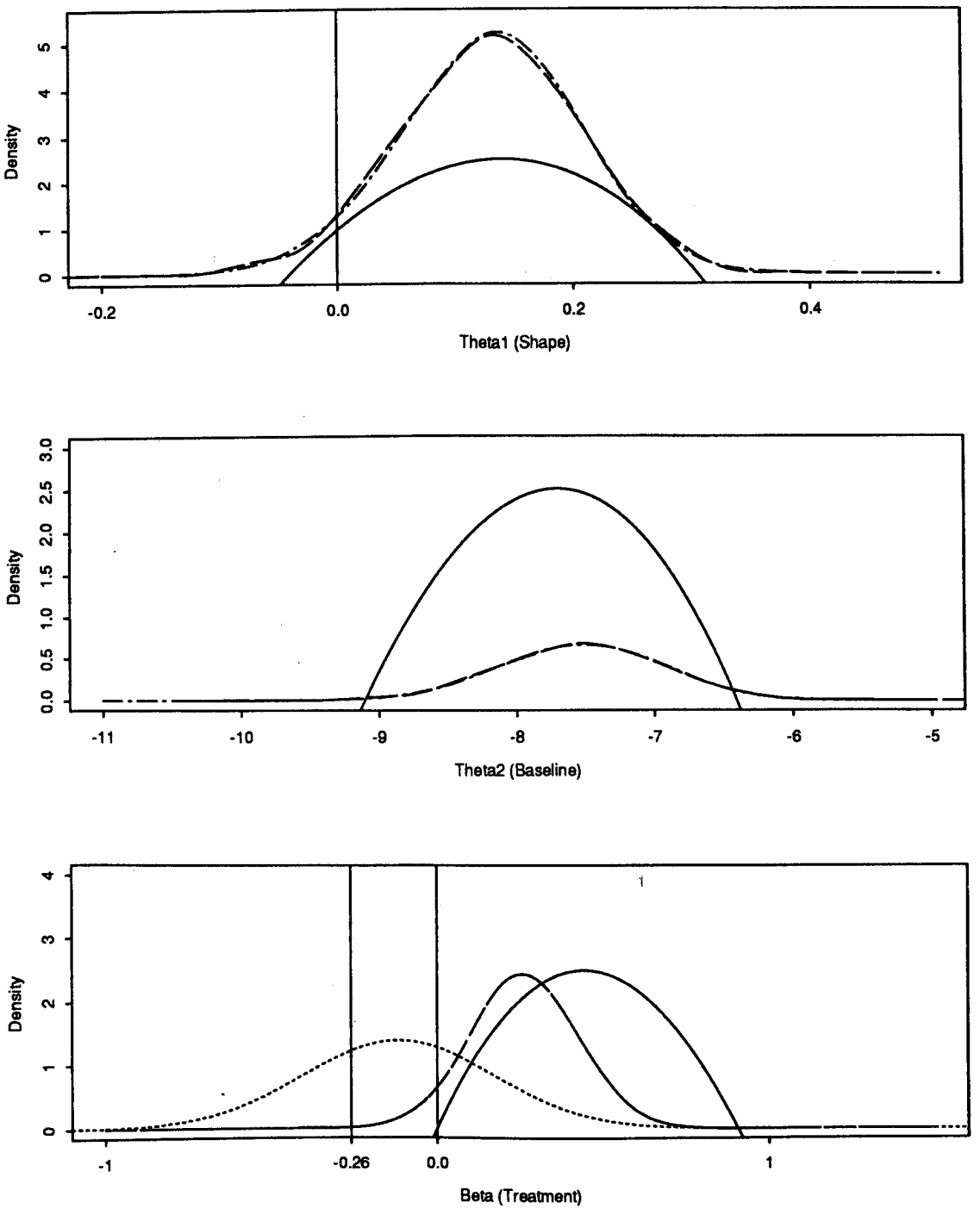
	$Cov(\theta_1, \theta_2)$	$Cov(\theta_1, \beta)$	$Cov(\theta_2, \beta)$
MLE	-0.045	0.001	-0.036
Prior	-	-	-
Posterior (L)[3]	-0.045	0.001	-0.034
Posterior (G-H)[4]	-0.044	0.001	-0.036
Prior	-	-	-
Posterior (L)	-0.044	0.001	-0.022
Posterior (G-H)	-0.044	0.001	-0.022
Prior	-	-	-
Posterior (L)	-0.048	0.000	-0.013
Posterior (G-H)	-0.048	0.000	-0.013

Table 6.12: Covariances for survival model using neutron data assuming Weibull baseline intensity and with various priors



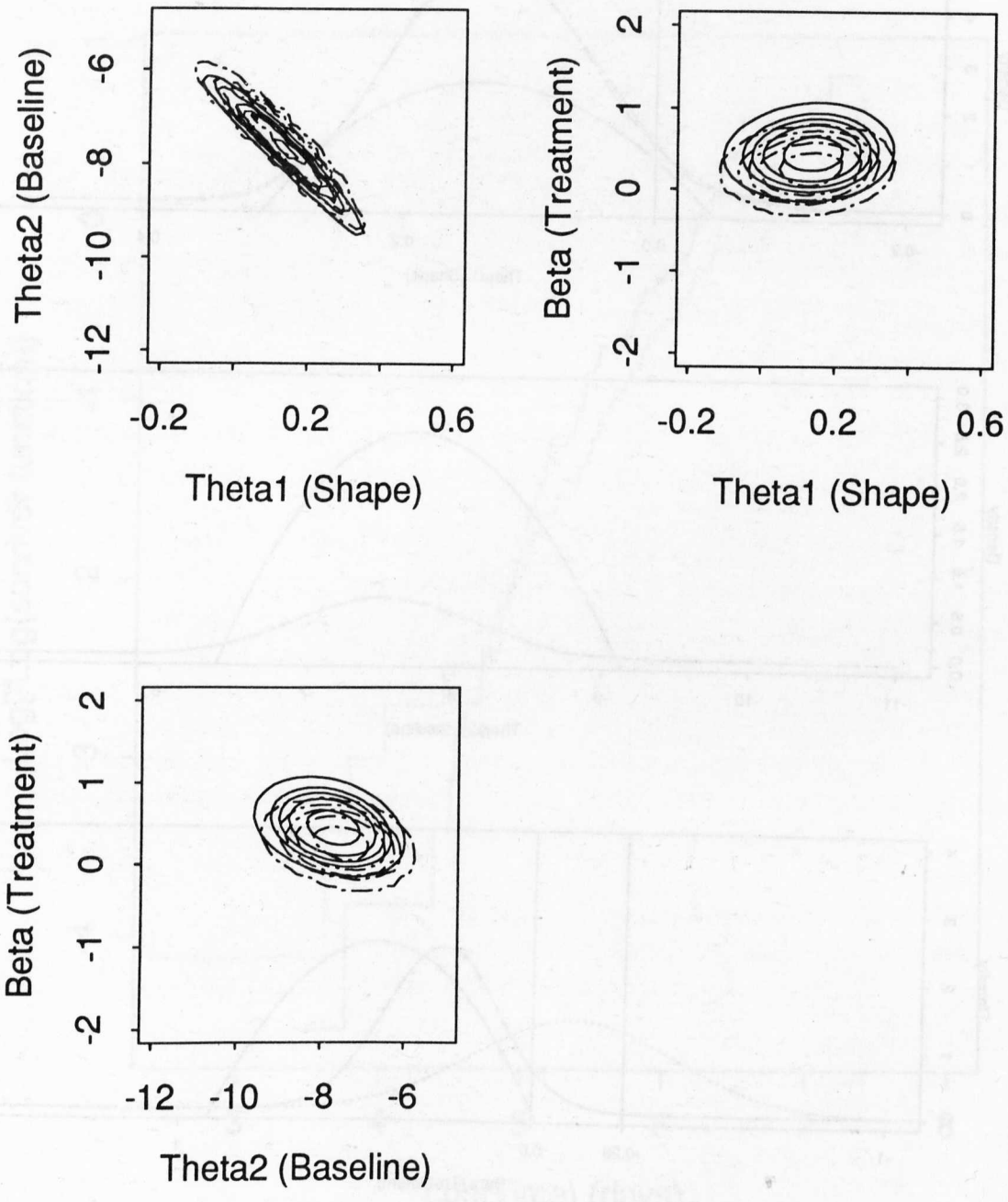
Key : — neutrons, ····· photons

Figure 6.14: Log[-Log(survivor function) for photons and neutrons.] based on Kaplan-Meier estimates of the survivor function for neutrons and photons for neutron therapy trial.



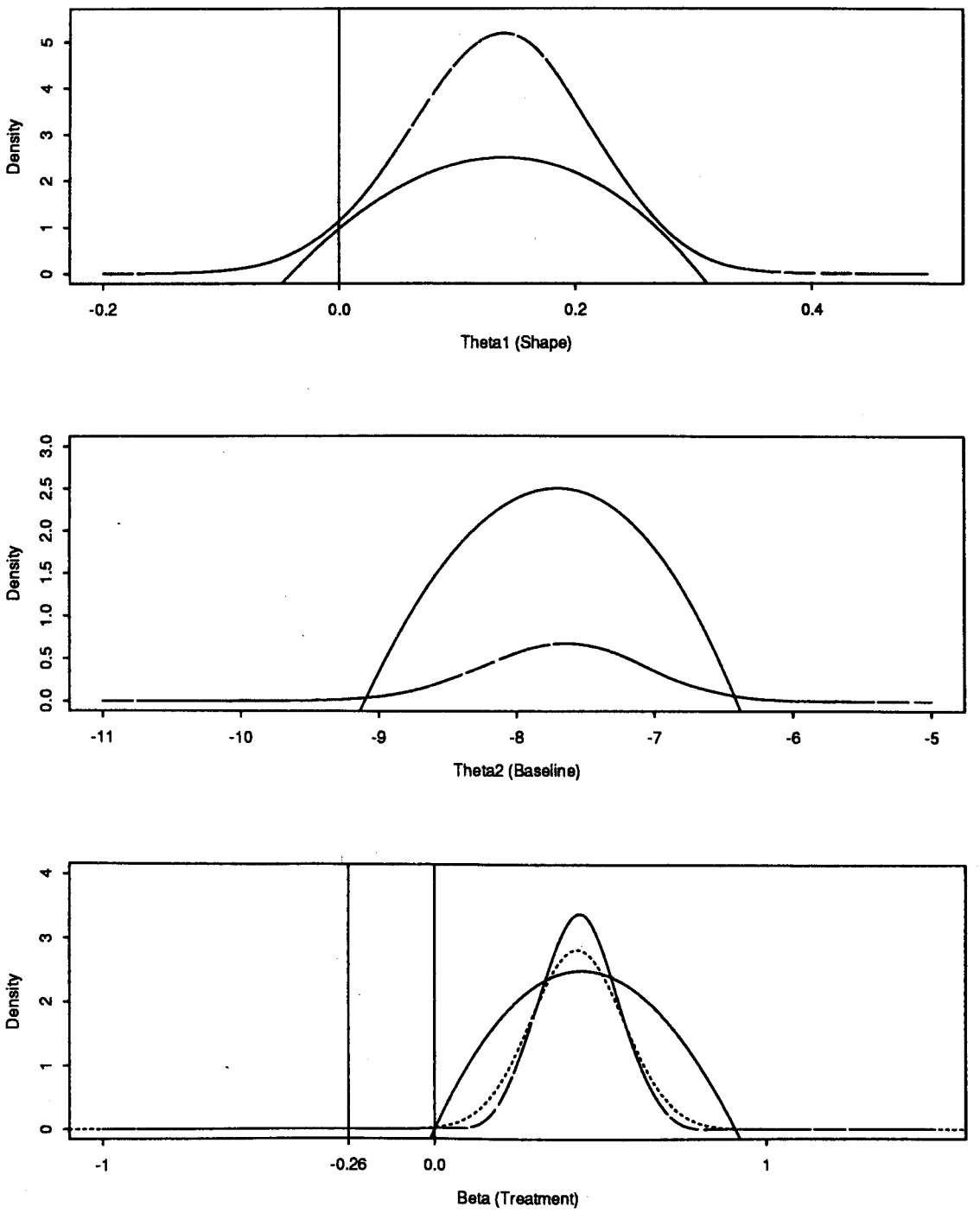
Key : — maximum relative log-likelihood, ····· clinical prior, – · – · – · – posterior (Laplace), – – – posterior (Gauss-Hermite)

Figure 6.15: Maximum relative log-likelihood and marginal posterior densities for survival model assuming Weibull baseline intensity with vague priors on θ_1 and θ_2 and clinical prior on β.



Key : —— maximum relative log-likelihood, ····· clinical prior, – · – · – · – posterior (Laplace), – – – posterior (Gauss-Hermite)

Figure 6.16: Joint maximum relative log-likelihood and joint posterior densities for survival model assuming Weibull baseline intensity with vague priors on θ_1 nd θ_2 and clinical prior on β, (50%, 80%, 90%, 95% and 99% credibility regions).



Key : —— maximum relative log-likelihood, ····· meta prior, – · – · – · – posterior (Laplace), – – – posterior (Gauss-Hermite)

Figure 6.17: Maximum relative log-likelihood and marginal posterior densities for survival model assuming Weibull baseline intensity with vague priors on θ_1 and θ_2 and meta prior on β.

6.8 Discussion

In this chapter we have considered making inferences about both the baseline intensity and the regression parameters in a fully parametric multiplicative intensity model when there is only one type of non-recurrent event. In particular we have considered the case when there may be prior information about one or more of the model parameters.

In the simplest case, when the baseline intensity is constant over time, there are at most two groups it is possible to obtain exact forms for the marginal posterior densities for the parameters. For more complicated situations when there may be several patient groups, defined by a number of covariates, a regression model was developed. Parameter estimation in this model required either asymptotic approximations such as those suggested by Tierney and Kadane (1986) or numerical integration techniques such as those advocated by Naylor and Smith (1982).

As an extension to the often unrealistic assumption that the baseline intensity is constant over time we extended the regression model so as to allow it to vary over a finite number of intervals, but to remain constant within each one. Though such a model necessarily was more complicated and had more parameters, by careful choice of the time intervals and use of more efficient optimisation routines parameter estimation was kept reasonably straightforward.

Finally, a regression model was considered in which the baseline intensity had a Weibull parametric form. This meant that the baseline intensity was a power transform of the time scale, the power parameters being considered as baseline parameters. Again asymptotic or numerical integration techniques were required for the estimation of model parameters.

All the three parametric models (constant intensity, piecewise constant intensity and Weibull intensity) were applied to the neutron therapy data described in Section 2.2 and the prior information and densities described in Sections 3.2 and 3.3 were used as prior inputs into the models. The results of these parametric models were compared to those of the much simpler non-temporal models described in Chapter 4 in which no account was explicitly taken of the time between the events, i.e treatment to death.

The results of all the models were in broad agreement. The common outcome is that the elicited prior beliefs of the clinicians used in the neutron therapy example should be updated in the light of the trial's results so that *a posteriori* they believe that neutrons are unlikely to be beneficial for patients with tumours of the pelvic region compared to photons. Beliefs based on the results of the meta analysis performed for the neutron study should change little in the light of the current study results. Overall a consensus was forming; that there was little belief in neutrons being more beneficial than photons for patients with cancer of the pelvic region. Similarly there was a consensus of belief forming that there was also negligible chance of a clinically significant difference in favour of neutrons.

There is considerable evidence to suggest that the baseline intensity is not

constant over time. The piecewise constant model with three time intervals; less than 6 months, 6 months to 12 months and greater than 12 months, indicates that the baseline intensity is greater in the second 6 months of follow-up than in either of the other two intervals. The Weibull intensity model also indicates evidence for a non-constant baseline intensity.

The Laplace approximations and Gauss-Hermite quadrature gave almost identical results for all three parametric models. Neither method proved difficult to implement for the models used in this chapter. However, Gauss-Hermite quadrature does require specialist software (Naylor and Shaw, 1985) whilst Laplace approximations only require the optimisation of the log-posterior density which can be implemented in a number of programming or statistical environments.

One of the main criticisms of the parametric models that we have described concerns the simplistic nature of the baseline intensity function. The baseline intensity could instead be modelled using a continuous smooth curve, such as a spline. This extension will be discussed further in Chapter 8.

Another criticism of the models considered in this chapter concerns the structure of the priors used. Rather than specifying, or eliciting prior information, about the model parameters directly, a hierarchical prior could have been used. In practice elicitation of prior information is often difficult, and a well elicited prior density for the main parameter of interest, a treatment effect say, is perhaps more useful than elicited estimates of hyper-parameters.

In summary, we have shown that Bayesian parametric multiplicative intensity models provide a coherent framework for the analysis of survival data. In the next chapter we will see how these simple two-state parametric models may be extended to the multi-state situation in which there is more than one type of event, and in which some of these events may be recurrent.

Chapter 7

Multi-State Models

7.1 Introduction

In this chapter we extend the two-state models of Chapter 6 to the case in which a patient may be in any one of a finite number of states at any point on the relevant time scale, and in which transitions from state to state may be reversible. A key assumption that we will make is that the transition times are known exactly, i.e transitions are assumed to take place at follow-up visits.

Several authors have considered the application of such models to medical settings, most notably Kay (1982, 1986) and Andersen (1986, 1988). More recently Longini *et al* (1989) and Van Druten *et al* (1990), have considered the use of such models in modelling the progression AIDS where a number of health states are possible. Hougaard and Madsen (1985) have also considered the application of multi-state models in the prognosis after myocardial infarction. All of this work has been from a classical perspective. Kirby (1991) though has considered the screening of women for cervical smear tests from a Bayesian perspective.

In the social sciences multi-state models have been used to study such phenomena as marital status, labour market dynamics and social unrest. Tuma, Hannan and Groeneveld (1979) describe the use of multi-state models to describe such social situations. De Stavola (1986) considered multi-state models for unemployment studies, in which individuals moved between employment states. As with medical applications of such model this work has been from a classical viewpoint. Lee, Judge and Zellner (1976) have considered the application of Bayesian multi-state models to macro and micro economic data.

The outline of this chapter is as follows; in Section 7.2 we consider the background to multi-state models, and their underlying assumptions. In Section 7.3 we extend the Bayesian two-state model of Chapter 6 for use with multi-state data. Section 7.4 describes the use of a Bayesian multi-state model in analysing the neutron therapy study, when we are interested in the modelling the development of metastases. Section 7.5 describes the application of multi-state models to the analysis of the quality of life data described in Chapter 2. Finally, we summarise and discuss the use of multi-state models in the two examples from cancer research in Section 7.6.

7.2 Classical Multi-State Models

Many authors have considered the use of time-homogeneous Markov processes in medical science, see Chiang (1968) for an overview. Though the approximations required for such models may seem strict, notably that the future is only dependent on the present and **not** on the past, and that the transition intensities between states are constant over time, estimation of the quantities of interest is relatively straight forward using the Chapman-Kolmogorov equations, see Cox and Miller (1965) and Chiang (1968) for details.

The assumption in such models that the intensities are constant over time is

equivalent to assuming that times between events are exponentially distributed, and the estimates obtained via maximum likelihood are equivalent to those obtained for the time-homogeneous Markov model. In the case of the intensities the maximum likelihood estimates are the number of events divided by the number of potential events. This highlights the link generally between stochastic processes and parametric survival models, i.e we can *count* the number of events in a specified time interval or we can consider the times between successive events.

The assumption of constant intensities may be relaxed by allowing them to be time varying, leading to a time-inhomogeneous Markov process. In the simplest case the time axis may be considered to be made up of a finite number of intervals and the intensity in each interval assumed to be constant.

Andersen (1986) amongst others, has noted that the assumption that the intensity of future events only depend on the state occupied at the current time is often an unrealistic one, and that we may also wish to allow for dependence on the time spent in the current state, i.e the sojourn time. For example in the case of the neutron therapy study considering the transition intensity from metastatic disease to death, a time-homogeneous or time-inhomogeneous Markov model assumes that the length of time a patient has had metastatic disease does not affect this intensity. We therefore require models that also allow for the inclusion of sojourn times. Such semi-Markov models may either use the sojourn times as covariates, or they may realign the time scale to be the onset of a particular event, e.g in the neutron therapy study this may be the onset of metastatic disease. As in the case of Markov models we may assume that the transition intensities are either time-homogeneous or time-inhomogeneous. When they are time-homogeneous the semi-Markov and Markov models coincide.

Figure 7.1 shows the implications of making Markov and semi-Markov assumptions. Consider a patient who say develops metastases at time t_1 and subsequently dies at time t_2, where time is measured from entry into the study, i.e $t = 0$. If we assume that the intensity is constant over time, say α, (Figure 7.1 (a)), then whether we assume a Markov or semi-Markov model the result is the same, i.e at time t^*, $t_1 < t^* < t_2$, the probability of dying in a short time interval $t^* + \delta t$ is α. Consider the case when we assume either a piecewise constant intensity with time grid $(0, \tau_1, \tau_2)$ (Figure 7.1 (b)) or a Weibull parametric form of the intensity (Figure 7.1 (c)). Under a Markov assumption the probability of dying in an interval $t^* + \delta t$ is $\alpha(t^*)$, which in the case of a piecewise constant intensity is α_3. Under a semi-Markov time-homogeneous assumption the probability of dying in the interval $t^* + \delta t$ is $\alpha(t^* - t_1)$.

The above rationale for modelling has been developed from a stochastic process point of view. A number of authors have considered the case when there are multivariate failures, see Kalbfleisch and Prentice (1980), Cox and Oakes (1984) and Clayton and Cuzick (1985). These authors have considered the case when the survival model is of a multiplicative intensity form, but with the baseline intensity left unspecified, thus yielding a multivariate version of a Cox proportional hazards model. Clayton (1991) has considered 'frailty models' in which

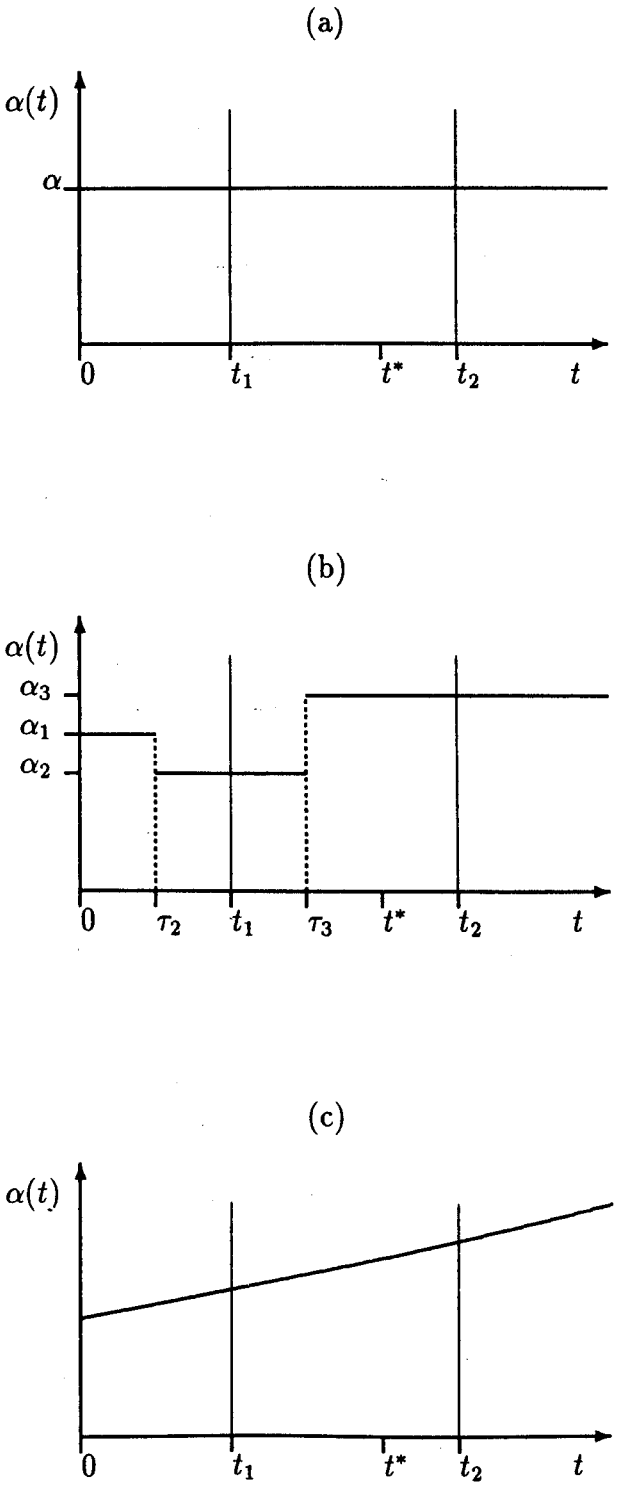


Figure 7.1: Models assuming (a) constant, (b) piecewise constant and (c) Weibull intensities.

the times to an event of interest for patients are not necessarily independent. The degree of dependence is governed by the frailty parameter.

Crowder, Kimber, Smith and Sweeting (1991) have reviewed work that has been done when the underlying intensity is not left unspecified. Beck (1979) has considered the case when the transition intensities are considered either to be constant or to have a Weibull parametric form. Lagakos (1976) also considered the special case when the transition intensities are assumed to be constant.

In we assume that the times at which transitions in a multi-state model take place are observed then we can consider the multiplicative intensity model described in Section 5.3, and its extension to the case when h types of transitions are possible. The intensity function becomes

$$\alpha_{ik}(t) = \lambda_{0k}(t)\, e^{\underline{\beta}_k^T \underline{z}_{ik}}\, Y_{ik}(t) \quad i = 1, \cdots, n \quad k = 1, \cdots, h \tag{7.1}$$

where $\lambda_{0k}(t)$ is the baseline intensity function for the kth transition, $\underline{z}_{ik}$ are the covariates in the model for the ith patient and the kth transition, and $Y_{ik}(t)$ is the 'at-risk' process for the ith patient and the kth transition. Andersen and Borgan (1985) and Fleming and Harrington (1991) have shown for both the case when $\lambda_{0k}(t)$ is left unspecified or when a particular parametric form is assumed, that the likelihood factorises into terms for each individual type of transition. This has also briefly been described in Section 5.3.3 Thus, we can consider each transition separately.

The overall likelihood corresponding to (6.2) is of the form

$$L(\underline{\phi}_1, \cdots, \underline{\phi}_h) = \prod_{i=1}^{n} \prod_{k=1}^{h} \left[\alpha_{ik}(t)^{dN_{ik}(t)}\, e^{-\int_0^t \alpha_{ik}(u)\, du} \right] \tag{7.2}$$

where $\underline{\phi}_k$ represents the model parameters for the kth type of event. This likelihood is a product over both patients *and* events.

The log-likelihood is of the form

$$\ell(\underline{\phi}_1, \cdots, \underline{\phi}_h) = \sum_{i=1}^{n} \sum_{k=1}^{h} \left\{ \log_e[\alpha_{ik}(t)]\, dN_{ik}(t) - \int_0^t \alpha_{ik}(u)\, du \right\} \tag{7.3}$$

where $\underline{\phi}_k$ represents the model parameters for the kth type of transition.

7.3 Bayesian Multi-State Models

In this section we develop a Bayesian multi-state model based on the fully parametric multiplicative intensity models of Chapter 6. We also consider other Bayesian approaches to multi-state models.

7.3.1 A Multiplicative Intensity Model

In this section we consider the case when $\lambda_{0k}(t)$ in (7.1) is assumed to have a particular parametric form, and that there is possibly prior information in the

form of a density function for a treatment parameter. The model takes the form

$$\alpha_{ik}(t) = \lambda_0(t|\underline{\theta}_k)\, e^{\underline{\beta}_k^T \underline{z}_{ik}}\, Y_{ik}(t) \quad i = 1, \cdots, n \quad k = 1, \cdots, h \tag{7.4}$$

where $\underline{\theta}_k$ are the baseline parameters for the kth transition. As in Chapter 6 we consider three possible forms for $\lambda_0(t|\underline{\theta}_k)$. These correspond to constant baseline intensity, piecewise constant or Weibull form baseline intensity. We will adopt the same parameterisation as in Chapter 6. This means that $\lambda_0(t|\underline{\theta}_k)$ is of the form e^{θ_k}, $e^{\theta_{jk}}$ or $e^{\theta_{1k}}\, t^{e^{\theta_{1k}}-1}\, e^{\theta_{2k}}$ depending on whether the baseline intensity is constant, piecewise constant or has a Weibull form respectively. This parameterisation assumes that only the values of the baseline parameters change with each different transition. It may be that the actual form of the baseline intensity needs to change. In this case $\lambda_0(t|\underline{\theta}_k)$ needs to be replaced by $\lambda_{0k}(t|\underline{\theta}_k)$. Since the likelihood has been shown to factorise with respect to the different transitions (Section 5.3.3), such an extension of (7.1) is relatively straightforward. We only consider the case when the parameters change, but consider assessing the fit of the different baseline intensities in the examples below.

In the case of there being little prior information about either the baseline parameters of any possible regression parameters, the joint posterior, $p(\underline{\phi}_1, \cdots, \underline{\phi}_h|H)$, is proportional to the overall likelihood (7.2).

Considering the kth transition, and that there is prior information in the form of a density about the pth regression parameter, $p(\beta_{pk})$, then as the likelihood factorises the joint posterior for the kth transition is of the form

$$p(\underline{\phi}_k|H) \propto \prod_{i=1}^{n}\Big[(\lambda_0(t|\underline{\theta}_k)\, e^{\underline{\beta}_k^T \underline{z}_{ik}}\, Y_{ik}(t))^{dN_{ik}(t)} e^{-\int_0^t \lambda_0(u|\underline{\theta}_k)\, e^{\underline{\beta}_k^T \underline{z}_{ik}}\, Y_{ik}(u)\, du}\Big]\, p(\beta_{pk}) \tag{7.5}$$

As in the case of the two state model if we assume that $p(\beta_{pk})$ is a Normal density, with mean μ_k and variance σ_k^2 then (7.5) becomes

$$p(\underline{\phi}_k|H) \propto \prod_{i=1}^{n}\Big[(\lambda_0(t|\underline{\theta}_k)\, e^{\underline{\beta}_k^T \underline{z}_{ik}}\, Y_{ik}(t))^{dN_{ik}(t)} e^{-\int_0^t \lambda_0(u|\underline{\theta}_k)\, e^{\underline{\beta}_k^T \underline{z}_{ik}}\, Y_{ik}(u)\, du}\Big]\, e^{-\frac{1}{2\sigma_k}(\beta_{pk}-\mu_k)^2} \tag{7.6}$$

We can see from the form of (7.6) that no closed form solution exists for the posterior marginal densities, in particular our parameter of interest, β_{pk}. As in the two state case we need to resort to some form of approximation methodology. Both the Laplace approximations suggested by Tierney and Kadane (1986) and Gauss-Hermite quadrature advocated by Naylor and Smith (1982) described in detail in Chapter 6 and successfully applied to the two state models, could be used in order to obtain posterior marginal densities and posterior moments.

7.3.2 Other Approaches

With the exception of Kirby (1991) little work has been published on Bayesian inference in multi-state models for medical applications. Lee, Judge and Zellner (1976) have considered Bayesian inference in multi-state models applied to economic data. The main aim of the work was to derive links between macro and micro economic models, and how the former could help to predict the performance in the latter. The case of micro economic data corresponds to the case when information about individual patients is available. Lee, Judge and Zellner's work is a generalisation of the 'Beta-Binomial' model considered in Chapter 4, and has also been described by Lindley (1965). Assuming a time-homogeneous Markov process, a multinomial likelihood is formed for the transition probabilities, and a multivariate Beta prior density is assumed for the probabilities. Analogous to the univariate case the corresponding posterior density is that of a multivariate Beta distribution. Inference may then be made about individual transitions.

Kirby (1991) considered modelling the screening histories of women in a cervical screening programme in the Grampian region of Scotland. She considered the data as a series of imperfect observations from a discrete state Markov process in continuous time, for which the extact transition times are not known. Graphical models were used to describe the inter-relationships between all of the relevant variables, both observed and unobserved. Estimation of the parameters of these graphical models used Gibbs sampling, briefly described in Section 6.4.3.

Aven (1986) has considered Bayesian inference for a parametric counting process. In particular he considered the case when the intensity associated with each component of a multivariate counting process was constant. In this situation assuming a conjugate Gamma prior density for the intensity yields a posterior Gamma density for the intensity. Though such a model can be used when there are a number of events and individuals, difficulty arises when a regression structure is assumed.

The Bayesian survival model advocated by Gamerman and West (1987a, 1987b), described in Chapter 6 relies on the factorisation of the likelihood at distinct failure times, and extension to the multi-state scenario is feasible.

Hjort (1986) has considered both Bayesian parametric and semi-parametric model for counting process models. Kalbfleisch (1978) has considered the case when the baseline hazard is assumed to be non-parametric and assumed a Gamma process prior for it, that setting this prior to be vague, yields 'partial likelihood'. If the model assumptions are valid we have a Bayesian semi-parametric model. We have seen from Section 5.3.2 that the (partial) likelihood for such a model factorises when there are a number of different types of event possible. Therefore, Kalbfleisch's method could be applied to a multi-state semi-parametric model.

7.4 Neutron Therapy Example

In this section a 3-state model shown in Figure 7.2 is used to study the effect of treatment on the risk of developing metastases, and for death.

7.4.1 Description of the data

Table 7.1 shows the number of transitions for each treatment group, and Figure 7.5 shows this diagrammatically. Figure 7.5 shows the dynamic evolution of metastatic disease with different types of line representing the different states. Of the 154 patients in the trial at this stage only 143 are in this subset, as only those patients who had one or more follow-up visits were considered at risk of any of the transitions. One patient was excluded because data was missing. It should be noted that metastases refers to clinically diagnosed metastases. Some patients were found to have developed metastases only at autopsy. However, there was a bias in autopsies towards neutron patients, and therefore metastases at death were not included.

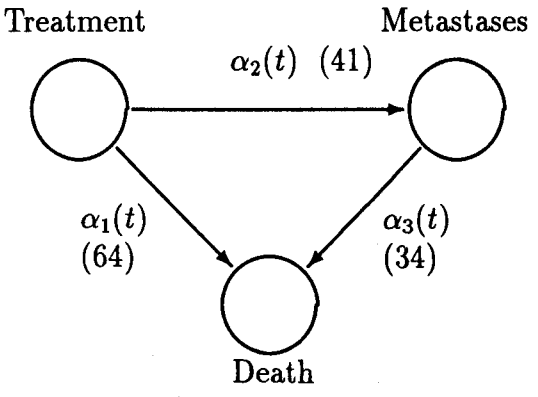


Figure 7.2: 3-state model for development of metastatic disease, $(\cdot)$ denotes the number of patients that made a particular transition, and $\alpha_k(t)$ is the transition intensity for the kth transition at time t.

We can represent this aspect of the neutron therapy data using counting process notation. Corresponding to the transition intensity $\alpha_1(t)$ in Figure 7.2 there is a counting process $N_{i1}(t)$ which is a 0/1 indicator function, zero until

a patient dies, and one after a patient has died, without having metastases, and similarly $Y_{i1}(t)$ is an at risk process that is one while a patient is at risk of dying (without metastases), and zero when they are not at risk. A patient may no longer be at risk of dying (without metastases) at a particular time because either they have already died or they have developed metastases. A similar situation is true for $\alpha_2(t)$, in that $N_{i2}(t)$ and $Y_{i2}(t)$ are both 0/1 indicator functions. $Y_{i2}(t)$ is one while patient i is at risk of developing metastases and zero when they are no longer at risk, either because they have died or they have already developed metastases. Similarly, $N_{i2}(t)$ is zero until patient i develops metastases, after which time it is one. The counting process and at-risk process corresponding to $\alpha_3(t)$ are again 0/1 indicator functions, but unlike the other processes they have to accommodate the case when there is delayed entry. At time zero $Y_{i3}(t)$ is zero, but if patient i develops metastases at some time after entry then $Y_{i3}(t)$ becomes one, and if patient i eventually dies $Y_{i3}(t)$ becomes zero again. The processes corresponding to $Y_{i3}(t)$ and $N_{i3}(t)$ for patient i who develops metastases at time t_1 and then dies at time t_2 are shown in Figures 7.3 and 7.4 respectively.

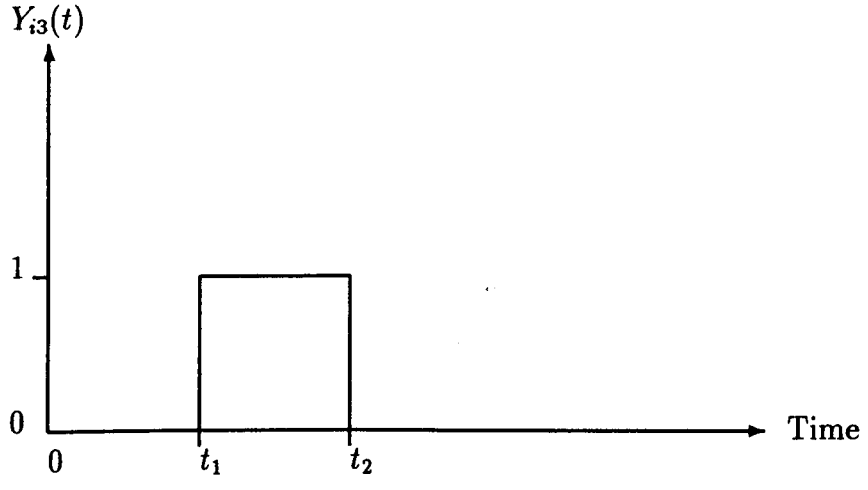


Figure 7.3: A simple risk process, $Y_{i3}(t)$, for metastases to death transition.

Transition	Neutrons	Photons	Total
Treatment to Death (without Mets)	41	23	64
Treatment to Metastases	30	11	41
Metastases to Death	26	8	34

Table 7.1: Transitional status of neutron and photon patients at 21st December 1990.

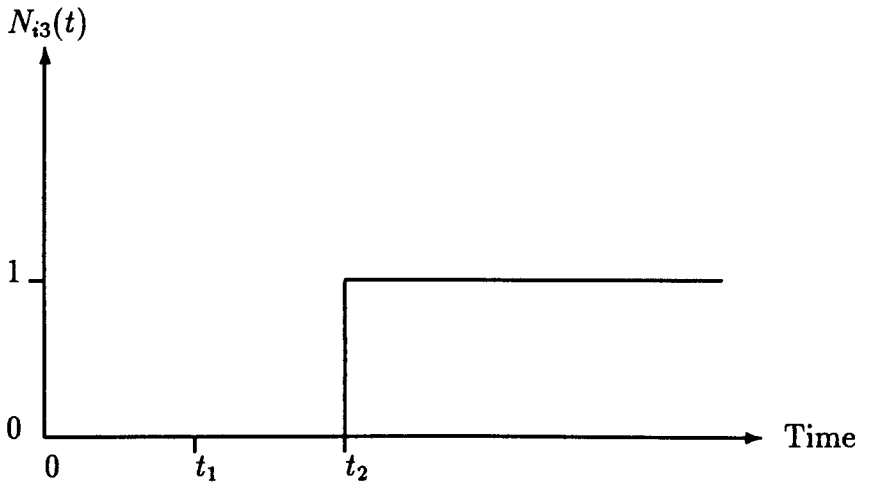


Figure 7.4: A simple counting process, $N_{i3}(t)$, for metastases to death transition.

7.4.2 Methods

In all the models that we shall consider Uniform densities were used for the baseline intensity parameters, $\underline{\theta}_1, \underline{\theta}_2, \underline{\theta}_3$, for all transitions. For the treatment effect, β, the prior distributions based on clinical beliefs described in Chapter 3 were used for transitions 1 and 2, i.e treatment to death and treatment to metastases respectively. The rationale for this was that in eliciting clinical beliefs the clinicians were asked for their beliefs about treatment failure, which included the development of metastases. For transition 3, from metastases to death, a Uniform density was used for all model parameters, as there was no prior information of relative treatment differences once a 'failure' had occurred.

For transitions 1 and 2, treatment to death and treatment to metastases respectively, we considered time-homogeneous and time-inhomogeneous Markov models. These were equivalent to semi-Markov processes since entry into the treatment state is by definition time of entry into the study. Baseline intensities were assumed to be constant, piecewise constant or to have a Weibull parametric form. For transition 3, metastases to death, we again fitted Markov models, both time-homogeneous and time-inhomogeneous, but we also considered semi-Markov models since the sojourn times, i.e the time that a patient spent in the metastatic state before making the transition to death, were also available. Therefore we considered time-homogeneous, but sojourn-inhomogeneous semi-Markov models, which in the case of a constant baseline intensity are equivalent to time-homogeneous Markov models.

As mentioned in Section 7.3 estimation of the posterior densities, means and variances, can be performed using either the Laplace approximations advocated by Tierney and Kadane (1986) or Gauss-Hermite quadrature described by Naylor and Smith (1982). Both methods were described in detail in Chapter 6.

Maximum likelihood estimates were also obtained directly by maximising the likelihood using `nlmin` in Splus (1990), and asymptotic estimates of the standard deviations were obtained using minus the inverse of the hessian matrix evaluated at the maximum likelihood value.

7.4.3 Results

Table 7.1 shows the transitional status of patients on the 21st December 1990. We can see that initial inspection of this table would lead us to believe that not only are neutron patients more likely to die, but they are more likely to have metastatic spread, and subsequently to die. However, this table makes allowance for neither the times taken for these events to occur or for the effect of censoring.

Table 7.2 shows parameter estimates in the case when the baseline intensity function is constant and clinical priors were used for the treatment effect, β, for transitions 1 and 2, and Uniform prior densities were used for all other parameters. Table 7.3 shows the parameter estimates assuming a piecewise constant baseline intensity function, with two intervals, less than or equal to 365 days, and greater than 365 days. Finally Table 7.4 gives the estimates assuming a Weibull parametric form for the baseline intensity. Figures 7.7 to 7.9 show the maximum relative log-likelihood, prior density and marginal posterior densities for β the treatment effect parameter. β is the logarithm of the ratio of the intensities for the two groups, so that $\beta = 0$ indicates a relative intensity ratio of one, i.e no treatment difference. Whilst a value of β less than zero indicates that neutrons have less risk than photons of a particular transition, correspondingly a value of β greater than zero indicates that photons have less risk than neutrons of the transition in question.

Transition	θ (Baseline)		β (Treatment)			
	Mean	SD	Mean	SD	$P_{(0)}$[a]	$P_{(-0.26)}$
Survival						
Prior	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)[b]	-6.753	0.147	0.313	0.169	0.032	0.000
Posterior (G-H)[c]	-6.753	0.147	0.313	0.169	0.032	0.000
MLE	-6.898	0.178	0.546	0.216	-	-
Treatment to Death ($k = 1$)						
Prior	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)	-6.904	0.167	0.192	0.190	0.080	0.009
Posterior (G-H)	-6.904	0.168	0.192	0.190	0.080	0.009
MLE	-7.041	0.209	0.437	0.261	-	-
Treatment to Metastases ($k = 2$)						
Prior	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)	-7.419	0.204	0.301	0.213	0.053	0.004
Posterior (G-H)	-7.419	0.204	0.301	0.214	0.053	0.004
MLE	-7.779	0.301	0.862	0.352	-	-
Metastases to Death ($k = 3$)						
Prior	-	∞	-	∞	-	-
Posterior (L)	-5.305	0.355	0.084	0.402	0.340	0.196
Posterior (G-H)	-5.306	0.365	0.085	0.416	0.341	0.203
MLE	-5.242	0.353	0.041	0.404	-	-

[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that β is less than 0 and -0.26 respectively.
[b] L denotes Laplace approximation.
[c] G-H denotes Gauss-Hermite quadrature.

Table 7.2: Parameter estimates for models with constant baseline intensities using various prior densities for β and Uniform prior densities for θ.

Survival

We can see from Tables 7.2 to 7.4 that for survival all three parametric models, that is constant, piecewise constant and Weibull form intensities, yield approximately the same results. These are such that the group of clinicians who prior to the trial being conducted strongly believed in the efficacy of neutrons, should revise their beliefs so that neutrons are unlikely to be beneficial compared to photons. All three models yield a posterior probability of neutrons being beneficial of 2-3%. These results are in broad agreement with those presented in Chapter 6 for the full data set, i.e $n = 154$.

Treatment to Death

For transition 1, treatment to death without developing metastases, there is considerable evidence to suggest that neutron patients do worse than photon

patients. This is in spite of the fact that an informative prior in favour of neutron therapy was used. All three models yielded a posterior relative risk (neutrons to photons) of 1.2 compared to the clinicians' prior relative risk of 0.89. The clinicians' have a posterior probability that neutrons were beneficial of 9-11% compared to a prior probability of 66%.

Treatment to Metastases

Considering transition 2, treatment to metastases, again there is broad agreement between the three parametric models. In the light of the trials results the clinicians should revise their beliefs so that *a posteriori* they believe that the relative risk (neutrons to photons) is approximately 1.3, and that the posterior probability of neutrons being beneficial, compared to photons, is 5-8%.

Metastases to Death

Using a time-homogeneous but sojourn-inhomogeneous Markov model for transition 3, metastases to death, we find there is again broad agreement between the models. The clinicians' posterior belief about the relative risk indicate an approximate estimate of 1.1, 95% credibility interval (0.49,2.50), using Laplace approximations. This estimate ignores the time at which patients developed metastases and only considers the length of time from onset of metastases to death. The clinicians' posterior probability that neutrons were better than photons in terms of death after metastases is 40%. So therefore once patients have developed metastases there is a slightly increased risk of dying if a patient originally received neutron therapy rather than photon therapy. If instead we consider a time-inhomogeneous Markov model for transition 3 the clinicians' posterior belief about the relative risk of death is centred on 0.9, 95% credibility interval (0.38,2.00), with a posterior probability that neutrons are beneficial of 40-45%. These results indicate that there is little evidence of difference between the two treatments in terms of death following metastatic spread even when the time at which metastases developed is taken into account.

Comparison with classical semi-parametric models

The results, for the metastases to death transition, obtained using the fully parametric models can be compared with those obtained using a semi-parametric multi-state model with the baseline intensity left unspecified, as have been applied by Andersen (1988) and Kay (1982). In the case of a semi-Markov model a relative risk of 1.02 was obtained together with a 95% confidence interval of (0.46,2.26). A Markov model, fitted by using time-dependent strata in the 21 program in BMDP[1] , yielded a relative risk of 0.83 with an approximate 95% confidence interval of (0.37,2.03). Comparing these results with Tables 7.3 and 7.4 we can see that they correspond closely to the maximum

[1]BMDP is a trademark of BMDP Statistical Software Inc.

Transition	θ_1 ($t \leq \tau_2$)		θ_2 ($t > \tau_2$)		β (Treatment)			
	Mean	SD	Mean	SD	Mean	SD	$P_{(0)}$[a]	$P_{(-0.26)}$
Survival								
Prior	-	∞	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)[b]	-6.780	0.172	-6.727	0.190	0.316	0.169	0.031	0.000
Posterior (G-H)[c]	-6.780	0.171	-6.727	0.190	0.316	0.169	0.031	0.000
MLE	-6.932	0.202	-6.853	0.213	0.553	0.218	-	-
Treatment to Death ($k = 1$)								
Prior	-	∞	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)	-6.863	0.191	-7.001	0.246	0.186	0.190	0.164	0.009
Posterior (G-H)	-6.863	0.191	-7.001	0.245	0.186	0.190	0.164	0.009
MLE	-7.003	0.231	-7.103	0.269	0.427	0.262	-	-
Treatment to Metastases ($k = 2$)								
Prior	-	∞	-	∞	-0.116	0.286	0.657	0.307
Posterior (L)	-7.211	0.223	-8.002	0.383	0.274	0.214	0.100	0.006
Posterior (G-H)	-7.211	0.222	-8.003	0.376	0.274	0.214	0.100	0.006
MLE	-7.559	0.318	-8.252	0.424	0.805	0.354	-	-
Metastases to Death (Semi-Markov) ($k = 3$)								
Prior	-	∞	-	∞	-	∞	-	-
Posterior (L)	-5.334	0.381	-5.300	0.642	0.102	0.408	0.401	0.187
Posterior (G-H)	-5.335	0.372	-5.302	0.642	0.102	0.413	0.402	0.190
MLE	-5.271	0.368	-5.088	0.620	0.059	0.409	-	-
Metastases to Death (Markov) ($k = 3$)								
Prior	-	∞	-	∞	-	∞	-	-
Posterior (L)	-4.794	0.407	-5.661	0.406	-0.132	0.413	0.625	0.378
Posterior (G-H)	-4.795	0.407	-5.662	0.408	-0.131	0.420	0.622	0.379
MLE	-4.726	0.402	-5.569	0.399	-0.174	0.415	-	-

[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that β is less than 0 and -0.26 respectively.
[b] L denotes Laplace approximation.
[c] G-H denotes Gauss-Hermite quadrature.

Table 7.3: Parameter estimates for time-inhomogeneous Markov and semi-Markov models with piecewise constant intensities, using various prior densities for β and Uniform prior densities for $\underline{\theta}$ with $\underline{\tau} = (0, 365, \infty)$ days.

likelihood estimates (MLE) obtained using the piecewise constant and Weibull form intensity models.

Clinical Interpretation

General interpretation of these results is that *a posteriori* the group of clinicians who *a priori* believed that neutrons were beneficial for treating tumours of the pelvic region, now should believe them not to be as effective as photons. Neutron patients both die at, and develop metastases at, a faster rate than photon patients, but that those who do develop metastases show no substantial difference in terms of subsequent survival even when the time at which metastases developed was taken into account.

Model Testing

In the above analysis we have assumed a constant, piecewise constant or Weibull baseline intensity for each of the possible transitions. It may be the case that a particular parametric form of the baseline intensity is more appropriate for a specific transition. We can assess the strength of evidence for each of the parametric intensity functions by looking at the posterior estimates of the baseline intensity parameters. In the case of the piecewise constant intensity assessing whether $\theta_1 = \theta_2$ and for the Weibull baseline intensity whether θ_1 is close to zero, indicating that a constant intensity is more appropriate.

Tables 7.2 to 7.4 indicate that for survival there appears to be little evidence to support the use of a piecewise constant intensity model, with $\underline{\tau} = (0, 365, \infty)$ days, whilst there does appear to be support for the use of a model with a Weibull form intensity over one with a constant intensity. The same situation holds in the case of the treatment to death transition, with little evidence for the piecewise constant intensity model, but more for the Weibull intensity model over the constant intensity model.

However for treatment to metastases the case is reversed with the shape parameter, θ_1, in the Weibull intensity model being very close to zero, indicating little difference from the constant intensity model. The piecewise constant intensity model shows some reduction in the baseline intensity in the second epoch, i.e after one year, suggesting that the baseline intensity is not constant over time.

Considering the semi-Markov models for the metastases to death transition there appears to be little evidence to suggest that the intensity is anything other than constant, with the baseline intensities in the piecewise constant model approximately equal and the shape parameter in the Weibull model being near zero.

The Markov models suggests that for the metastases to death transition the baseline intensity is not constant over time. The intensities in the piecewise constant model are different, and the shape parameter in the Weibull model is not near zero. Both models indicate that the baseline intensity for this transition is reducing with time, measured from entry into the study.

Transition	θ_1 (Shape)		θ_2 (Baseline)		β (Treatment)			
	Mean	SD	Mean	SD	Mean	SD	$P_{(0)}$ [a]	$P_{(-0.26)}$
Survival								
Prior	-	∞	-	∞	-0.116	0.082	0.657	0.307
Posterior (L)[b]	0.229	0.080	-8.468	0.272	0.347	0.169	0.019	0.000
Posterior (G-H)[c]	0.229	0.080	-8.460	0.679	0.347	0.169	0.019	0.000
MLE	0.243	0.080	-8.705	0.697	0.602	0.218	-	-
Treatment to Death ($k = 1$)								
Prior	-	∞	-	∞	-0.116	0.082	0.657	0.307
Posterior (L)	0.260	0.096	-8.974	0.838	0.231	0.190	0.110	0.005
Posterior (G-H)	0.105	0.098	-7.650	0.720	0.208	0.190	0.137	0.007
MLE	0.278	0.096	-9.250	0.862	0.511	0.262	-	-
Treatment to Metastases ($k = 2$)								
Prior	-	∞	-	∞	-0.116	0.082	0.657	0.307
Posterior (L)	-0.046	0.114	-7.206	0.807	0.297	0.214	0.082	0.004
Posterior (G-H)	-0.043	0.128	-7.206	0.799	0.297	0.213	0.083	0.004
MLE	-0.010	0.126	-7.716	0.858	0.859	0.354	-	-
Metastases to Death (Semi-Markov) ($k = 3$)								
Prior	-	∞	-	∞	-	∞	-	-
Posterior (L)	-0.077	0.138	-4.956	0.797	0.084	0.402	0.428	0.201
Posterior (G-H)	-0.077	0.138	-4.956	0.784	0.084	0.407	0.418	0.199
MLE	-0.063	0.138	-4.905	0.788	0.037	0.404	-	-
Metastases to Death (Markov) ($k = 3$)								
Prior	-	∞	-	∞	-	∞	-	-
Posterior (L)	-0.736	0.326	-1.653	1.682	-0.020	0.412	0.612	0.324
Posterior (G-H)	-0.758	0.667	-1.928	2.202	-0.034	0.451	0.530	0.308
MLE	-0.501	0.397	-2.385	1.855	-0.067	0.412	-	-

[a] $P_{(0)}$ and $P_{(-0.26)}$ denote the probability that $\beta < 0$ and $\beta < -0.26$ respectively.
[b] L denotes Laplace approximation.
[c] G-H denotes Gauss-Hermite quadrature.

Table 7.4: Parameter estimates for time-inhomogeneous Markov and semi-Markov models with Weibull parametric intensities, using clinical prior density for β and Uniform prior densities for $\underline{\theta}$.

7.4.4 Discussion of Neutron Therapy Example

In this example we have been able to assess the role that metastatic disease spread played in subsequent death. Using these models we have been able to show how prior beliefs about treatment failure held by the clinicians in the study should be revised in the light of the trial results. We saw how even though *a priori* the clinicians believed that patients treated with neutrons would fare better in terms of treatment failure, *a posteriori* they should revise their beliefs so that patients treated with photons had a better prognosis both in terms of death without metastatic spread and the development of metastases itself. Though there was no prior information on events after metastatic spread *a posteriori* there was considerable evidence to suggest little difference between the two treatment groups in terms of subsequent risk of death regardless of either sojourn time with metastases or time of development in trial time. These findings concur with the simple classical analysis presented in Chapter 2, in which the development of metastases was treated as a time-dependent covariate. In that analysis although metastases on there own seemed to influence future survival there was little evidence to suggest a treatment-metastases interaction.

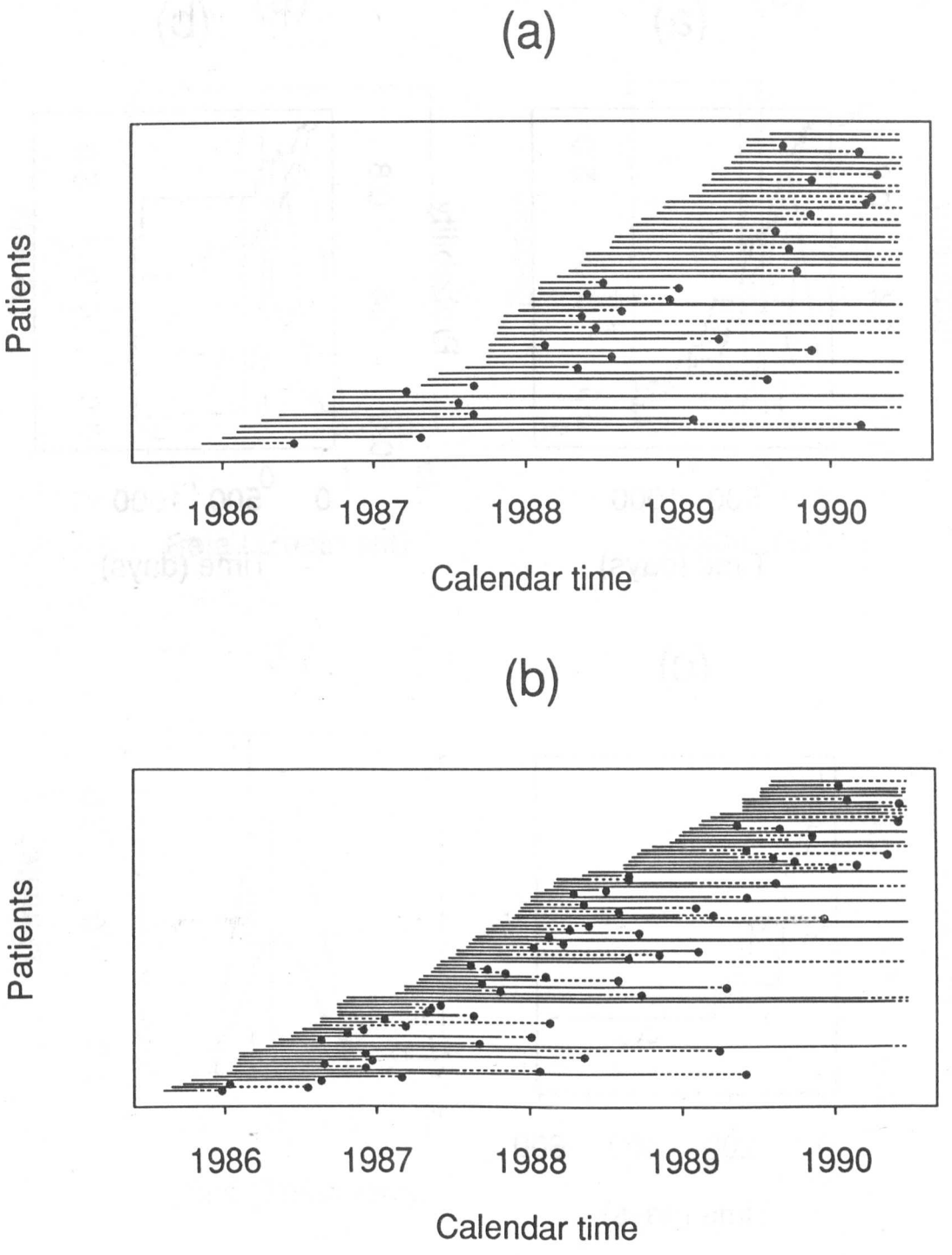
We can see from Tables 7.2 to 7.4 and Figures 7.7 to 7.9 that both the Laplace approximations suggested by Tierney and Kadane (1986) and the method of Gauss-Hermite quadrature advocated by Naylor and Smith (1982) yield approximately the same results.

As has already been stated although the neutron therapy trial was stopped because of the mortality results initial interest in the study surrounded morbidity. In this chapter we have only considered one aspect of morbidity, that is the development of metastases. There are two other aspects of morbidity left to investigate, treatment toxicity and tumour progression and regression. Unfortunately at the time data of sufficient quality is not available on either of these outcomes, but it is hoped that such analyses will be carried out in the future. Figure 7.10 shows the modelling of tumour progression and regression diagrammatically.

7.5 Quality of Life Example

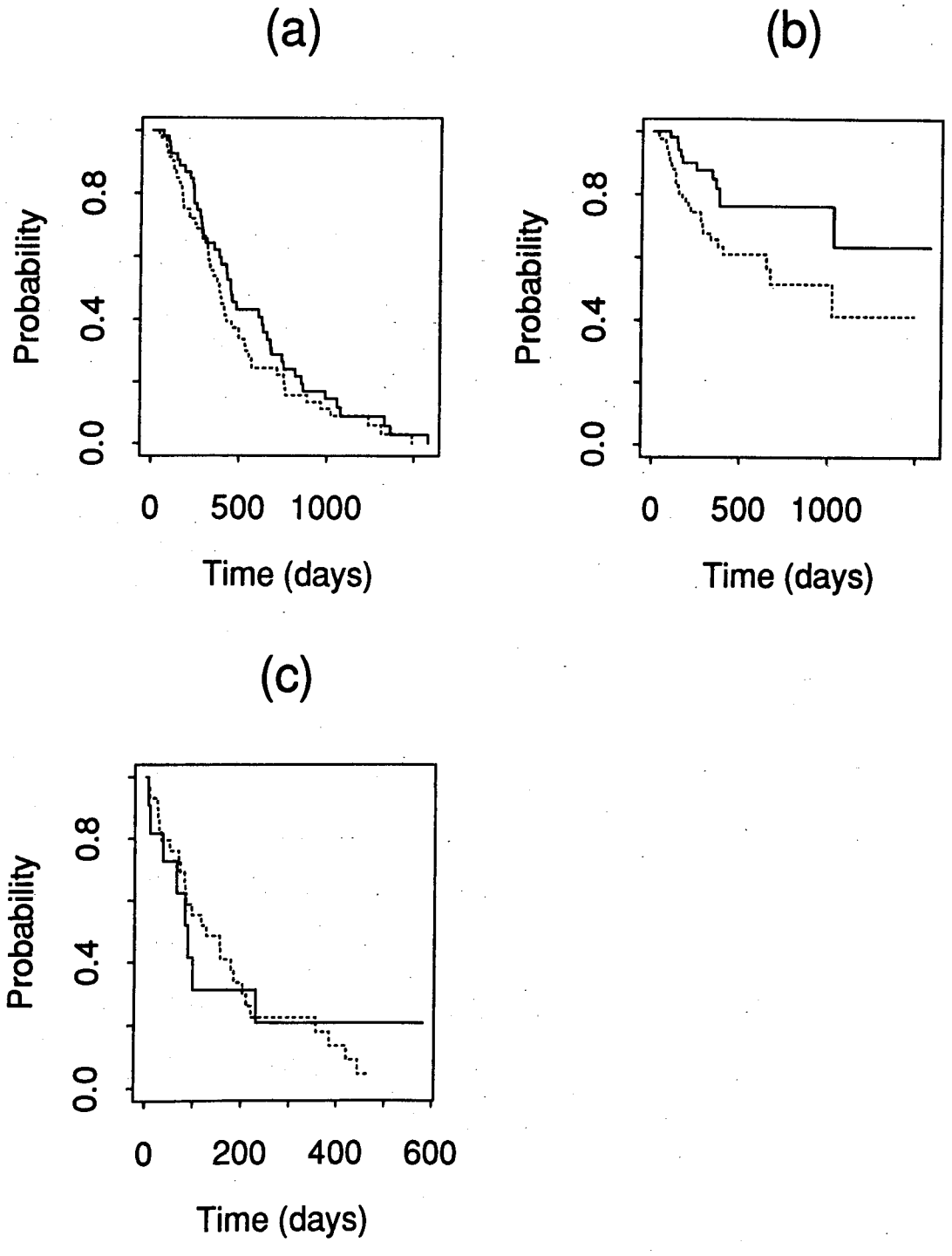
In this section we consider the use of multi-state models in analysing event history data that arises in quality of life studies. Over recent years there has been increased interest in quality of life studies, especially in cancer clinical trials, (Fayers and Jones, 1983, Olschewski and Schumacher, 1990, Fayers *et al.*, 1991, Pocock, 1991). Traditionally analysis of cancer clinical trials has focused on survival or disease-free survival, but with treatments only marginally increasing survival interest has focused on other issues such as treatment side-effects. Palliative treatment has also focused attention on other measures.

Having justified the need for measures other than survival much work has been done to develop measuring instruments that can measure a patient's quality of life, Bowling (1991) reviews many measuring instruments, whilst Fayers



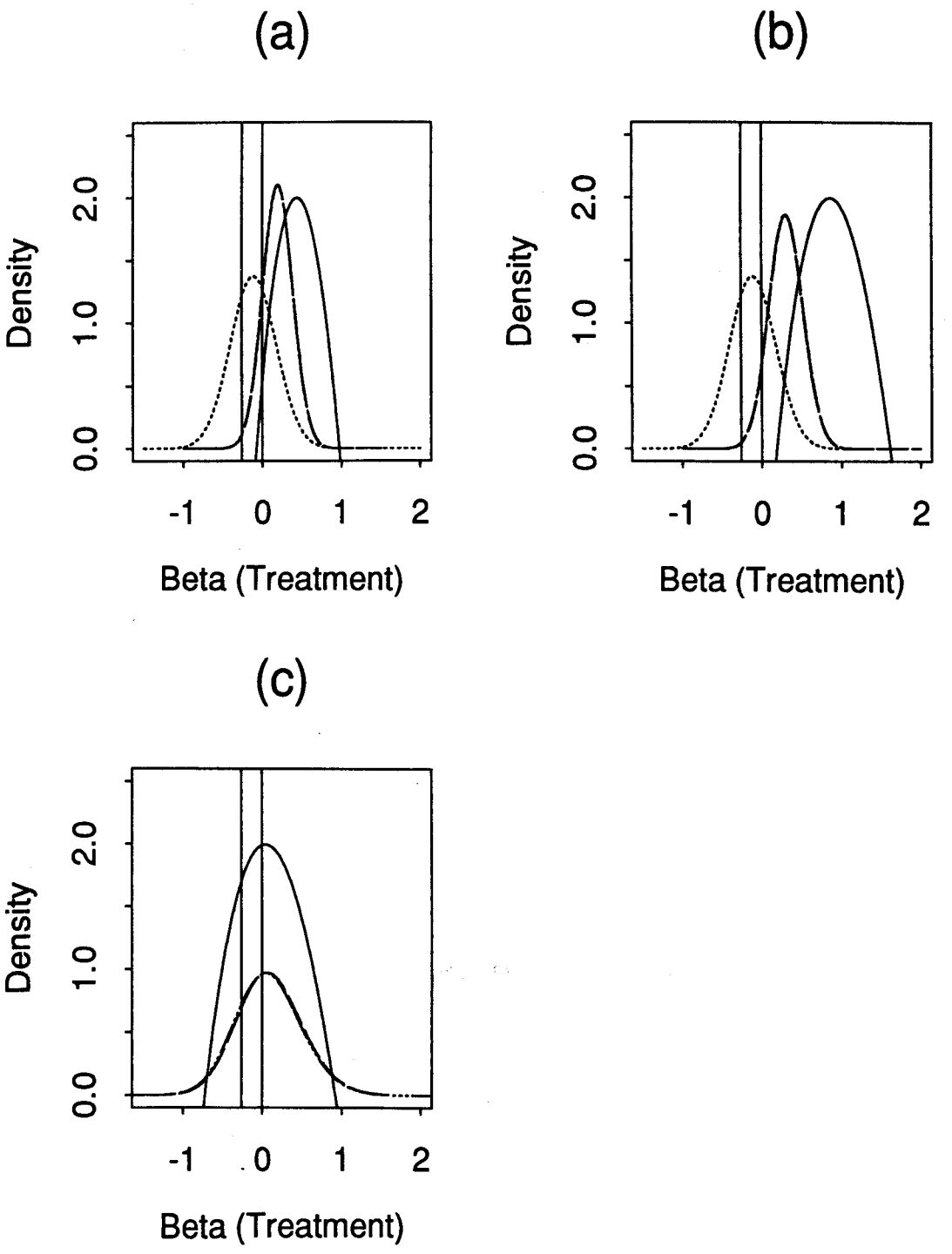
Key : ——— Metastases-Free, – – – – Metastases, • Death

Figure 7.5: Dynamic evolution of metastatic disease for neutron therapy example; (a) photon patients, (b) neutron patients.



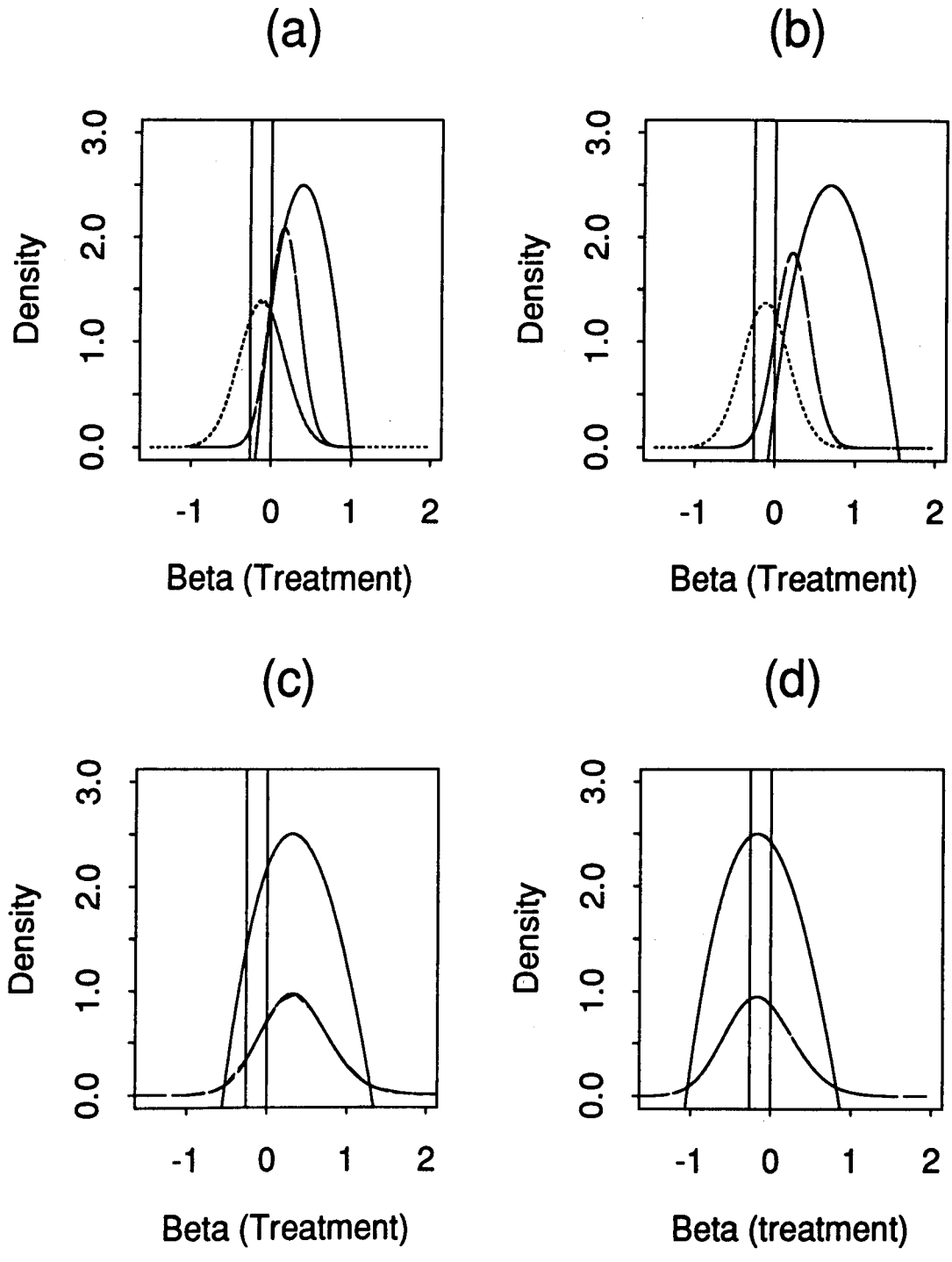
Key : —— photons, ····· neutrons; left and right vertical lines are at -0.26 and 0.0 respectively.

Figure 7.6: Kaplan-Meier estimated survival curves for various end-points; (a) treatment to death, (b) treatment to metastases and (c) metastases to death.



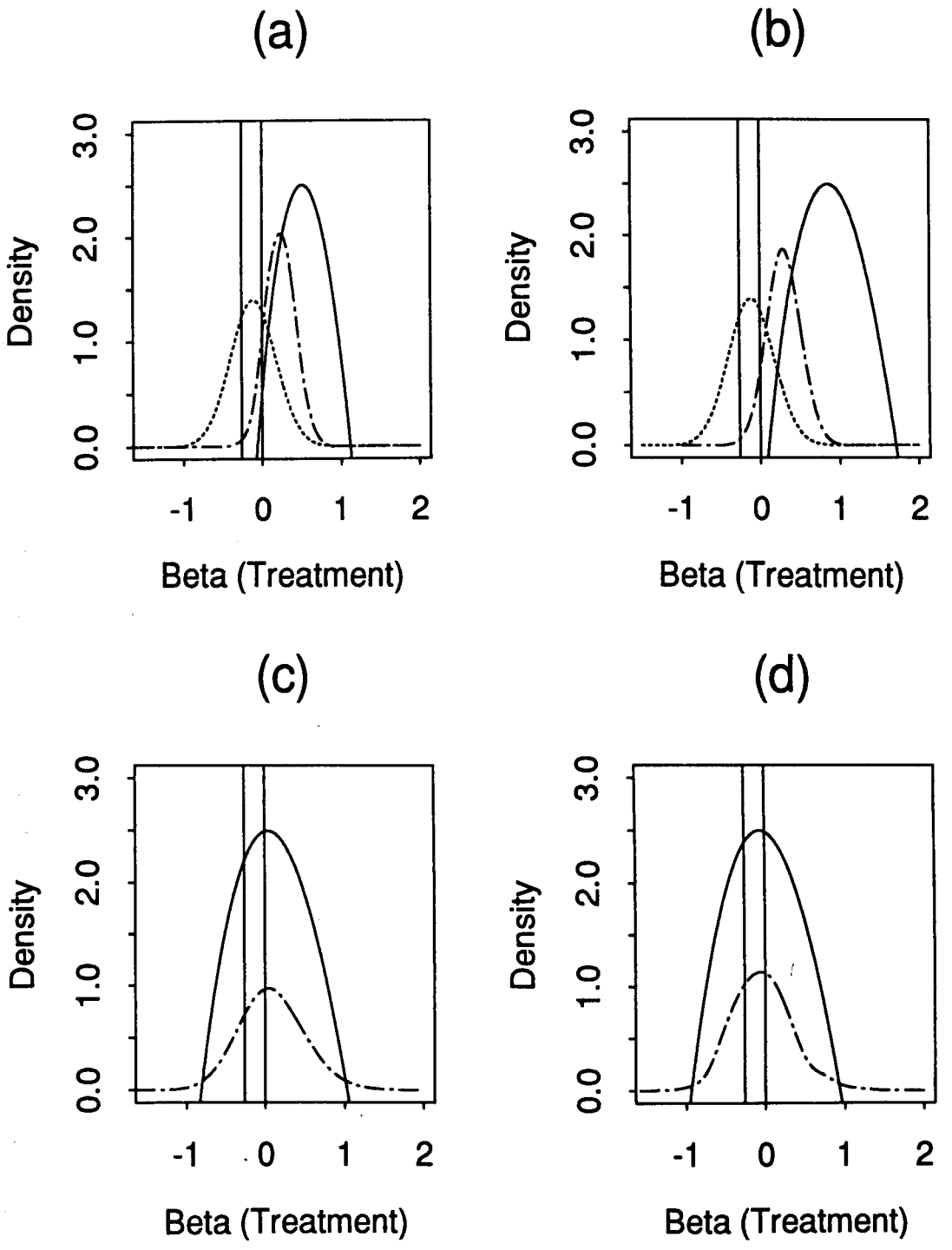
Key : — Maximum relative log-likelihood, ····· clinical prior, – · – · – · – clinical posterior (Laplace), — — clinical posterior (Gauss-Hermite); left and right vertical lines are at -0.26 and 0.0 respectively.

Figure 7.7: Maximum relative log-likelihood, clinical prior and marginal posteriors for β, treatment effect, in neutron therapy study assuming constant baseline intensities for (a) treatment to death (b) treatment to metastases and (c) metastases to



Key : —— Maximum relative log-likelihood, ····· clinical prior, – · – · – · – clinical posterior (Laplace), — — clinical posterior (Gauss-Hermite); left and right vertical lines are at -0.26 and 0.0 respectively.

Figure 7.8: Maximum relative log-likelihood, clinical prior and marginal posteriors for β, treatment effect, in neutron therapy study assuming piecewise constant baseline intensities for (a) treatment to death (b) treatment to metastases, (c) metastases to



Key : — Maximum relative log-likelihood, ····· clinical prior, – · – · – · – clinical posterior (Laplace), — — clinical posterior (Gauss-Hermite)

Figure 7.9: Maximum relative log-likelihood, clinical prior and marginal posteriors for β, treatment effect, in neutron therapy study assuming Weibull form baseline intensities for (a) treatment to death (b) treatment to metastases, (c) metastases to death (semi-Markov) and (d) metastases to death (Markov).

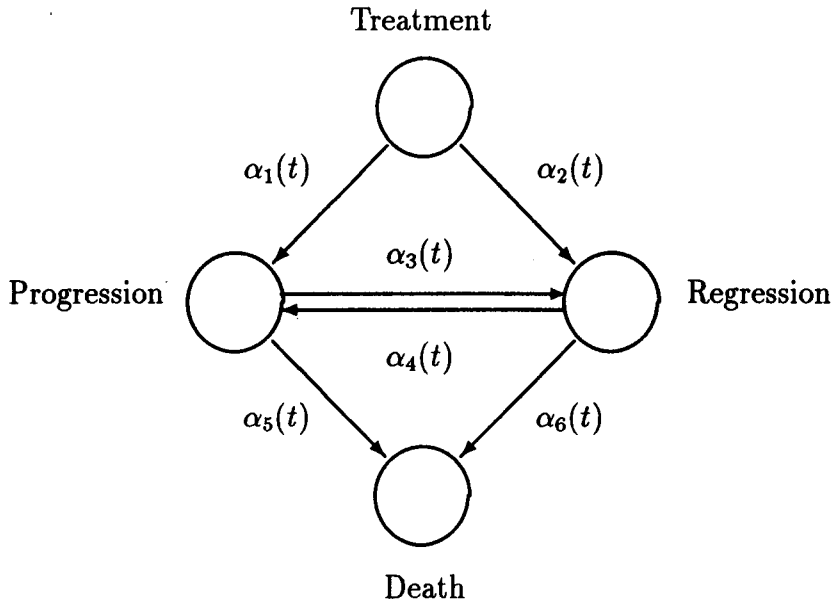


Figure 7.10: Multi-state model for tumour response, $\alpha_k(t)$ is the transition intensity for event k at time t.

and Jones (1983) review many of the scales that are used in cancer. The type of data that is generated by using such measuring instruments is not exclusively found in quality of life studies, but also in many sociological settings such as panel studies (Kalbfleisch and Lawless, 1985, DeStavola, 1986). Such data is often in the form of several *items* i.e question responses, on each of a number of *dimensions* e.g physical, sexual, social etc, and possibly recorded on a large number of *occassions*. There are two problems; the dimensionality of the data and the fact that there may be many observations over time.

Good graphical presentation of the data is fundamental, as has been advocated by Fayers (1991). Simple analysis within dimensions can be carried out using paired t-tests, as advocated by Cox *et al* (1992). As a simple analysis concentrates on the dimensionality of the data, so the types of more complex analysis that have been suggested rely on a reduction in the dimensionality of the data, but concentrate on the dynamic nature. There are essentially two methods that have been advocated; quality adjusted survival and multi-state models. We shall only briefly outline the former method, but will discuss in detail the latter.

Galsziou, Simes and Gelber (1990) advocate the use of quality adjusted survival analysis in which a patient's total survival time is split into a finite number of intervals. Each interval is then weighted according to the patient's quality of life in that interval. A quality adjusted survival time is then obtained by taking

the sum of the weighted interval lengths. The advantage of this method is that once the quality adjusted survival times have been formed, traditional survival analysis techniques may then be used. The disadvantage of such a method is that the parameter estimates that are eventually obtained are difficult to interpret. In common with multi-state models this method also relies upon there being a single measure of quality of life. This is a contentious issue. A number of authors have argued in favour of a global measure, including Tandon (1990) and Schumacher, Olschewski and Schulgen (1991), but more recently Cox *et al* (1992) have argued against the use of such a measure, especially across dimensions.

Multi-state methods have been advocated by Olschewski and Schumacher (1990) and Abrams (1992). They too have the disadvantage that they require there to be a single measure. This measure may either be a global quality of life measure or it may be a measure within a dimension, but across items in that dimension. The longitudinal nature of the data is then reflected in patients moving between different quality of life states. Though these assumptions have been critised they do enable easily interpretable quantities to be obtained, e.g the probability of deterioration/improvement in quality of life or mean time spent in a specific quality of life state, both conditional on covariates. Another disadvantage that has been noted is the quality and quantity of the data required for such models. This is becoming less of an issue as interest in quality of life increases and such data is *routinely* collected as part of cancer clinical trials. As an illustration of the use of such multi-state models we will consider the pilot quality of life study described in Chapter 2.

7.5.1 Methods

In Chapter 2 the formation of a global index was described, so that patients could be in either a 'good', 'medium' or 'poor' quality of life state at any observed time. In this pilot study we are interested in comparing the quality of life experienced by lung cancer patients with that of other cancer patients, i.e those with cancer of the testis, ovary and skin (melanoma). Table 7.5 shows the number of transitions between states that were observed for the two patient groups. We can see that for two of the transitions, 'good' to death and 'poor' to 'medium' (for at least one of the patient groups) there were no transitions. We shall therefore ignore these two transitions. Figure 7.11 shows the possible transitions diagrammatically, with $\alpha_k(t)$ denoting the transitions intensity for the kth transition. Figure 7.12 shows the dynamic evolution of the patients' quality of life with different line types representing the different quality of life states. As we will consider the case of a time-homogeneous Markov model, $\alpha_k(t)$ will be constant. $\alpha_k(t)$ is parameterised as in the case of the metastases example in Section 7.3, so that for non-lung cancer patients the intensity for the kth transition is e^{θ_k} and for lung cancer patients it is $e^{\theta_k+\beta_k}$. Therefore e^{β_k} represents the relative risk of the transition for lung cancer patients compared to non-lung cancer patients. Accordingly β_k is the logarithm of the relative risk

for the kth transition with a value of zero indicating no patient group difference.

Using this parameterisation of $\alpha_{ik}(t)$ (7.1), and the likelihood (7.2) described in Section 7.2 and assuming uniform prior probability density functions for all parameters we can obtain the joint posterior density function. Parameter estimates and marginal posterior densities can be obtained using either the Laplace approximations suggested by Tierney and Kadane (1986) or Gauss-Hermite quadrature advocated by Naylor and Smith (1982) both methods were described in detail in Chapter 6.

7.5.2 Results

Table 7.6 shows the posterior parameter estimates obtained using both Laplace approximations and Gauss-Hermite quadrature, and as a comparison those obtained using maximum likelihood. Figures 7.13 and 7.14 show the maximum relative log-likelihood and posterior marginal probability density functions for the β parameters.

We can see from Table 7.6 that considering overall survival, lung cancer patients are at increased risk of dying, relative to non-lung cancer patients, with virtually no posterior probability that the relative risk is less than one, i.e a relative risk greater than one indicating that lung cancer patients are at increased risk compared to non-lung cancer patients. Considering patients who are in the 'good' quality of life state, we can see that lung cancer patients have increased risk of deteriorating and entering the 'medium' quality of life state. Again there is virtually no posterior probability that the relative risk is less than one. As may be expected considering patients who are in a 'medium' quality of life state the relative risk of improving, favours non-lung cancer patients with a posterior probability of approximately 66%. Still considering patients who are in a 'medium' quality of life state, there is an increased risk of deterioration to a 'poor' quality of life state for lung cancer patients compared to non-lung cancer patients, with a posterior probability of approximately 80%. Lung cancer patients are also at an increased relative risk of dying from a 'medium' quality of life state with a posterior probability of approximately 65%. Finally, considering those patients who are in a 'poor' quality of life state there is an increased relative risk of dying for non-lung cancer patients with a posterior probability of approximately 77%.

7.5.3 Clinical Interpretation

Thus, overall lung cancer patients not only have an increased relative risk of dying, but they also have an increased relative risk of their quality of life deteriorating regardless of their present quality.

7.5.4 Discussion of Quality of Life Example

This analysis was intended as an illustration of the use of multi-state models in quality of life studies. The interpretation of the parameter estimates needs care, as they are based on relatively small numbers of transitions. This explains the relatively large standard deviations seen in Table 7.6. Although both the Laplace approximations and Gauss-Hermite quadrature gave results which were the same for all practical purposes, there were larger discrepancies than had been seen in previous applications in this thesis. Convergence of both `nlmin` and BAYES4 were noticably slower and more difficult to achieve than in other examples, especially considering that for each transition only two parameters were estimated. Laplace approximations require that n in (6.11) be 'moderate'; both large and small n can seriously affect parameter and density estimation, (Tierney and Kadane, 1986, Tierney *et al.*, 1989). In the case of this study n is small, being at most 18.

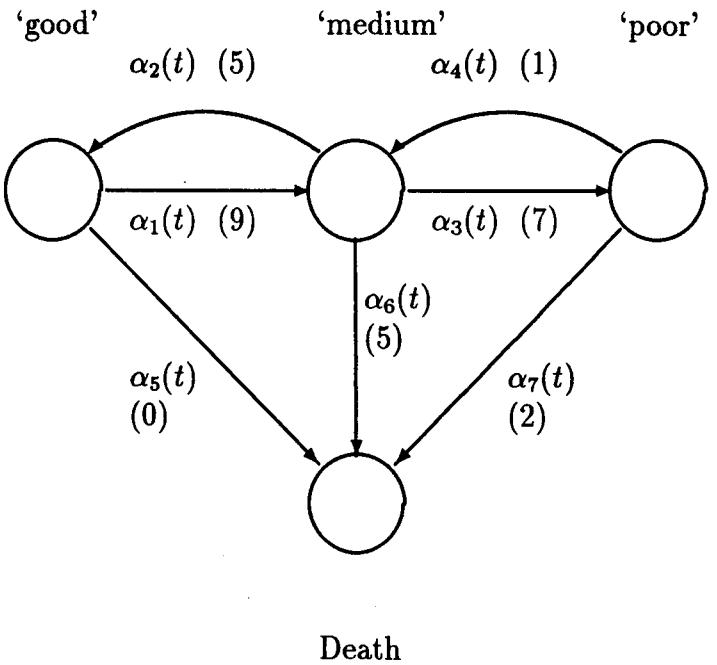


Figure 7.11: Multi-state model for pilot quality of life study, $(\cdot)$ denotes the number of patients who made a particular transition. and $\alpha_k(t)$ is the transition intensity for event k.

		'good'	'medium'	'poor'	death
'good'	non-lung	8	5	0	0
	lung	1	4	0	0
'medium'	non-lung	2	8	3	2
	lung	3	5	4	3
'poor'	non-lung	0	0	1	2
	lung	0	1	3	0

Table 7.5: Transitional status for lung and non-lung cancer patients in pilot quality of life study.

7.6 General Discussion of Multi-State Models

In this chapter we have seen how the two-state models of Chapter 6 can be extended to the case when there are more than two states, if we are prepared to assume that the times of transitions are known. In clinical trials such an assumption would seem reasonable as patients are followed-up more closely than in obervational studies. We have been able to consider both time-homogeneous and time-inhomogenous Markov models in which the risk of particular transition is dependent only on the current state that a patient is in, and does not depend on the length of time that has been spent in that state. We have also considered the case when we do allow for the sojourn times in a time-homogeneous, but sojourn inhomogeneous semi-Markov model. A key aspect of these models has been the incorporation of prior information about specific transitions into the formal statistical analysis by adopting Bayesian methodology.

The advantage of the models developed in this chapter is that transition specific covariates may be included in the model so that their influence on intermediate events may be evaluated. This is the key concept of multi-state models; that the overall survival process is decomposed into its constituent components, and a clearer picture of the disease/treatment process is obtained.

We have applied Bayesian multi-state models to the development of metastases in the neutron therapy study. Using such models we have been able to address the clinically important issue of what role the development of metastases plays in helping to explain the treatment differences.

A further extension of the models presented in this chapter would be to allow for a semi-Markov model in continuous time which is inhomogeneous with respect to both trial time and sojourn time. Thus for the general multiplicative model (7.1) the intensity would be of the form

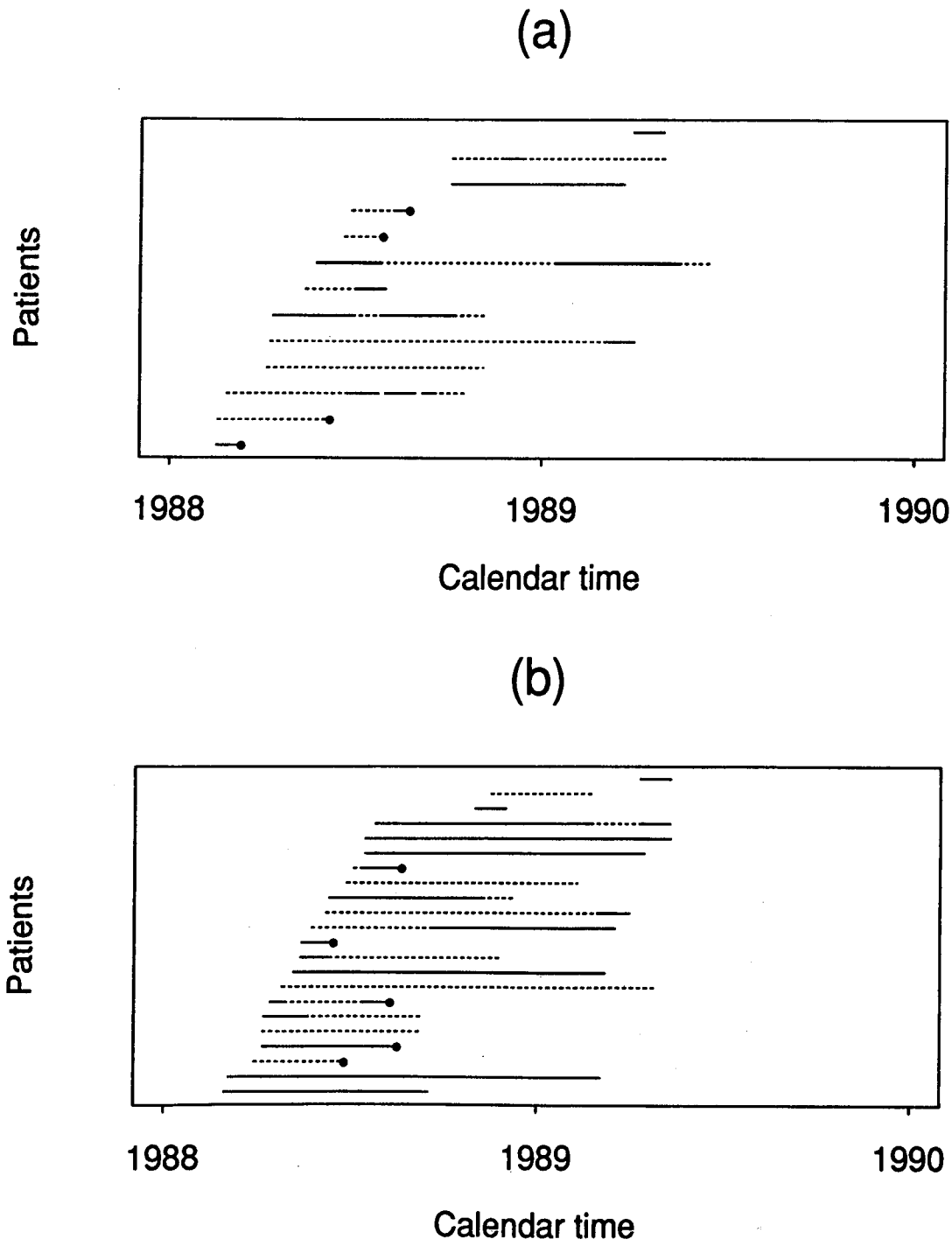
$$\alpha_{ik}(t) = \lambda_{0k}(t)\, e^{\gamma_k x_{ik}(s,t)}\, e^{\underline{\beta}_k^T \underline{z}_{ik}(t)}\, Y_{ik}(t)$$

where s is the time of event k in overall trial time, $x_{ik}(s,t)$ is the sojourn time for patient i at time t for event k, and $\underline{z}_{ik}(t)$ are other possibly time-dependent covariates specific to event k for patient i. In the metastases example above, given that a patient had already developed a metastases, the probability that they would die within a short space of time would be a function of both the

Transition	θ (Baseline)		β (Lung)		
	Mean	SD	Mean	SD	$P_{(0)}$[a]
Survival					
Prior	-	∞	-	∞	-
Posterior (L)[b]	-7.344	0.318	1.801	0.428	0.000
Posterior (G-H)[c]	-7.345	0.324	1.801	0.438	0.000
MLE	-7.294	0.316	1.792	0.428	-
'good' to 'medium' ($k = 1$)					
Prior	-	∞	-	∞	-
Posterior (L)	-5.932	0.437	2.502	0.670	0.000
Posterior (G-H)	-5.935	0.470	2.500	0.711	0.001
MLE	-5.831	0.446	2.526	0.670	-
'medium' to 'good' ($k = 2$)					
Prior	-	∞	-	∞	-
Posterior (L)	-6.057	0.412	-0.314	0.728	0.667
Posterior (G-H)	-6.196	0.455	7 -0.316	0.748	0.664
MLE	-6.094	0.445	-0.247	0.728	-
'medium' to 'poor' ($k = 3$)					
Prior	-	∞	-	∞	-
Posterior (L)	-6.627	0.563	0.617	0.763	0.169
Posterior (G-H)	-6.636	0.628	0.621	0.823	0.220
MLE	-6.460	0.575	0.575	0.762	-
'medium' to death ($k = 6$)					
Prior	-	∞	-	∞	-
Posterior (L)	-7.279	0.684	0.343	1.001	0.366
Posterior (G-H)	-7.286	0.738	0.343	1.044	0.371
MLE	-7.029	0.700	0.342	0.993	-
'poor' to death ($k = 7$)					
Prior	-	∞	-	∞	-
Posterior (L)	-4.581	0.560	-0.692	0.910	0.778
Posterior (G-H)	-4.584	0.594	-0.696	0.947	0.769
MLE	-4.415	0.577	-0.609	0.912	-

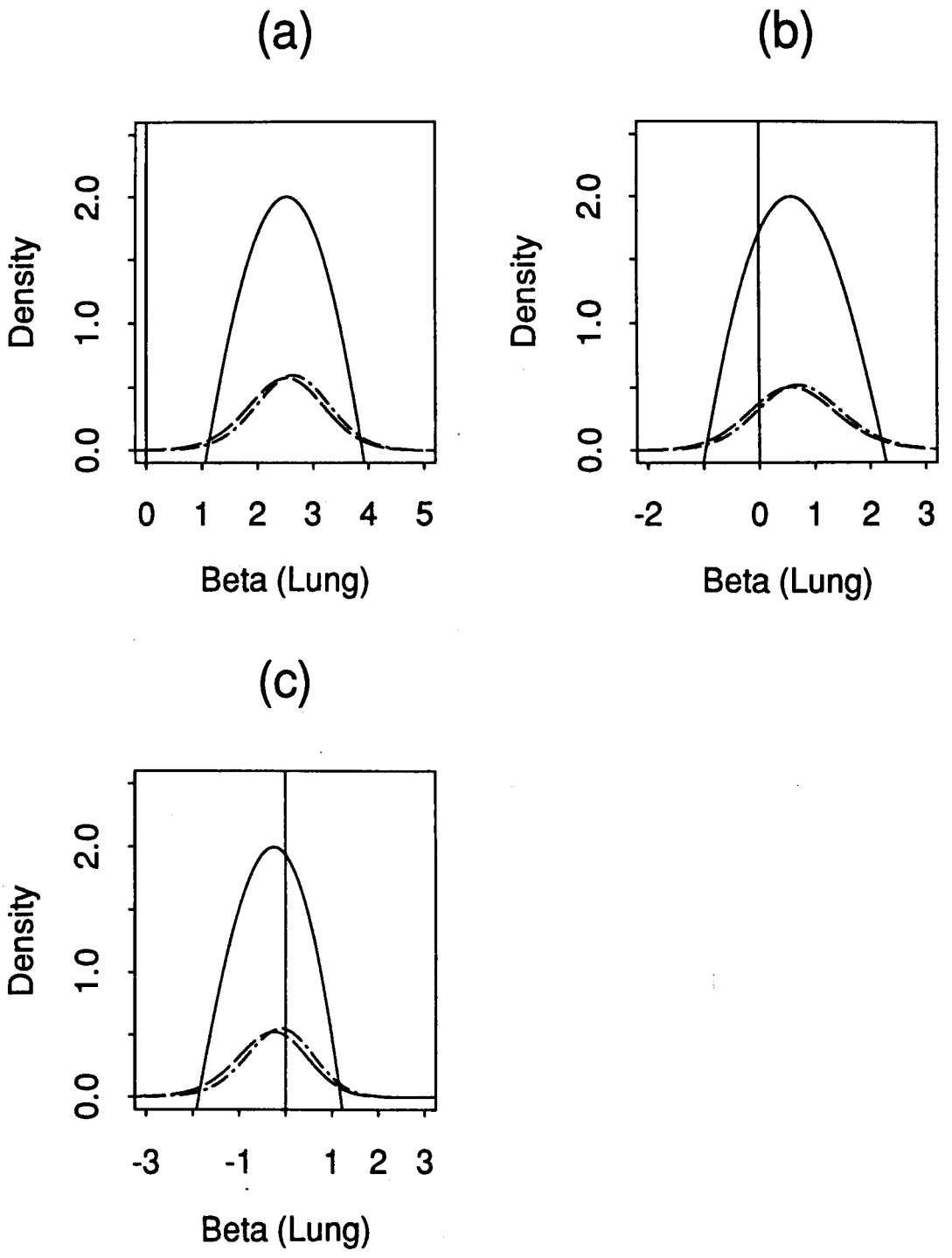
[a] $P_{(0)}$ denotes the probability that β is less than 0.
[b] L denotes Laplace approximation.
[c] G-H denotes Gauss-Hermite quadrature.

Table 7.6: Parameter estimates for time-homogeneous Markov model for pilot quality of life study.



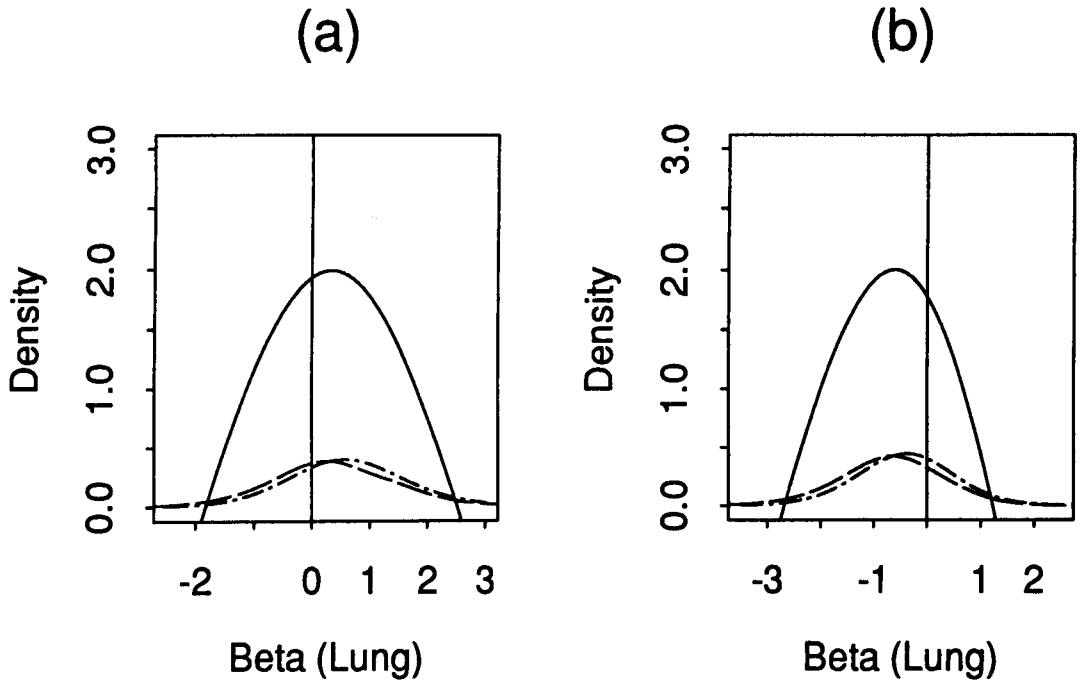
Key : ——— 'good', ····· 'medium', – – – 'poor', • death.

Figure 7.12: Dynamic evolution of event histories for quality of life study; (a) lung cancer patients and (b) non-lung cancer patients.



Key : —— maximum relative log-likelihood, – · – · – · – posterior (Laplace), — — posterior (Gauss-Hermite).

Figure 7.13: Maximum relative log-likelihood and marginal posterior densities for β in quality of life study; (a) 'good' to 'medium', (b) 'medium' to 'good', (c) 'medium' to 'poor'.



Key : —— maximum relative log-likelihood, – · – · – · – posterior (Laplace), — — posterior (Gauss-Hermite).

Figure 7.14: Maximum relative log-likelihood and marginal posterior densities for β in quality of life study; (a) 'medium' to death and (b) 'poor' to death.

current position in trial time and the length of time for which they had had a metastases, i.e

$$x_{ik}(s,t) = \begin{cases} t-s & \text{if } t > s \\ 0 & \text{if } t \leq s \end{cases}$$

We have also considered the application of multi-state models to data from a pilot quality of life study. This is an important and expanding area in cancer clinical trials, in which survival is not always the most appropriate endpoint. Multi-state models allow more clinically meaningful quantities to be estimated than those using traditional quality-adjusted survival analysis techniques. Multi-state models can be applied either globally or within particular dimensions, e.g to estimate the relative risk of a patient's physical ability improving, given their current status, on one treatment compared to another. This example also demonstrates the need for 'good' quality data if the modelling assumptions we have had to make can be justified.

Chapter 8

Summary and Conclusions

In this chapter we summarise the results of this thesis, both statistical and clinical, discuss the possible application of the methods developed to epidemiology, and outline areas for further work.

8.1 Statistical summary

We saw in Chapter 3 that in cancer clinical trials there is often prior information in the form of both clinical beliefs and results from previous studies. Careful consideration has to be given to the combination of clinical beliefs, especially when there are major discrepancies between individuals. The combination of the results from previous studies is helped by the methodology that has been developed for conducting meta analyses, though Bayesian methods are potentially useful. The elicitation of the clinical beliefs in the neutron therapy study was done primarily to establish whether continued randomisation of patients into the study was ethical. Such an elicitation process serves not only to quantify beliefs about possible treatment difference but also the size of a difference required to change clinical practice.

In a wider context, in the pharamceutical industry, elicitation and uses of prior information have a role to play. Within individual companies there is a very real role for Bayesian methods in terms of drug development, i.e identifying which drug is worth pursuing, when the evidence may come from a number of dispirit sources. In the area drug regulation acceptance of Bayesian methods is largely governed by the regulatory authorities, i.e Committee on Saftey of Medicines in the U.K, and the Food and Drug Administration in the United States, and in principle there is now acceptance of them. However, use is currently limited by the lack of suitable well understood models and software.

Having obtained and quantified prior information, the non-temporal models described in Chapter 4 can be used in an initial analysis. In particular an odds model approach, based on assuming Beta conjugate prior distributions for the failure rates in each of the two treatment groups separately, and a Normal theory approach, assuming the hazard ratio to be Normally distributed with unknown mean, but known variance, and a Normal prior distribution for the mean, are straight forward to implement and do not necessarily require sophisticated software. The major disadvantages with such models is that they can only accommodate two patient groups, and that they cannot explicitly allow for differential follow-up or censoring. The first disadvantage may be addressed by using a logistic regression model, but the complexity of implementation is greatly increased, whilst still not addressing the second problem of differential follow-up. Non-temporal models also suffer from the disadvantage that they are not able to consider more than one end-point per patient.

In order to address both the problem of the inclusion of covariates and that of differential follow-up a fully parametric multiplicative intensity model was developed. A fully parametric model was developed for two reasons; firstly, there are many situations where there is evidence to suggest that a particular

parametric form may be appropriate, and secondly, as a set of alternative models to the now almost *de facto* semi-parametric proportional hazards model. We considered three different parametric models. The first, and simplest, was a model in which the baseline intensity was constant over time. This model corresponds to assuming the times to death, in a survival analysis problem, are exponentially distributed, or in generalised model terms a Poisson regression model. As an extension of this model, we considered the case when there are a number of time intervals, which are chosen arbitarily, and the baseline intensity was constant within each one. Finally, a model in which the baseline intensity is a power transform of the time scale was developed. This model corresponds to assuming that the times to death, in a survival analysis problem, follow a Weibull distribution.

In the case when there were two states, the second of which was absorbing, the model was applied to the survival data in the neutron therapy trial. The model parameters were estimated using both Laplace approximations and Gauss-Hermite quadrature. The two methods gave virtually identical results in terms of both moment estimates and marginal posterior densities. It was also shown that parameter estimation in this class of models does not require the use of computationally intensive simulation methods such as Gibbs sampling. Though Gauss-Hermite quadrature requires sophisticated software the Laplace approximations only require the ability to maximise a function of several variables, and therefore could be implemented easily using standard software. The conclusions about the efficacy of neutron therapy compared to photon therapy for tumours of the pelvic region were not different to those obtained using the non-temporal models of Chapter 4. This can be explained by the fact that we found little evidence for the baseline intensity varying with time. For data sets in which this is not the case then a difference between the temporal and non-temporal models could be expected.

The two state model, developed in Chapter 6, was extended to the multi state case where there are a number of states a patient may be in at a specific time point. Such models allow for the survival process to be decomposed into constituent components, and enable a greater depth of understanding. By developing the two-state models in Chapter 6 using exclusively counting process notation the extension to the multi-state scenario was straight forward. The multi-state models of Chapter 7 were applied to two examples. The first was a three state illness-death/disability model, to assess the effect of metastatic disease spread on survival in the neutron therapy study. The three different parametric forms for the baseline intensity allowed both Markov and semi-Markov model assumptions to be tested. The second example was a multi state model applied to a pilot quality of life study. Due to the size of the study only a time-homogeneous Markov model was possible. In both examples parameter estimation was performed using Laplace approximations and Gauss-Hermite quadrature. As in the two state model the results using these methods were for practical purposes the same.

We have shown that Bayesian multi-state models provide a coherent frame-

work for the analysis of clinical trials, yielding clinically meaningful summaries of current state of knowledge about the disease/treatment process.

8.2 Clinical interpretation

Neutron therapy example

The clinical beliefs that were elicited indicated that nine out of the ten clinicians thought that, for the treatment of pelvic tumours, high energy neutron therapy was going to be more beneficial than photon therapy. Only one clinician thought that photon therapy was likely to be more beneficial than neutron therapy. A meta analysis of five published and one unpublished studies indicated that low energy neutron therapy were unlikely to be more beneficial, than photon therapy, for these patients.

In the light of the trial, aggregated clinical beliefs should change considerably so that *a posteriori* the clinicians, as a group, would only consider there to be a small chance of neutrons being more beneficial that photons in terms of overall survival. Beliefs based on the results of the six previous studies would only change slightly in the light of the current study.

Further analysis shows that neutron patients are at increased risk of developing clinically detected metastases or dying without having developed them. Once they have developed there appears little evidence of a treatment difference even after allowing for time of development. Though information about metastatic disease found at autopsy was available, patients who received neutron therapy were more likely to have an autopsy. Therefore, use of this data could introduce serious bias into any analysis.

In using a Bayesian approach in the analysis of this study we have assessed the current weight of evidence for the use of neutron therapy for tumours of the pelvic region based not only on the *current* study, but also on previous trial results and clinical opinion. At an individual, or group, level such an approach formalises what often happens in practice, i.e that the results from the *current* study are interpreted in the light of previous findings. The Bayesian approach also has the advantage that we are formally able to synthesize information from dispirit sources about the role neutron therapy has to play in treating pelvic tumours in a coherent manner.

Quality of life example

Analysis of the pilot quality of life study showed that lung cancer patients were not only at increased risk of death relative to other patients, but also of a deteriorating quality of life, regardless of their present quality.

Though these results are not surprising, the quantities that were obtained from the multi-state models are easily interpretable. Such models allow an analysis either globally, i.e ignoring dimensions, or within specific dimensions. Thus, for a complex analysis of quality of life studies, multi-state models have

an advantage over quality-adjusted survival techniques. A fundamental need for good graphical presentation of the results was also highlighted. As an example the use of Lexis type diagrams can clearly show patients' changing quality of life status, though further work in this area is required.

8.3 Limitations of the Bayesian approach

We have shown that the Bayesian approach aids the clinical understanding of complex survival data. However there are a number of limitations of such an approach. The first major limitation of the Bayesian approach over the classical approach is that there is no clear modelling strategy to adopt, either in terms of model selection or model comparison. Whilst in classical survival analysis 'goodness of fit' of models has been addressed by some authors, see Kay (1984) for a review, in Bayesian survival analysis, with the exception of Chaloner (1991), little work has been done.

In the analysis of the neutron therapy study a number of different parametric models were considered. Whilst none of the different models gave radically different results, assessing which is the 'best' model is difficult. For this particular study, of the models considered, the most appropriate is one in which there are three time intervals; less than six months, six months to a year and greater than a year. However, further analysis may show that there is evidence of a time-treatment interaction, thus violating the multiplicative intensity assumption.

Another limitation of the Bayesian approach is that it requires careful interpretation. We saw in Section 6.5.4 that arbitary use of only the marginal posterior densities can lead to unusual results and can be misleading. A full analysis requires that both the marginal densities and the bivariate densities are examined.

8.4 Links with epidemiology

Good medical practice develops by the accumulation of information over time. In particular, the process of sequential accumulation of information about a particular hypothesis occurs in epidemiology. An initial hypothesis may be generated by laboratory experiments or biological theory, a case-control study may be performed to initially test this hypothesis, and subsequently a large prospective cohort study may be conducted. After each study the belief about the hypothesis should be updated. Therefore Bayesian methodology has a potential role to play in epidemiology.

The non-temporal models of Chapter 4 may be applied to any type of study in which there are two patient groups and the response is dichotomous. In particular models that use the odds ratio will be invariant to the sampling scheme used and are particularly appealing. The odds models and the Normal theory models are both convenient tools for an initial analysis. When either

there are a number of explanatory factors or when patients are matched, the logistic regression models outlined in Chapter 4 would be suitable, though these would typically involve the use of approximate integration techniques.

In prospective cohort studies the multiplicative intensity models developed here would be appropriate in modelling the risk of an event using time from exposure. In particular, the piecewise constant intensity model would allow an analysis by age cohort. The multi-state multiplicative models would be useful when there is information on changing exposure patterns in an occupational cohort, and individuals can be thought to move between a finite number of exposure states.

8.5 Further work

There are a number of areas for future work.

A fruitful area for future work is the use of Bayesian methodology for summarising previous trial results. One way in which such an approach may help is in highlighting the sequential nature of meta analyses. This is an area that is not accommodated in either the fixed effects model or the random effects model. Lau, Antman, Jimenez-Silva, Kupelnick, Mosteller, and Chalmers (1992) have demonstrated the clinical significance of using such sequential methods in trials to the assess the efficacy of streptokinase for acute myocardial infarction.

The two-state multiplicative intensity models of Chapter 6 were extended to the case when there were at most three explanatory factors. In theory there is no difficulty extending them further to accommodate a larger number of covariates. As we have seen when the number of model parameters increases, use of optimisation routines such as those in NAG is required in order to implement the Laplace approximations discussed in this thesis. Therefore further work is required to produce a general Splus function, which makes use of NAG, for optimisation and numerical estimates of the hessian matrix, but allows a general model to be specified as a Splus object. However, as we have seen for models with many parameters, careful parameterisation is necessary together with careful interpretation.

Although the potential use of Markov Chain Monte Carlo, and in particular Gibbs sampling, was mentioned, it was not used as the presense of censoring meant that sampling from the conditional distributions was not straightforward. Recently Kuo and Smith (1992) and Smith and Roberts (1993) have shown that this particular problem can be overcome by treating the censored observations as further unknown model parameters that need to be estimated. Further work is required to compare the results using Gibbs sampling with those obtained using Laplace approximations and Gauss-Hermite quadrature. Experience of the use of Gibbs sampling in the relatively well behaved problems studied in this thesis may give insight into its potential use in more complex and less tractable problems.

The baseline intensity in all the models developed in this thesis can be con-

sidered restrictive. One possible way of relaxing this assumption is to assume that the baseline intensity can be approximated by a continuous smooth curve, such as a spline (Wegman and Wright, 1983). Taulbee (1979), Gilks (1986), Durrleman and Simon (1989) and Sleeper and Harrington (1990) have all considered the use of splines in regression models for survival data from a classical perspective. Only Shaw (1988) has considered taking a Bayesian approach to the use of splines in survival models. This area would appear to be a fruitful one for further work.

The models considered here all assume that the effect of model covariates remains constant over time clearly this may not be the case. In the piecewise constant intensity model this can be allowed for in theory by assuming coefficients to vary between epochs but to be constant within each epoch. This is similar to the approach of Gamerman (1987a) though without necessarily assuming an auto-regressive structure. Further work is also required to determine the 'best' method for defining the time intervals in peicewise constant intensity models, i.e whether the should be pre-specified or whether they should be data dependent.

The multi-state models developed here use one of two time scales; either the length of time a patient has been in a particular state (sojourn time) or the trial time. Ideally we would like to be able to accommodate both scales within the same model. The most obvious way of doing this is to allow the baseline intensity to be on the trial time scale, and for the sojourn times to enter the model as time-dependent covariates.

As we have already mentioned one of the key assumptions of the models developed here is that baseline intensity can be adequately approximated by a parametric form, such as a piecewise constant function or a power transform. An alternative approach is to allow the baseline intensity to be non-parametric, resulting in a semi-parametric model similar Cox's proportional hazards model. A number of authors have discussed such a model from a Bayesian perspective, including Hjort (1986), Clayton (1991) and Carlin, Chaloner, Church, Louis and Matts (1992). However, one of the contributions of this thesis has been to provide a number of alternative parametric survival models to the now almost *de facto* semi-parametric proportional hazards model.

Appendix A

MRC Neutron Therapy Trial

A.1 Column Descriptions

Column 1 Patient Number

Column 2 Survival Time (days), using 21st December 1990 as censoring date.

Column 3 Death, 0 = Alive, 1 = Dead

Column 4 Treatment, 0 = Neutrons, 1 = Photons

Column 5 Site, 1 = Cervix, 2 = Rectum, 3 = Bladder, 4 = Prostate

Column 6 Phase, 0 = randomised before 10th January 1988, 1 = randomised after 10th January 1988.

Column 7 Metastases diagnosed before death, 0 = No Metastases, 1 = Metastases

Column 8 Time to Metastases (days) or death whichever is the sooner

Note

One patient has a missing value for time to metastases, denoted 'NA', and ten patients did not have sufficient follow-up for the detection of metastases, i.e time to metastases of zero.

A.2 Data

1	139	1	0	2	0	1	71
2	325	1	0	2	0	1	147
3	111	1	0	2	0	1	87
4	309	1	0	3	0	0	287
5	549	1	0	2	0	1	129
6	219	1	1	3	0	1	137
7	451	1	0	2	0	0	386
8	320	1	0	1	0	1	288
9	1241	1	0	2	0	1	1021
10	316	1	0	2	0	0	213
11	477	1	1	1	0	0	449
12	734	1	0	2	0	0	534
13	1490	1	1	2	0	0	1058
14	209	1	0	3	0	1	202
15	1597	0	1	1	0	0	1581
16	825	1	0	2	0	0	757
17	301	1	0	3	0	1	274
18	1115	1	0	4	0	1	671
19	177	1	0	2	0	0	0
20	1036	1	1	3	0	1	1031
21	1533	0	0	1	0	0	1489
22	493	1	0	2	0	0	456
23	99	1	0	3	0	0	80
24	465	1	1	2	0	1	377
25	574	1	0	3	0	0	496
26	127	1	0	3	0	1	122
27	145	1	0	2	0	0	115
28	222	1	0	2	0	1	38
29	152	1	0	2	0	0	49
30	43	1	0	2	0	0	0
31	1375	0	1	1	0	0	1331
32	310	1	1	3	0	0	274

33	325	1	0	3	0	0	284
34	215	1	0	3	0	0	170
35	1366	0	1	1	0	0	1363
36	165	1	1	1	0	1	157
37	224	1	0	3	0	0	179
38	1348	0	0	3	0	0	1312
39	247	1	0	3	0	0	171
40	1340	0	0	2	0	0	1238
41	269	1	1	3	1	0	240
42	587	1	0	3	0	0	495
43	11	1	1	3	0	0	0
44	125	1	1	2	0	0	101
45	291	1	0	2	0	0	250
46	176	1	0	3	0	1	104
47	228	1	0	3	0	0	176
48	769	1	0	3	0	0	715
49	154	1	0	2	0	0	138
50	474	1	0	4	0	1	272
51	814	1	1	2	0	0	612
52	126	1	0	3	0	1	76
53	79	1	0	3	0	0	77
54	1114	0	0	4	0	1	648
55	172	1	0	2	1	1	76
56	1110	0	1	1	0	0	1078
57	430	1	0	1	0	0	415
58	486	1	0	2	0	1	129
59	577	1	0	3	1	0	570
60	161	1	0	3	1	0	79
61	37	1	1	3	1	0	NA
62	1037	0	0	3	1	0	1021

63	224	1	0	4	1	1	142
64	173	1	0	3	1	0	158
65	364	1	0	1	1	1	283
66	199	1	0	2	1	0	136
67	959	0	1	2	1	1	377
68	229	1	0	3	1	1	113
69	995	0	1	3	1	0	990
70	299	1	1	3	1	0	280
71	773	1	1	3	1	0	743
72	133	1	1	2	1	1	97
73	968	0	1	1	1	0	854
74	472	1	0	2	1	0	384
75	969	0	0	3	1	0	964
76	232	1	1	3	1	1	167
77	535	1	1	1	1	0	451
78	186	1	1	2	1	0	139
79	246	1	1	3	1	0	186
80	44	1	1	2	1	0	0
81	417	1	0	1	1	0	329
82	744	1	0	2	1	0	399
83	149	1	1	2	1	0	100
84	239	1	0	3	1	1	29
85	141	1	0	3	1	0	37
86	886	0	1	3	1	0	865
87	92	1	0	3	1	0	89
88	332	1	1	1	1	0	255
89	897	0	0	3	1	0	886
90	118	1	1	3	1	0	93
91	515	1	0	2	1	0	425
92	335	1	1	3	1	0	306
93	145	1	0	3	1	0	122
94	843	0	0	1	1	0	764

95	528	1	0	2	1	0	333
96	823	0	1	2	1	0	819
97	753	0	1	4	1	0	681
98	545	1	1	2	1	0	459
99	759	0	0	3	1	0	754
100	757	0	1	3	1	0	678
101	110	1	0	3	1	1	85
102	764	0	1	3	1	0	753
103	427	1	1	3	1	0	286
104	690	0	1	1	1	0	658
105	683	0	1	1	1	0	637
106	499	1	0	2	1	1	372
107	395	1	0	4	1	0	371
108	556	1	0	1	1	1	400
109	336	1	0	1	1	0	244
110	30	1	0	3	1	0	0
111	347	1	1	3	1	0	238
112	596	1	0	3	1	1	212
113	638	0	1	3	1	0	629
114	249	1	0	3	1	0	146
115	619	0	1	3	1	1	337
116	605	0	0	4	1	0	568
117	250	1	1	2	1	0	0
118	153	1	1	2	1	0	0
119	366	1	1	3	1	0	291
120	563	0	0	3	1	0	552
121	212	1	1	1	1	0	0
122	477	1	1	1	1	1	246
123	568	0	1	3	1	0	482
124	316	1	0	2	1	0	174
125	270	0	0	2	1	0	212
126	434	1	1	2	1	0	153
127	533	0	0	2	1	0	530
128	213	1	0	1	1	1	128
129	134	1	0	2	1	0	0
130	89	1	0	3	1	0	86
131	468	1	0	3	1	0	417
132	393	1	1	2	1	0	240
133	473	0	1	3	1	0	429
134	472	0	1	1	1	0	426
135	465	0	0	3	1	1	168
136	237	1	1	3	1	1	139
137	442	0	0	2	1	0	393
138	421	0	1	3	1	0	385
139	395	0	1	2	1	1	359

140	403	0	1	2	1	0	394
141	388	0	0	3	1	0	321
142	388	0	0	3	1	0	350
143	372	1	0	1	1	0	316
144	248	1	0	1	1	1	94
145	279	1	1	3	1	0	216
146	56	1	1	3	1	0	0
147	366	0	1	1	1	0	355
148	85	1	1	3	1	0	62
149	347	0	0	1	1	1	331
150	344	0	0	3	1	0	319
151	346	0	0	3	1	0	321
152	319	0	0	2	1	1	184
153	165	1	0	2	1	0	122
154	311	0	1	1	1	0	232

Appendix B

Pilot Quality of Life Study

B.1 General Questionnaire

Miscellaneous

1.	Chemotherapy toxicity score	0	0
		1-3	1
		4-6	2
		7-10	3
2.	Weight loss (Kg)	None	0
		1-3	1
		4-6	2
		>6	3
3.	Visits to G.P.	Every 4 weeks or more	0
		Every 2-4 weeks	1
		Every 1-2 weeks	2
		G.P. calls	3
4.	Fitness for self-care	Self-caring	0
		Minimum assistance needed	1
		Some assistance needed	2
		Unable, needs maximum assistance	3
5.	Elimination	No problem	0
		Some constipation (diet controlled)	1
		Constipation (requires aperients)	2
		Constipation (requires enema)	3

Physical

6.	Pain	No pain	0
		Well controlled	1
		Some control	2
		Uncontrolled	3
7.	Ambulatory status	Mobile	0
		In bed/chair ≤50% of day	1
		In bed/chair >50% of day	2
		Confined to bed/chair	3
8.	Leisure	As before	0
		Goes out occasionally, lots of friends	1
		Friends visit, does not go out	2
		Does not go out, no visitors	3
9.	Sexual function	No change	0
		Decreased	1
		Impaired	2
		Impotent	3
10.	Diet	Normal	0
		Fair, solids	1
		Poor, supplements	2
		No appetite, fluids only	3

Stress

11.	Relationship with partner	Open	0
		Some talk	1
		Limited talk	2
		Unable to talk	3
12.	Family relationships	Good	0
		Coping fairly well	1
		Some friction	2
		Very strained	3
13.	Sleep at night	> 6 hours	0
		4-6 hours	1
		2-4 hours	2
		< 2 hours	3
14.	Employment	Full	0
		Moderately heavy work	1
		Light work only	2
		Unable to work	3
15.	Finances	No problem	0
		Managing (no savings)	1
		Some difficaulty (using savings)	2
		Depleted, DSS assistance	3

Appendix C

BMJ Paper

High energy neutron treatment for pelvic cancers: study stopped because of increased mortality

R D Errington, D Ashby, S M Gore, K R Abrams, S Myint, D E Bonnett, S W Blake and T E Saxton

British Medical Journal Volume 302, May 1991, pages 1045 to 1051.

Accreditation of Work

R D Errington Overall study co-ordinator, consultant in radiotherapy

D Ashby Co-ordinator of statistical analysis, morbidity analysis

S M Gore Trial design, 'trial roulette', meta analysis

K R Abrams Mortality analysis, statistical computing

S Myint Consultant in oncology

D E Bonnett Senior physicist

S W Blake Physicist

T E Saxton Cyclotron Manager

Reprinted from the BRITISH MEDICAL JOURNAL, 4th May 1991, Vol. 302, Pages 1045-1051

PAPERS

High energy neutron treatment for pelvic cancers: study stopped because of increased mortality

R D Errington, D Ashby, S M Gore, K R Abrams, S Myint, D E Bonnett, S W Blake, T E Saxton

Abstract

***Objective*—To compare high energy fast neutron treatment with conventional megavoltage x ray treatment in the management of locally advanced pelvic carcinomas (of the cervix, bladder, prostate, and rectum).**

***Design*—Randomised study from February 1986; randomisation to neutron treatment or photon treatment was unstratified and in the ratio of 3 to 1 until January 1988, when randomisation was in the ratio 1 to 1 and stratified by site of tumour.**

***Setting*—Mersey regional radiotherapy centre at Clatterbridge Hospital, Wirral.**

***Patients*—151 patients with locally advanced, non-metastatic pelvic cancer (27 cervical, 69 of the bladder, seven prostatic, and 48 of the rectum).**

***Intervention*—Randomisation to neutron treatment was stopped in February 1990.**

***Main outcome measures*—Patient survival and causes of death in relation to the development of metastatic disease and treatment related morbidity.**

***Results*—In the first phase of the trial 42 patients were randomised to neutron treatment and 14 to photon treatment, and in the second phase 48 to neutron treatment and 47 to photon treatment. The relative risk of mortality for photons compared with neutrons was 0·66 (95% confidence interval 0·40 to 1·10) after adjustment for site of tumour and other important prognostic factors. Short term and long term complications were similar in both groups.**

***Conclusions*—The trial was stopped because of the increased mortality in patients with cancer of the cervix, bladder, or rectum treated with neutrons.**

Medical Research Council Cyclotron Unit, Mersey Regional Centre for Radiotherapy and Oncology, Clatterbridge Hospital, Wirral L63 4JY
R D Errington, FRCR, *consultant in radiotherapy and oncology*
S Myint, FRCR, *consultant in radiotherapy and oncology*
D E Bonnett, PHD, *Medical Research Council senior physicist*
S W Blake, MSC, *Medical Research Council physicist*
T E Saxton, BSC, *cyclotron manager*

Departments of Statistics and Computational Mathematics and Public Health, University of Liverpool, Liverpool L69 3BX
D Ashby, PHD, *lecturer in medical statistics*
K R Abrams, MSC, *research student in medical statistics*

Medical Research Council Biostatistics Unit, Cambridge, CB2 2BW
S M Gore, PHD, *senior statistician*

Correspondence to: Dr Errington.

BMJ 1991;302:1045-51

Introduction

Compared with conventional megavoltage radiotherapy (with photons) high linear energy transfer radiation such as neutron beams has potential biological advantages that may lead to improved overall results when used to treat locally advanced tumours.[1] These advantages were not shown, however, in randomised studies on the treatment of rectal cancer and cancer of the bladder with low energy neutron beams[2-6] and carcinoma of the cervix with mixed photon and neutron treatment schedules.[7] Of the studies of mixed photon and neutron treatment only the one in patients with cancer of the prostate showed an advantage when neutrons were used as part of the radiation treatment,[8,9] but the validity of this observation has been questioned.[10]

Mixed photon and neutron treatment schedules were evolved to overcome the logistic problems posed by limited access to non-hospital based cyclotrons and were not based on any radiobiological rationale.[11] Low energy neutrons are associated with excess morbidity when used alone to treat pelvic tumours, which may obscure any benefit to be derived from neutrons if they could be used to irradiate pelvic tumours with dose distributions similar to those obtained with megavoltage x rays.[12,13]

More recently, hospital based cyclotrons capable of producing high energy neutrons have been developed, and there is registry based evidence that these are associated with greatly reduced morbidity when used to treat patients with pelvic tumours.[14] In view of this, further phase three randomised studies of treatment with high energy neutrons versus treatment with photons were initiated in patients with locally advanced pelvic tumours (of the cervix, bladder, prostate, and rectum) to define the role of neutron treatment at these sites and determine whether or not the potential biological advantages are real in terms of clinical outcomes.

The research programme at Clatterbridge Hospital had the following positive features: high energy neutron treatment was compared with modern megavoltage photon treatment; patient follow up on site and by research clinicians; randomisation from the outset, with a ratio of patients allocated to neutron treatment compared those allocated to photon treatment of 3 to 1 from 6 February 1986 until 11 January 1988, when the ratio was changed to 1 to 1 randomisation by permuted blocks of variable length and stratified by site of tumour; dual planning of eligible patients to avoid "non-evaluable" bias—patients were randomised only if the plans for both neutron and photon treatment were acceptable; a careful informed consent procedure.

Poor patient accrual was one argument for combining patients with cancer at different sites (cervix, bladder, rectum, and prostate) within a single randomised trial; site specific trials would have been entirely lacking in statistical power. A second reason was that although the specific tissues affected by morbidity due to radiation might differ among patients with cancer at different sites, the pooled data would give a clearer indication of whether morbidity was becoming the serious problem it had been shown to be when low energy neutrons were used. Collaboration with the American College of Radiology Radiation Therapy Oncology Group was not possible as in March 1988 poor patient recruitment led to their trials in patients with cervical and rectal cancer being abandoned. The group's trial of high energy neutrons versus photons in patients with cancer of the prostate, was closed in October 1990 with 178 patients entered.

In preparation for a mid-term review of the Clatterbridge cyclotron's research programme in December 1989 by a Medical Research Council subcommittee an ad hoc analysis of mortality and morbidity results was prepared. Randomisation in the council's trial of patients with pelvic tumours to high energy neutron treatment or photon treatment was suspended on 12 February 1990. The decision was ratified by the cancer therapy committee of the council on 8 March 1990. In

this paper we describe the sequence of events leading to that decision and the results on which it was based.

Methods

ELIGIBILITY

To be eligible for randomisation patients had to have histologically confirmed adenocarcinoma of the rectum or prostate, squamous cell carcinoma of the cervix, or transitional cell carcinoma of the bladder not previously treated by radiotherapy or chemotherapy. Patients who were 80 years old or over had a Karnofsky performance score of 40 or less, or were otherwise unfit for radical pelvic radiotherapy were not eligible; nor were those with a history of malignancy at another site or evidence of distant metastases. Patients entering the study all had locally advanced disease as determined by appropriate clinical (including examination under anaesthesia, cystoscopy, and sigmoidoscopy) and radiological investigations (including computed tomography). By TNM staging randomised patients therefore had T_{3a}, T_{3b}, or T_4 and N_0, N_1 or N_2, N_x stage cancer in all sites studied.

RADIATION TECHNIQUES AND DOSAGE

Treatment in all patients was planned with a treatment simulator that incorporated information from diagnostic and planning computed tomograms. In all patients an initial large volume of the pelvis was treated to cover the primary tumour and pelvic lymph nodes. This was followed by a second phase, entailing treatment of a small volume of the pelvis, replanned on the basis of computed tomograms, to treat the primary tumour site with a 2 cm margin around the area of known macroscopic disease. In patients with cervical cancer intracavity treatment was given whenever possible after the first phase of pelvic radiotherapy and was followed, where appropriate, with a boost treatment to a small volume of the parametrium. Table I gives the dose schedules for each site.

For both neutron and photon treatment comparable dose distributions were achieved as confirmed by dual planning in the early phase of the study (see below). Three and four field techniques were used and plans accepted only when the variation of dose within the target volume did not deviate by more than 7·5% of the target absorbed dose (as specified in section 33 of International Commission on Radiation Units and Measurements report No 29). Isocentric treatment techniques were used with 8 MV x rays (source to axis distance 100 cm) from a linear accelerator or with a neutron beam generated by bombarding a beryllium target with 62 MeV protons.[15]

RANDOMISATION AND DUAL PLANNING

During the initial phase (phase 1: from 6 February 1986 to 10 January 1988) 56 patients were randomised in the ratio of 3 to 1—42 to neutron treatment and 14 to photon treatment—but without stratification by site of tumour. Block length (eight) was not disclosed to clinicians, and randomisation was performed at the hospital by using sealed envelopes prepared by one of us (SMG).

Dual planning (for neutron and photon treatment) for eligible patients was practised for the first 50 patients; all were in fact randomised because they had adequate radiotherapy plans for both modalities. Dual planning was then suspended. Analysis by intention to treat was adhered to strictly with no patient excluded from the analysis retrospectively on grounds of inadequacy of radiotherapy plan, thus avoiding non-evaluable bias.

From 11 January 1988, 1 to 1 randomisation (again by using sealed envelopes) by permuted blocks of length four or six (determined by simple randomisation) and stratified by site of tumour was adopted for patients with tumours of the cervix, bladder, rectum, or prostate.

PATIENT FOLLOW UP

During treatment all patients were seen weekly to record reactions to treatment. During the first year after treatment patients were seen monthly and computed tomography, and cystoscopy in those with bladder tumours, repeated every three months. For subsequent years patients were seen at intervals of two to three months. At each follow up visit tumour response was assessed; any morbidity due to radiation was scored by the European Organisation for Research on Treatment of Cancer/Radiation Therapy Oncology Group scoring criteria.

OUTCOMES

For the interim analysis the primary outcome was mortality from all causes. The secondary end point was severe toxicity caused by treatment as defined by the recurrence of a reaction scoring grade 3 or higher by the European Organisation for Research on Treatment of Cancer/Radiation Therapy Oncology Group scoring criteria.

TRIAL SIZE

During the first phase of unstratified randomisation accrual was 29 patients per annum. Referrals for cancers of the bladder and cervix doubled in the second phase of randomisation, when accrual increased to 46 patients per annum. Even so, a minimum target (to give a 50% power to detect a relative risk of 1·30; see below) of 300 randomised patients was not likely to be reached until 1993; the Medical Research Council mid-term review of the programme was scheduled for December 1989. In the absence of a formal plan for interim analysis ad hoc analysis (of 134 patients randomised up to 12 September 1989) was undertaken before the council's visit to the hospital and intended mainly as a check on data quality.

Trial size was assessed in 1988 soon after the start of 1 to 1 randomisation stratified by site of tumour. Five randomised trials of treatment with low energy neutrons for carcinoma of the bladder or rectum had been published by the end of 1987.[2,4,5,6] In March 1988 a non-random sample of 10 clinicians and physicists

TABLE I—*Dose schedules with neutrons and photons in patients treated for pelvic cancer*

Type of radiation	Radiation to large pelvic volume (phase 1)	Radiation to small pelvic volume (phase 2)		Total dose to tumour
Patients with cancer of the bladder, rectum, or prostate:				
Neutrons	14·4 Gy (9 fractions over 21 days: 3 fractions/week)		4·8 Gy (3 fractions over 7 days)	19·2 Gy (12 fractions over 28 days)
Photons	44 Gy (22 fractions over 30 days: 5 fractions/week)		20 Gy (10 fractions over 14 days)	64 Gy (32 fractions over 44 days)
Patients with cancer of the cervix:				
Neutrons	14·4 Gy (9 fractions over 21 days)	Intracavity caesium (by a selectron) 16 Gy (point A) if possible or	4·8 Gy (3 fractions over 7 days)	
Photons	50 Gy (25 fractions over 35 days)		16 Gy (8 fractions over 10 days)	

TABLE II—*No of patients randomised to treatment with neutrons or photons by phase of randomisation (1 or 2) and site of tumour, with (actuarial) death rates at one year*

Site of tumour	Phase 1		Phase 2		Total (death rate at one year)	
	Neutrons	Photons	Neutrons	Photons	Neutrons	Photons
Cervix	3	6	9	9	12 (32%)	15 (22%)
Rectum	19	4	12	13	31 (52%)	17 (44%)
Bladder	17	4	24	24	41 (71%)	28 (44%)
Prostate	3	0	3	1	6 (17%)	1 (0%)
Total (death rate at one year)	42 (60%)	14 (32%)	48 (48%)	47 (41%)	90 (55%)	61 (38%)

with an interest or involvement in neutron treatment were asked to quantify their current belief about the failure rate of treatment with high energy neutrons in patients with pelvic cancer. The consensus belief was a modest advantage with neutron treatment (median relative risk was 1·14 for failure of photon treatment compared with high energy neutron treatment). Respondents were also asked what relative risk they would accept for high energy neutrons to be recommended routinely for treating cancer of the pelvis. The consensus was a relative risk of 50/38·5=1·30—that is, a 30% greater failure rate with photons than with high energy neutrons. Randomisation of 600 patients was indicated for 80% power to detect such a moderate difference in failure rates as 50% *v* 38·5% and randomisation of 300 patients for 50% power.

Respondents' belief about the failure rate of treatment with high energy neutrons also established that randomisation of patients was ethical: in March 1988 respondents put 26% of their belief that neutron treatment had a failure rate of 38·5% or less, but 28% of their belief that there were as many or more failures with high energy neutrons as with photons (a reference failure rate of 50%). Disparity between repondents' belief in treatment with high energy neutrons and the posterior distribution derived from subsequent statistical overview (see appendix) of the five randomised trials of low energy neutrons in treating cancer of the bladder and rectum[2-6] was considerable.

DATA MONITORING COMMITTEE

Acting on the interim analysis reports (analysis date 12 September 1989) prepared for the Medical Research Council subcommittee's site visit on 4 December 1989, the neutron subgroup met in January 1990 to suggest establishing an independent data monitoring committee to advise on stopping randomisation in the pelvic cancer trial. The data monitoring committee was presented with formal analysis of mortality and morbidity for all patients randomised up to 26 January 1990; interim analysis reports up to 12 September 1989; statistical overviews of published randomised trials of treatment with low energy neutrons versus photons for pelvic cancer and for cancers of the head and neck; and a summary of beliefs (in March 1988) about the failure rate of high energy neutron treatment for pelvic tumours.

Randomisation was suspended in February 1990 by one of us (RDE); this decision was ratified by the data monitoring committee and approved by the cancer therapy committee on 8 March 1990.

STATISTICAL ANALYSIS

Comparison of mortality by using life tables was used throughout, which makes proper allowance for differential follow up times induced by initial 3 to 1 randomisation. Randomisation was not stratified in the initial phase and was stratified only by site of tumour in the second phase, when the randomisation ratio was 1 to 1, and so retrospective covariate adjustment was performed by Cox's proportional hazards model.[16,17] Covariate adjustment allowed us to check that imbalances (especially in the first phase, when randomisation was not stratified) between patients randomised to treatment with neutrons and those randomised to treatment with photons did not cause bias in the estimation of the relative risks of death for the two treatments. Covariates were firstly design variables (site of tumour and phase of randomisation) and secondly those variables specified in advance by one of us (RDE) as being important prognostic indicators (T stage, N stage, and Karnofsky index). Although T_{4a} stage has a different definition for tumours of the bladder and rectum, our anlaysis makes the reasonable assumption of similarity of relative risk (for example, for T_{3b} *v* T_{3a}) across all sites of tumour. Observations on mortality were censored on 26 January 1990, when the vital status of every patient was ascertained. For analyses of morbidity due to radiation observations were censored at the last recorded follow up visit. All analyses were by intention to treat. Protocol violations occurred in 11 patients (six treated with neutrons and five treated with protons) of 151 patients—namely, two patients with cervical cancer (incorrect histology), six with rectal cancer (five failed to receive protocol treatment and one with incorrect results on histological examination), and three with cancer of the bladder (two failed to receive protocol treatment and one with incorrect staging).

Results

By 26 January 1990, 151 patients had been randomised. Table II shows how they were distributed by site of tumour, phase of randomisation, and treatment. Accrual was 70% higher in the second phase compared with the first, with recruitment doubled for patients with cancer of the cervix or bladder. Figure 1 shows the estimated survivor function by treatment for cancers at all sites combined. Table II gives the (actuarial) death rate at one year by treatment group and phase of randomisation, and also by site of tumour and treatment group. Only one patient with cancer of the prostate was randomised to treatment with photons. The relative risk of mortality for treatment with photons compared with treatment with neutrons was estimated by using Cox's proportional hazards model. Comparing the two treatment regimens without adjusting for any covariates yielded a relative risk of 0·59 ($p<0·025$)—that is, patients treated with photons had an estimated time specific risk of dying 0·59 times that of patients treated with neutrons (95% confidence interval 0·36 to 0·95).

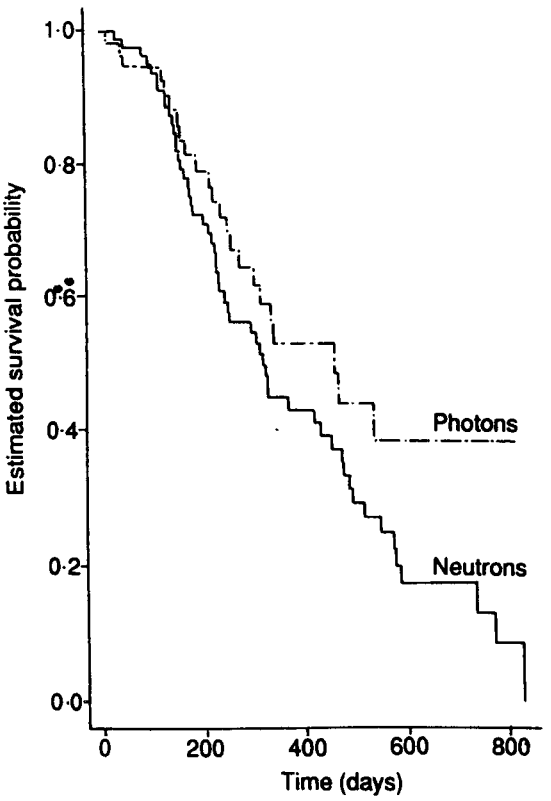
Allowance for site of tumour changes the relative risk to 0·62, and allowance for phase of randomisation changes it to 0·65. However, once the other important prognostic factors (T stage, N stage, and Karnofsky index) are included, the effect of phase of treatment is negligible ($p=0·88$), indicating that differences in survival by phase in table II are due to small imbalances in the randomised groups (table III) rather than to real changes in survival. The relative risk adjusted for site of tumour and other important prognostic factors was 0·66 (0·40 to 1·10).

Table IV shows important prognostic characteristics by site of tumour and treatment. Using Cox's proportional hazards model a combined risk score predicting survival can be calculated for each patient. Table V shows the risk score summation for site of tumour, T and N stage, and Karnofsky index. Table VI shows the distribution of this risk score for both treatment groups. On average the patients treated with neutrons had a worse prognosis than those treated with photons in the first phase, when randomisation was not stratified by site of tumour, and patients with cancers of the rectum and bladder were overrepresented in the neutron group. This explains why the adjusted relative risk is slightly nearer to unity than the unadjusted relative risk.

Table VII summarises cause of death according to treatment and subdivided into deaths within a year of treatment and later deaths. The distribution by cause of all deaths was similar for the two treatment groups ($\chi^2=4{\cdot}74$, df=3); a relative excess of patients with rectal cancer had metastases (with or without local progression) compared with patients with cancer of the bladder ($\chi^2=7{\cdot}93$, df=3).

Table VIII summarises the morbidity due to radiation by site of tumour for early and late severe reactions (Radiation Therapy Oncology Group/ European Organisation for Research on Treatment of Cancer score 3, 4, or 5). There was very little difference between the treatments for early reactions and none for late reactions: by Cox's proportional hazards model (unadjusted for covariates) the relative risk of early severe reactions for photon treatment compared with neutron treatment was 0·77 (95% confidence interval 0·29 to 2·04) and of late severe reactions 0·85 (0·39 to 1·89).

Patients with cancer of the prostate showed no early or late severe reactions, and for the others only three sites were involved for early reactions: the upper gastrointestinal tract, the lower gastrointestinal tract, and the bladder. For severe late reactions the sites most frequently involved were lower gastrointestinal tract and bladder.



Estimated probability of survival in patients with pelvic cancer randomised to treatment with neutrons or photons (analysis of results until 26 January 1991)

TABLE III—*Mean (SD) risk score [number of patients] by treatment and phase of randomisation*

Phase of treatment	Neutrons	Photons	Total
1*	−1·79 (0·59) [42]	−2·18 (0·53) [14]	−1·89 (0·60) [56]
2	−2·15 (0·63) [48]	−2·06 (0·58) [47]	−2·10 (0·60) [95]
1 and 2	−1·98 (0·64) [90]	−2·09 (0·57) [61]	−2·02 (0·61) [151]

*Randomisation was not stratified by site of tumour—a major prognostic factor. Patients randomised to neutron treatment had a significantly worse prognosis than those randomised to photon treatment (z=2·32, p=0·02), which can be adjusted for by Cox regression analysis.

TABLE VI—*Number (percentage) of patients in each category of risk score for treatment with photons v treatment with neutrons*

Category of risk score	Neutrons	Photons
−3·99 - −3·50	1 (1)	0 (0)
−3·49 - −3·00	6 (7)	5 (8)
−2·99 - −2·50	10 (11)	5 (8)
−2·49 - −2·00	29 (32)	20 (33)
−1·99 - −1·50	25 (28)	24 (39)
−1·49 - −1·00	16 (18)	6 (10)xp%
−0·99 - −0·50	2 (2)	1 (2)
−0·49 - −0·00	1 (1)	0 (0)
Total	90 (100%)	61 (100%)

TABLE IV—*Major prognostic factors by site of tumour and treatment (neutrons or photons) in patients with pelvic cancer. Figures are numbers (percentages) of patients*

	Cervix		Bladder		Rectum		
	Neutrons	Photons	Neutrons	Photons	Neutrons	Photons	Prostate
T stage:							
3a	2 (17)		9 (22)	8 (29)	4 (13)		1
3b	6 (50)	13 (88)	23 (56)	14 (50)	1 (3)		
4	4 (33)	2 (12)	9 (22)	6 (21)	26 (84)	17 (100)	6*
N stage:							
0	11 (92)	10 (67)	34 (83)	26 (93)	27 (87)	16 (94)	6*
1	1 (8)	5 (33)	5 (12)	2 (7)	3 (10)	1 (6)	
2 or missing			2† (4)		1 (3)		1
Karnofsky index:							
50			1 (2)	1 (4)	2 (6)	1 (6)	
60	1 (8)	2 (14)	9 (22)	3 (11)	6 (19)	1 (6)	
70	4 (33)	1 (6)	5 (12)	2 (7)	7 (23)	12 (71)	1*
80	2 (17)	5 (33)	19 (46)	14 (50)	13 (42)	3 (17)	4
90	4 (33)	5 (33)	4 (10)	8 (29)	2 (6)	0	2
100	1 (8)	2 (14)	3 (7)				

*Prognostic factors of the one patient with cancer of the prostate randomised to photon treatment.
†N stage was missing for one of these patients.

TABLE V—*Risk score summation for site of tumour, stage, and Karnofsky index*

Site of tumour	Risk score	T stage	Risk score	N stage	Risk score	Karnofsky index Risk score
Cervix	−1·12	3a	0	0	0	
Bladder	0	3b	0·27	1	0·73	−0·026 × Karnofsky
Rectum	0·04	4a	0·09	2	0·83	
Prostate	−0·81	4b	0·01			

Example: risk score for patient with cancer of the bladder, stage T_{3B}, N_0 and Karnofsky index of 80=0+0·27+0−0·026=−1·81.

Discussion

When designing a clinical trial to compare two treatments the first priority is unbiased comparison, and the second is a powerful comparison. The trial of neutron treatment versus photon treatment achieved the first of these, but given the anticipated effects, combined with patient accrual, the second was always likely to be a problem.

Clinical opinion was elicited formally and indicated a median expectation of results for neutron treatment being favourable compared with photon treatment, but there was sufficient uncertainty to justify randomisation. Nevertheless, the conflict between clinical opinion and the results of the studies on treatment with low energy neutrons (see appendix) should have meant that formal stopping rules were incorporated to guard against the possibility of adverse mortality or morbidity.

TABLE VII—*Cause of death by treatment (neutrons or photons) and survival time, and by site of tumour for patients with cancer of the bladder or rectum*

Cause of death	Deaths within one year after randomisation		Deaths after one year		Total No of deaths		Patients with cancer of the bladder	Patients with cancer of the rectum
	Neutrons	Photons	Neutrons	Photons	Neutrons	Photons		
Metastases	5			1	5	1	5	1
Metastases and local tumour progression	21	5	7	2	28	7	12	20
Local progression	13	12	4	1	17	13	16	8
Morbidity due to treatment			3	1	3	1	2	1
New primary cancer			2		2			
Intercurrent disease	2	2			2	2	3	1
Total	41	19	16	5	57	24	38	31

TABLE VIII—*Early (within 90 days) and late morbidity due to radiation by Radiation Therapy Oncology Group/European Organisation for Research on Treatment of Cancer scoring by (neutrons or photons) and site of tumour. Figures are numbers of patients unless stated otherwise*

	Cervix		Bladder		Rectum			All sites	
	Neutrons	Photons	Neutrons	Photons	Neutrons	Photons	Prostate*	Neutrons	Photons
Early morbidity:									
No randomised	12	15	41	28	31	17	7	90	61
No with ≥1 severe early reaction	1	1	9	2	2	3	0	12	6
Actuarial % with ≥1 severe reaction within three months after randomisation	11	8	27	9	7	23	0	16	12
Late morbidity:									
No of survivors at 90 days	9	14	35	23	30	14	7	80	52
No with ≥1 severe late reaction	2	3	8	4	6	3	0	16	10
Actuarial % with ≥1 severe reaction within one year after randomisation	21	15	40	24	16	27	0	24	21

*Only one of seven patients was randomised to photon treatment.

Because the trial was randomised the main results were presented for the comparison between neutron treatment and photon treatment, unadjusted for any other factor. However, because the trial was randomised in two phases and only in the second phase was stratified by site of tumour the relative risk of photon treatment compared with neutron treatment was re-estimated taking these and other important covariates into account. This modified the relative risk but did not substantially alter the conclusions of the unadjusted analysis. Covariate adjustment for other important prognostic factors had a similar effect. The 95% confidence interval extends to 1·10 but clearly excludes a relative risk of 1·30, at which neutrons would be recommended for treatment, and is consistent with a relative risk of 0·65 from the statistical overview of randomised trials of low energy neutron treatment for tumours of the rectum and bladder.

Another consideration is whether the estimate of the relative risk should be adjusted for early stopping of the trial. This was not done formally in this study because the informal analysis was undertaken for a scheduled visit to the hospital and was not motivated by any knowledge of the results. Pocock and Hughes have shown that clinical trials that stop early are prone to exaggerate the difference between treatments.[18] Continuing the trial just to obtain unbiased estimates of the treatment effect would, however, have been unethical given that the test treatment (neutrons) was proving inferior.

Because of the trial's design—we studied patients with cancer at four separate sites but with similar protocols and initial joint randomisation—it is debatable whether the data should be analysed as one trial or four. Had the trials continued to completion interest would undoubtedly have been focused on the relative risk or benefits of neutrons at each site. However, the power to look at each site separately is weak; indeed, it had been questioned whether there were enough data for a combined analysis. Because of the magnitude of effect, combined with internal consistency across patients with tumours at different sites and external consistency with results for treatment with low energy neutrons, the data monitoring committee concluded that there was sufficient evidence of adverse mortality in patients treated with neutrons to stop the trials, although there were insufficient data to quantify the excess risk with any accuracy.

Another issue is clinical significance versus statistical significance. In this trial the difference in mortality was large, but the primary difference between the results of the first informal analysis and those presented formally for the data monitoring committee was that the statistical significance of the relative risk for mortality changed from $p=0·07$ to $p=0·025$, which gave stronger evidence against the null hypothesis of no treatment difference. The important yardstick, however, is not equality of effect but the difference that would be required to change clinical practice. This was estimated in March 1988 to be a 30% greater failure rate for photon treatment by the respondents to the inquiry made by one of us (SMG). This difference was well outside the 95% confidence interval, even at the first informal analysis, and it could be argued that the trials should have been suspended on those grounds alone.

There was an intrinsic asymmetry between the two treatments, with one being an established treatment readily available to patients not in the trial and the other being an experimental treatment available at only one site in the United Kingdom. Had the difference been in the reverse direction, in favour of neutron treatment, the trial would be continuing in order to get a better estimate of the superiority of the new treatment. Ethically this would have been justified by considerations of limited resources—that is, that neutron treatment could not currently be offered to all patients. This is in contrast with drugs trials, where it is often feasible quickly to make a new treatment widely available. However, in this trial, with the new treatment showing an adverse effect, to have continued with the trial to gain a better estimate of the difference between neutron and photon treatments would have

entailed continuing to randomise some patients to a treatment with an almost certainly worse prognosis than the treatment they would have received if they did not enter the trial. Randomisation was suspended by one of us (RDE) in February 1990 because it was thought to be unethical to continue to randomise patients. Monitoring of patients continues, and to that extent more information on neutron treatment versus photon treatment will become available, especially as regards morbidity in the survivors.

A lesson learnt from this trial is that all trials of this kind, with mortality as an outcome and potentially causing severe toxicity, should have a data monitoring committee in place from the start of the trial. Its role is to decide how and when to monitor results and to share responsibility for decisions to stop or continue trials. This avoids the difficult situation where individual clinicians are forced to make these decisions with little external support or guidance.

In the low energy neutron studies in patients with cancer of the bladder the increased mortality associated with neutrons was related to increased morbidity compared with that caused by photons rather than differences in local tumour control or metastatic relapse.[2,3] This contrasts with the present study, in which the morbidity data showed no significant difference between neutron treatment and photon treatment. This observation, however, needs to be interpreted with caution as follow up times were short and the poor survival of patients treated with neutrons depleted the number available for the full assessment of more serious late complications. Despite this reservation the assumption that high energy neutrons, with improved physical dosimetry, would be associated with less normal tissue morbidity may be correct. This does not, however, lead to any benefit from high energy neutron treatment in terms of survival or freedom from metastatic disease.

We emphasise that these conclusions are not applicable at this stage to patients with locally advanced prostatic adenocarcinoma in view of the small number of such patients recruited to this study, only one of whom was randomised to treatment with photons. There is evidence supporting the use of mixed beam therapy in patients with carcinoma of the prostate,[4] and accrual to a phase three Radiation Therapy Oncology Group study of treatment with neutrons alone versus treatment with photons was closed in October 1990 with 178 patients entered, though it will be several years before the results of this trial are available. In conclusion, the results of this study do not support the continued use of the stated schedules (table I) of high energy neutrons in the treatment of locally advanced carcinomas of the cervix, bladder, or rectum.

We thank Sister Rose Dixon, cyclotron nurse; Mrs J Neale, cyclotron superintendent radiographer; Miss S Minnis, trials secretary; and Mrs E Batty for their contributions. We also thank Professor G R H Sealy, clinical director of the Mersey regional centre for radiotherapy and oncology for advice; the Medical Research Council data monitoring committee (chairman Mr R R Hall with Dr J A Lewis, chief statistician; the medical affairs department, Imperial Chemical Industries; and Dr T J Priestman, consultant in radiotherapy and oncology, Wolverhampton); the neutron therapy steering committee (chairman Professor C A F Joslin); and Dr F Spencer (Medical Research Council head office). The Clatterbridge cyclotron facility was established as a jointly funded initiative between the Medical Research Council, Imperial Cancer Research Fund, Cancer Research Campaign, Clatterbridge Cancer Research Trust, and the Cancer and Polio Research Fund in collaboration with Mersey Regional and Wirral Health Authorities and the University of Liverpool. Professor H M Warenius was Medical Research

APPENDIX

Statistical overview of randomised trials of low energy neutron treatment versus photon treatment for cancer of the bladder or rectum

(1) Batterman[1]: 34 patients with advanced tumours of the bladder or rectum were randomised to treatment with photons, 31 to treatment with neutrons at 17 Gy, and 26 to treatment with neutrons at 19 Gy. Results were given as an artist's impression of Kaplan-Meier survival curves, from which we read one year survival rates as 39% for neutrons at 17 Gy, 44% for neutrons at 19 Gy, and 47% for photons. Analysis was without exclusion of patients after randomisation.

(2) Pointon *et al*[5]: 53 patients with stage T_2 or T_3 cancer of the bladder were randomised (by telephone) to treatment with photons, 28 to low dose neutrons, and 27 to high dose neutrons. Analysis was without exclusion of patients after randomisation; 16 patients did not, however, complete the treatment to which they were randomised. From life tables, shown as a step function for all patients randomised to neutron treatment we read one year survival rates as 70% for neutron treatment and 80% for photon treatment.

(3) Duncan *et al*[6]: 60 patients with transitional cell carcinoma of the bladder stratified by T stage and histological grade into four groups were randomised (by using sealed envelopes) to treatment with photons and 53 to treatment with neutrons. Treatment had to be started within 14 days of randomisation; no patient was excluded from analysis. From life tables we read one year survival rates as 50% for the neutron group and 72% for the photon group.

(4,5, and 6) Duncan *et al*[6]: 10 out of 77 patients were excluded after randomisation (patients were stratified, and sealed envelopes were used), of whom four had drawn photon treatment and six neutron treatment. Of the excluded patients, one out of four had survived photon treatment at one year after randomisation and two out of six had survived neutron treatment (G R Kerr, personal communication). Actuarial survival rates at one year were tabulated separately for patients analysed in the trial of patients with inoperable cancer and for patients analysed in the trial of patients with recurrent cancer. One year survival rates were 15% (neutron group), 62% (photon group) and 33% (neutron group), 69% (photon group) respectively.

Patients with inoperable cancer were stratified according to whether the inoperability was due to the extent of the tumour (15 were assigned to photon treatment, 16 to neutron treatment) or age or general condition (one in the photon group, four in the neutron group). In the trial of patients with recurrent cancer 16 of those randomised to photon treatment and 15 of those randomised to neutron treatment were analysed. Both trials closed in May 1984 because it was feared that similar radiation morbidity or mortality, or both, might be experienced by patients in these trials as had been seen in a concurrent trial of neutron treatment for bladder cancer (trial 6).

Trials 3, 4, and 5 were terminated because of concerns about radiation morbidity or mortality, or both, in patients randomised to low energy neutron treatment. A figure of 99 survivors of low energy neutron treatment at one year after randomisation is more than 3·3 standard errors fewer than expected (that is, 115·3) if death rates were identical with photon treatment and low energy neutron treatment. The pooled relative risk (95% confidence interval) of death within one year after randomisation (photon *v* low energy neutrons) was 0·65 (0·50 to 0·84).[20]

Mantel-Haenszel[19] overview of six trials of neutron v photon treatment

Trial	Site of tumour		Outcome at one year: Alive	Outcome at one year: Dead	Survivors of neutron treatment at one year: Observed	Expected	Variance
Batterman[1]	Bladder, rectum	Neutrons	23	34	23	24·429	5·273
		Photons	16	18			
Pointon[5]	Bladder	Neutrons	39	16	39	41·250	5·108
		Photons	42	11			
Duncan[6]	Bladder	Neutrons	27	26	27	32·832	6·693
		Photons	43	17			
Duncan[6]	(Inoperable) Rectum	Neutrons	3	17	3	7·222	2·109
		Photons	10	6			
Duncan[6]	(Recurrent) Rectum	Neutrons	5	10	5	7·742	1·998
		Photons	11	5			
Duncan 1987	Bladder (unpublished)	Neutrons	2	4	2	1·800	0·560
		Photons	1	3			
Total					99	115·275	21·741

$\chi^2 = \frac{[(115{\cdot}275 - 99) - 0{\cdot}5]^2}{21{\cdot}741} = 11{\cdot}45$ ($z = 3{\cdot}38$).

Odds ratio $= \frac{14{\cdot}490}{30{\cdot}765} = 0{\cdot}47$ (95% confidence interval 0·30 to 0·73).

Derived relative risk[20] = 0·65 (0·50 to 0·84).

Council honorary clinical coordinator for part of the study period. The Clatterbridge neutron therapy research programme is managed and funded by the Medical Research Council. K R Abrams is supported by a grant from the North West Cancer Research Fund. We also thank the following clinical colleagues for their support: Mr D A Berstock, Dr A J Slater, Dr J E S Brock, Dr M A Coe, Dr B Cottier, and Dr B Jones.

1 Wambersie A, Bewley DK, Lalanne CM. Prospects for the application of fast neutrons in cancer therapy. *Bull Cancer (Paris)* 1986;73:546-61.
2 Duncan W, Arnott SJ, Jack WJL, *et al*. A report of a randomized trial of d(15)+Be neutrons compared with megavoltage x-ray therapy of bladder cancer. *Int J Radiat Oncol Biol Phys* 1985;11:2043-9.
3 Duncan W, Williams JR, Kerr GR, *et al*. An analysis of the radiation related morbidity observed in a randomized trial of neutron therapy for bladder cancer. *Int J Radiat Oncol Biol Phys* 1986;12:2085-92.
4 Duncan W, Arnott SJ, Jack WJL, Orr JA, Kerr GR, Williams JR. Results of two randomised clinical trials of neutron therapy in rectal adenocarcinoma. *Radiother Oncol* 1987;8:191-8.
5 Batterman JJ. Results of d+T fast neutron irradiation on advanced tumors of bladder and rectum. *Int J Radiat Oncol Biol Phys* 1982;8:2159-64.
6 Pointon RS, Read G, Green D. A randomised comparison of photons and 15 MeV neutrons for the treatment of carcinoma of the bladder. *Br J Radiol* 1985;58:219-24.
7 Maor MH, Gillespie BW, Peters LJ, *et al*. Neutron therapy in cervical cancer: results of a phase III RTOG study. *Int J Radiat Oncol Biol Phys* 1988;14:885-91.
8 Laramore GE, Krall JM, Thomas FJ, *et al*. Fast neutron radiotherapy for locally advanced prostate cancer: results of an RTOG randomized study. *Int J Radiat Oncol Biol Phys* 1985;11:1621-7.
9 Laramore GE. The cyclotron saga. *BMJ* 1990;301:344.
10 MacDougall RH, Arnott SJ. The cyclotron saga. *BMJ* 1990;300:1721.
11 Peters LJ, Maor H, Laramore GE, *et al*. Review of clinical results of fast neutron therapy in the USA. In: Maruyama Y, Beach JL, Feola JM, eds. Californium-252 brachytherapy and fast neutron beam therapy. *Nuclear Science Applications Section B* 1986;2:243-60.
12 Griffin TW, Wambersie A, Laramore GE, Castro J. High LET: heavy particle trials. *Int J Radiat Oncol Biol Phys* 1988;14:583-92.
13 Zink S, Antoine J, Mahoney FJ. Fast neutron therapy clinical trials in the United States. *Am J Clin Oncol* 1989;12:277-82.
14 Griffin T, Pajak T, Laramore G, Davis L. Analysis of neutron radiotherapy treatment complications. *Bull Cancer (Paris)* 1986;73:582-6.
15 Bonnett DE, Blake SW, Shaw JE, Bewley DK. The Clatterbridge high-energy neutron therapy facility specification and performance. *Br J Radiol* 1988;61:38-46.
16 Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 1972;34:187-220.
17 Elashoff JD. Surviving proportional hazards. *Hepatology* 1983;73:1031-5.
18 Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Controlled Clinical Trials* 1989;10: 209-21S.
19 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell Scientific Publications, 1987:460.
20 Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley, 1981:68-75.

(Accepted 1 March 1991)

References

Aalen, O., (1978). Non-parametric inference for a family of counting processes. *Annals of Statistics*, 6,701–726.

Aaronson, N. K., Bullinger, M., and Ahmedzai, S., (1988). A modular approach to quality of life assessment in cancer clinical trials. *Recent Results in Cancer Research*, 111,231–249.

Abrams, K. R., Ashby, D., and Errington, R. D., (Submitted). Bayesian approach to neutron therapy trial. *Controlled Clinical Trials.*

Abrams, K. R., (1989). Bayesian survival analysis–a review. Master's thesis, University of Leicester.

Abrams, K. R., (1992). In discussion of Cox *et al*: 'Quality-of-life: can we keep it simple?'. *Journal of the Royal Statistical Society (A)*, 155,353–393.

Achcar, J. A. and Smith, A. F. M., (1990). *Aspects of reparameterization in approximate Bayesian inference.* North Holland, Amsterdam. In Essays in honour of G.A. Barnard.

Achcar, J. A., Brookmeyer, R., and Hunter, W. G., (1985). An application of Bayesian analysis to medical follow-up data. *Statistics in Medicine*, 4,509–520.

Achcar, J. A., Bolfarine, H., and Pericchi, L. R., (1987). Transformation of survival data to an extreme value distribution. *The Statistician*, 36,229–234.

Aitchison, J. and Bacon-Shone, J., (1981). Bayesian risk ratio analysis. *The American Statistician*, 35,4,254–257.

Aitkin, M. and Clayton, D. G., (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics*, 29,156–163.

Altham, P. M. E., (1969). Exact Bayesian analysis of a 2×2 contingency table and Fisher's 'exact' significance test. *Journal of The Royal Statistical Society (B)*, 31,261–269.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N., (1992). *Statistical models based on counting processes.* Springer, New York.

Andersen, P. K., (1986). Time-dependent covariates and Markov processes. In Moolgavkar, S. and Prentice, R. L., editors, *Modern Statistical Methods in Chronic Disease Epidemiology*, pages 82–103. Wiley, New York.

Andersen, P. K., (1988). Multi-state models in survival analysis : A study of nephropathy and mortality in diabetes. *Statistics in Medicine*, 7,661–670.

Anderson, P. K. and Borgan, O., (1985). Counting process models for life history data–a review (with discussion). *Scandinavian Journal of Statistics*, 12,97–158.

Anderson, P. K. and Gill, R. D., (1982). Cox's regression model for counting processes–a large sample study. *Annals of Statistics*, 10,1100–1120.

Armitage, P. and Berry, G., (1987). *Statistical Methods in Medical Research.* Blackwell Scientific Publications, second edition.

Aven, T., (1986). Bayesian inference in a parametric counting process. *Scandinavian Journal of Statistics*, 13,87–97.

Baker, R. J. and Nelder, J. A., (1978). *GLIM Release 3.* Oxford University. Numerical Algorithms Group.

Batterman, J. J., (1982). Results of d+T fast neutron irradiation on advanced tumours of bladder and rectum. *Int J Rad Oncol Biol Phys*, 8,2159–2164.

Bayes, T., (1783). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of The Royal Society*, 53,370–418.

Beck, G. J., (1979). Stochastic survival models with competing risks and covariates. *Biometrics*, 35,427–438.

Berger, J. O., (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, second edition.

Berry, D. A., Wolff, M. C., and Sack, D., (1992). Public health decision making : a sequential vaccine trial. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 4*, Oxford. Oxford University Press.

Besag, J. and Green, P. J., (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society (B)*, 55. To appear.

Bhattacharya, S. K., (1967). Bayesian approach to life-testing and reliability estimation. *Journal of the American Statistical Association*, 62,48–62.

Borgan, O., (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, 11,1–16.

Bowling, A., (1991). *Measuring health: A Review of Quality of Life Measurement Scales.* Open University Press, Milton Keynes.

Box, G. E. P. and Tiao, G. C., (1967). Bayesian estimation of means for the random effects model. pages 174–181.

Box, G. E. P. and Tiao, G. C., (1973). *Bayesian inference in statistical analysis.* Addison-Wesley.

Breslow, N. E., (1972). Discussion of regression model and life table. *Journal of The Royal Statistical Society (B)*, 34,216–217.

Breslow, N. E., (1974). Covariance analysis of censored survival data. *Biometrics*, 30,89–99.

Bulpitt, C. J., (1988). Meta-analysis. *The Lancet*, ii,93–94.

Burridge, J., (1980). Empirical Bayes estimates of survival time data. *Journal of the Royal Statistical Society (B)*, 43,65–75.

Canavos, G. C. and Tsokos, C. P., (1973). Bayesian estimation of life parameters in the Weibull distribution. *Operations Research*, 21,3,755–763.

Carlin, B., Chaloner, K., Church, T., Louis, T. A., and Matts, J. P., (1992). Bayesian approaches for monitoring clinical trials with an application to Toxoplasmic encephalitis prophylaxis. *Technical Report.* University of Minnesota.

Carlin, J. B., (1992). Meta-analysis for 2×2 tables : a Bayesian approach. *Statistics in Medicine*, 11,141–158.

Chaloner, K. M., Church, T., Louis, T. A., and Matts, J. P., (1992). Graphical elicitation of a prior distribution for a clinical trial. *Technical Report.* Address: Division of Biostatistics, University of Minnesota.

Chaloner, K. M., (1991). Bayesian residual analysis in the presence of censoring. *Biometrika*, 78,637–644.

Chiang, C. L., (1968). *Introduction to Stochastic Processes in Biostatistics.* John Wiley and Sons.

Chilvers, C. E. D., Fayers, P. M., Freedman, L. S., Greenwood, R. M., Machin, D., Palmer, N., and Westlake, A. J., (1988). Improving the quality of data in radomized clinical trials : the COMPACT computer package. *Statistics in Medicine*, 7,1165–1170.

Clayton, D. G. and Cuzick, J., (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society (A)*, 148,82–117.

Clayton, D. G., (1988). The analysis of event history data : A review of progress and outstanding problems. *Statistics in Medicine*, 7,819–841.

Clayton, D. G., (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, 47,467–485.

Cornfield, J. and Detre, K., (1977). Bayesian life table analysis. *Journal of the Royal Statistical Society (B)*, 39,86–94.

Cox, D. R. and Hinkley, D. V., (1974). *Theoretical Statistics.* Chapman and Hall, London.

Cox, D. R. and Miller, H. D., (1965). *The Theory of Stochastic Processes.* Methuen, London.

Cox, D. R. and Oakes, D., (1984). *Analysis of Survival Data.* Chapman and Hall, London.

Cox, D. R. and Reid, N., (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society (B)*, 49,1,1–39.

Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J., and Jones, D. R., (1992). Quality-of-life assessment : can we keep it simple? *Journal of the Royal Statistical Society (A)*, 155,353–393.

Cox, D. R., (1970). *Analysis of Binary Data.* Monographs in Statistics and Applied Probability. Chapman and Hall, London.

Cox, D. R., (1972). Regression models and life-tables. *Journal of the Royal Statistical Society (B)*, 39,187–220.

Cox, D. R., (1975). Partial likelihood. *Biometrika*, 62,2,269–276.

Crowder, M. J., Kimber, A. C., Smith, R. L., and Sweeting, T. J., (1991). *Statistical analysis of reliability data.* Chapman and Hall, London.

Davidson, J., (1988). *The reliability of mechanical systems.* Mechanical Engineering Publications Limited, London.

de Bruijn, N. G., (1958). *Asymptotic methods in analysis.* Dover, New York.

deBoor, C., (1978). *A practical guide to splines.* Springer-Verlag, New York.

DeGroot, M. H., (1986). *Probability and Statistics.* Addison-Wesley, Massachusetts, U.S.A, second edition.

Dellaportas, P. and Wright, D., (1991). Numerical prediction for the two parameter Weibull distribution. *The Statistician*, 40,365–372.

Dellaportas, P. and Wright, D., (1991). Positive embedded integration in Bayesian analysis. *Statistics and Computing*, 1,1–12.

Dennis, J. E. and Wei, H. H. W., (1979). The new unconstrained optimization algorithms which use function and gradient values. *Journal of OPtimization Theory and Applications*, 28,453–483.

DerSimoniam, R. D. and Laird, N., (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7,177–188.

DeStavola, B. L., (1986). Sampling designs for short panel data. *Econometrica*, 54,415–424.

Dixon, D. O. and Simon, R., (1992). Bayesian subset analysis in a colorectal cancer clinical trial. *Statistics in Medicine*, 11,13–22.

DuMouchel, W., (1990). Bayesian metaanalysis. In Berry, D., editor, *Statistical Methodology in the Pharmaceutical Sciences*, pages 509–529. Marcel-Dekker, New York.

Duncan, W., Arnott, S. J., Jack, W. J. L., Orr, J. A., Kerr, G. R., and Williams, J. R., (1987). Results of two randomized clinical trials of neutron therapy in rectal adenocarcinoma. *Rad Oncol*, 8,191–8.

Durrleman, S. and Simon, R., (1989). Flexible regression models with cubic splines. *Statistics in Medicine*, 8,551–561.

Erdelyi, A., (1956). *Asymptotic expansions.* Dover Publications, New York.

Errington, R. D., Ashby, D., Gore, S. M., Abrams, K. R., Myint, S., Bonnett, D. E., Blake, S. W., and Saxton, T. E., (1991). High energy neutron treatment for pelvic cancers - study stopped because of increased mortality. *British Medical Journal*, 302,1045–1051.

Fayers, P. M. and Jones, D. R., (1983). Measuring and analysing quality of life in cancer clinical trials : a review. *Statistics in Medicine*, 2,429–446.

Fayers, P. M., Bleehen, N. M., Girling, D. J., and Stephens, R. J., (1991). Assessment of quality of life in small-cell lung cancer using a daily diary card developed by the Medical Research Council lung cancer working party. *British Journal of Cancer*, 64,299–306.

Fisher, R. A., (1958). *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

Fleiss, J. L., (1981). *Statistical Methods for Rates and Proportions.* Probability and Mathematical Statistics. Wiley, New York, second edition.

Fleming, T. R. and Harrington, D. P., (1991). *Counting processes and survival models.* Wiley.

Freedman, L. S. and Spiegelhalter, D. J., (1983). The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *The Statistician*, 32,153–160.

Freedman, L. S. and Spiegelhalter, D. J., (1992). Application to decision making during a clinical trial. *Statistics in Medicine*, 11,13–35.

Freedman, B., (1987). Equipoise and the ethics of clinical research. *New England Journal of Medicine*, 317,141–145.

Frydman, H., (1991). Nonparametric maximum likelihood estimation for a periodically observed markov illness-death process, with application to diabetes survival data. *Research Report, University of Copenhagen*, 3.

Gamerman, D. and West, M., (1987b). An application of dynamic survival models in unemployment studies. *The Statistician*, 36,269–274.

Gamerman, D., (1987a). Dynamic inference on survival functions. In *Probability and Bayesian Statistics*, pages 183–192. Plenum, New York.

Gamerman, D., (1991). Dynamic Bayesian models for survival data. *Applied Statistics*, 40,63–79.

Geisser, S., (1984). On prior distributions for binary trials (with discussion). *The American Statistician*, 38,244–251.

Gelfand, A. E. and Smith, A. F. M., (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85,398–409.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M., (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. Research report, Nottingham University.

Geman, S. and Geman, D., (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE*, 6,721–741.

Genest, C. and Zidek, J. V., (1986). Combining probability distributions. *Statistical Science*, 1,114–147.

Gilks, W. R. and Wild, P., (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41,337–348.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D., and Kirby, A. J., (1993). Modelling complexity : Applications of Gibbs sampling in medicine. *Journal of The Royal Statistical Society (B)*, 55. To Appear.

Gilks, W. R., (1986). The relationship between birth history and current fertility in developing countries. *Population Studies*, 40,437–455.

Gill, R. D., (1984). Understanding Cox's regression model : A Martingale approach. *Journal of the American Statistical Association*, 79,441–447.

Glasziou, P. P., Simes, R. J., and Gelber, R. D., (1990). Quality adjusted survival analysis. *Statistics in Medicine*, 9.

Gore, S. M., (1987). Biostatistics and the Medical Research Council. *Medical Research Council News*, pages 19–20.

Greenhouse, J., (1992). On some applications of Bayesian methods in cancer clinical trials. *Statistics in Medicine*, 11,37–53.

Griffin, T., Pajak, T., Laramore, G., and Davis, L., (1986). Analysis of neutron radiotherapy treatment complications. *Bull Cancer*, 73,582–586.

Harmer, M. H., (1978). *TNM - classification of malignant tumours.* International Union Against Cancer, Geneva.

Hills, S. E. and Smith, A. F. M., (1992). Parameterization issues in Bayesian inference. In Berger, J. O., Bernardo, J. M., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, Oxford. Oxford University Press.

Hjort, N. L., (1986). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scandinavian Journal of Statistics*, 13,63–85.

Holford, T. R., (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, 36,299–305.

Hougaard, P. and Madsen, E. B., (1985). Dynamic evaluation of short-term prognosis after myocardial infarction. *Statistics in Medicine*, 4,29–38.

Hughes, M. D., (1991). Practical reporting of Bayesian analysis of clinical trials. *Drugs Infomation Bulletin*, 25,3,381–393.

Jacobsen, M., (1984). Maximum likelihood estimation in the multiplicative intensity model : A survey. *International Statistical Review*, 52,193–207.

Jeffreys, H., (1961). *Theory of probabitity.* Oxford University Press, Oxford.

Johansen, S., (1983). An extension of Cox's regression model. *International Statistical Review*, 51,165–174.

Johnson, N. L. and Kotz, S., (1970). *Continuous univariate distributions - 1.* John Wiley and Sons, New York.

Kalbfleisch, J. D. and Lawless, J. F., (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80,863–871.

Kalbfleisch, J. D. and MacKay, R. J., (1978). Remarks on a paper by Cornfield and Detre. *Journal of the Royal Statistical Society (B)*, 40,175–177.

Kalbfleisch, J. D. and Prentice, R. L., (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60,267–278.

Kalbfleisch, J. D. and Prentice, R. L., (1980). *The Statistical Analysis of Failure time data.* John Wiley and Sons, New York.

Kalbfleisch, J. D., (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society (B)*, 40,214–221.

Kalbfleisch, J. D., (1985). *Probability and statistical inference*, volume 2. Springer-Verlag, New York.

Kaplan, E. L. and Meier, P., (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53,457–481.

Karnofsky, D. A. and Burchenall, J. H., (1949). The clinical evaluation of chemotherapeutic agents in cancer. *In: Evaluation of Chemotherapeutic Agents*, pages 199–205. Columbia University Press.

Kass, R. E. and Slate, E. H., (1992). Reparameterization and diagnostics of posterior non-Normality. In Berger, J. O., Bernardo, J. M., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, Oxford. Oxford University Press.

Kay, R. D., (1982). The analysis of transition times in multi-state stochastic processes using proportional hazard regression models. *Communications in Statistics-A*, 11,1743–1756.

Kay, R. D., (1984). Goodness of fit methods for the proportional hazards regression model: a review. *Revue Epidemiologie et Sante Publique.*, 32,185–198.

Kay, R. D., (1986). A Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics*, 42,855–865.

Keiding, N. and Andersen, P. K., (1987). Nonparametric estimation of transition intensities and transition probabilities : a case study of a two-state markov process. *Research Report, University of Copenhagen*, 8.

Keiding, N., (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions of the Royal Society London*, 332,487–509.

Kirby, A. J., (1991). *Statistical Modelling for the Precursors of Cervical Cancer.* University of Cambridge. Unpublished Ph.D. Thesis.

Kuo, L. and Smith, A. F. M., (1992). Bayesian computations in survival models via the Gibbs sampler. In Klein, J. P. and Goel, P. K., editors, *Survival analysis: state of the art*, pages 11–24. Kluwer Academic Publishers.

Lagakos, S. W., (1976). A stochastic model for censored-survival data in the presence of an auxiliary variable. *Biometrics*, 32,551–559.

Laird, N. and Oliver, D., (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76,231–240.

Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., and Chalmers, T. C., (1992). Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine*, 327,248–254.

Lawless, J. F., (1987). Regression methods for Poisson process data. *Journal of The American Statistical Association*, 83,808–815.

Lee, T. C., Judge, G. C., and Zellner, A., (1976). *Estimating the parameters of the Markov probability model from aggregate time series data.* North-Holland, second edition.

Leonard, T., (1972). Bayesian methods for binomial data. *Biometrika*, 59,581–589.

Leonard, T., (1975). Bayesian estimation methods for two-way contingency tables. *Journal of The Royal Statisticsl Society (B)*, 37,23–37.

Lexis, W., (1875). Einleitung in die theorie der bevölkerungsstatistik. Translated into English by N. Keyfitz, in *Mathematical Demography* (1977), editors D.Smith & N.Keyfitz. Berlin : Springer.

Lindley, D. V., (1964). Bayesian analysis of contingency tables. *Annals of Mathematical Statistics*, 35,1622–1643.

Lindley, D. V., (1965). *Introduction to Probability and statistics from a Bayesian viewpoint. Part 2 - Inference.* Cambridge University Press, Cambridge.

Lindley, D. V., (1971). The estimation of many parameters. *Foundations of Statistical Inference*, pages 435–455. Editors: V. P. Godamle and D. A. Sprott.

Lloyd, D., Duffty, P., Durbin, G., Morgan, I., Morley, C., Greenough, A., South, M., Hume, R., Marshall, T., Turner, T., Sutton, A., Poole, J., Gamsu, H., Cooke, R., Rennie, J., Hey, E., Thurlbeck, S., McIntosh, N., Dear, P., Wheeler, D., Gore, S., Miller, N., and Bangham, A., (1987). Ten centre trial of artificial surfactant (artificial lung expanding compound) in very premature babies. *British Medical Journal*, 294,991–996.

Longini, I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., Lemp, G. F., and Hethcote, H. W., (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine*, 8,831–843.

Mantel, N. and Haenszel, W., (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22,719–748.

Maritz, J. S. and Lwin, T., (1989). *Empirical Bayes Methods.* Chapman and Hall, London, second edition.

Marshall, R. J., (1988). Bayesian analysis of case-control studies. *Statistics in Medicine*, 7,1223–1230.

Martz, H. F. and Waller, R. A., (1982). *Bayesian Reliability Analysis.* Wiley, New York.

McCullagh, P. and Nelder, J. A., (1989). *Generalized Linear Models.* Monographs on Statistics and Applied Probability 37. Chapman and Hall, London, second edition.

Morris, C. N. and Mormand, S. L., (1992). Hierarchical models for combining information and for meta-analyses. In Bernardo, J., Berger, J. O., Lindley, D., and Smith, A., editors, *Bayesian Statistics 4*, Oxford. Oxford University Press.

Naylor, J. C. and Shaw, J. E. H., (1985). *Bayes Four - User Guide.* University of Nottingham.

Naylor, J. C. and Smith, A. F. M., (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, 31,214–225.

Numinen, M. and Mutanen, P., (1987). Exact Bayesian analysis of two proportions. *Scandinavian Journal of Statistics*, 14,67–77.

Office of Population Censuses and Surveys (OPCS), (1990). *Review of the national cancer registration system.* Her Majesty's Stationery Office, London.

O'Hagan, A., Woodward, E. G., and Moodaley, L. C., (1990). Practical Bayesian analysis of a simple logistic regression : predicting corneal transplants. *Statistics in Medicine*, 9,1091–1101.

O'Hagan, A., (1988). *Probability : methods and measurement.* Chapman and Hall, London.

Olschewski, M. and Schumacher, M., (1990). Statistical analysis of quality of life data in cancer clinical trials. *Statistics in Medicine*, 9,749–763.

Peto, R. and Lee, P., (1973). Weibull distributions for continuous-carcinogenesis experiments. *Biometrics*, 29,457–470.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., and Smith, P. G., (1976). Design and analysis of randomised clinical trials requiring prolonged observation of each patient. i: Introduction and design. *British Journal of Cancer*, 34,585–612.

Peto, R., (1987). Why do we need systematic overviews of randomized trials? *Statistics in Medicine*, 6,233–240.

Pocock, S. J. and Hughes, M. D., (1990). Estimation issues in clinical trials and overviews. *Statistics in Medicine*, 9,657–671.

Pocock, S. J., (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29,175–188.

Pocock, S. J., (1983). *Clinical trials - a practical approach.* Wiley, Chichester.

Pocock, S. J., (1991). A perspective on the role of quality-of-life assessment in clinical trials. *Controlled Clinical Trials*, 12,257–265.

Pointon, R. S., Read, G., and Greene, D., (1985). A randomized comparison of photons and 15 MeV neutrons for the treatment of carcinoma of the bladder. *British Journal of Radiology*, 58,219–224.

Prentice, R. L., Williams, B. J., and Petersen, A. V., (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68,373–379.

Prentice, R. L., (1975). Discimination among some parametric models. *Biometrika*, 62,607–614.

Raghunathan, T. E., (1991). Pooling controls from different studies. *Statistics in Medicine*, 10,1417–1426.

Raiffa, H. and Schlaifer, R., (1961). *Applied statistical decision theory.* Havard Business School, Boston.

Ripley, B. D., (1987). *Stochastic simulation.* Wiley, Chichester.

Rogatko, A., (1991). Bayesian approach for meta-analysis of controlled clinical trials. Research report, IARC, Lyons, France.

Rothschild, V. and Logothetis, N., (1985). *Probability Distributions.* John Wiley and Sons, New York.

Rubin, D., (1987). Comment: a non-iterative sampling/importanceresampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. *Journal of the American Statistical Association*, 82,543–546.

Savage, L. J., (1971). Elicitation of personal probabilities and expectations. *The Journal of the American Statistical Association*, 66,783–801.

Schumacher, M., Olschewski, M., and Schulgen, G., (1991). Assessment of quality of life in clinical trials. *Statistics in Medicine*, 10,1915–1930.

Shaw, J. E. H., (1988). *Numerical Integration and Display Methods for Bayesian Inference.* University of Nottingham. Unpublished Ph.D. Thesis.

Silverman, B. W., (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Singpurwalla, N. D. and Song, M. S., (1987). The analysis of Weibull lifetime data incorporating expert opinion. In *Probability and Bayesian Statistics*, pages 431–442. Plenum, New York.

Skene, A. M. and Wakefield, J. C., (1990). Hierarchical models for multicentre binary response studies. *Statistics in Medicine*, 9,919–929.

Sleeper, L. A. and Harrington, D. P., (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85,412,941–949.

Smith, A. F. M. and Roberts, G. O., (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte carlo methods. *Journal of The Royal Statistical Society (B)*, 55. To Appear.

Smith, A. F. M., Skene, A. M., Shaw, J. E. H., Naylor, J. C., and Dransfield, M., (1985). The implementation of the Bayesian paradigm. *Communications in Statistics*, 14,1079–1102.

Smith, A. F. M., Skene, A. M., Shaw, J. E. H., and Naylor, J. C., (1987). Progress with numerical and graphical methods for practical Bayesian statistics. *The Statistician*, 36,75–82.

Smith, J. Q., (1985). *Decision Analysis - A Bayesian Approach.* Chapman and Hall, London.

Spiegelhalter, D. J. and Freedman, L. S., (1988). Bayesian approaches to clinical trials. In Bernardo, J., DeGroot, M., Lindley, D., and Smith, A., editors, *Bayesian Statistics 3*, pages 453–477, Oxford. Oxford University Press.

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. B., (In press). Applying Bayesian thinking in drug development and clinical trials. *Statistics in Medicine.*

Statistical Sciences Inc., (1990). *Splus for Sun workstations.* Oxford, U.K. Version 2.3.

Stigler, S. M., (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, 1,359–362.

Sweeting, T. J., (1981). Scale parameters : A Bayesian treatment. *Journal of the Royal Statistical Society (B)*, 43,3,333–338.

Sweeting, T. J., (1982). A Bayesian analysis of some pharmacological data using a random coefficient regression model. *Applied Statistics*, 31,205–213.

Sweeting, T. J., (1984). Approximate inference in location-scale regression models. *Journal of the American Statistical Association*, 79,847–850.

Sweeting, T. J., (1987). Approximate Bayesian analysis of censored survival data. *Biometrika*, 74,809–816.

Tandon, P. K., (1990). Applications of global statistics in analysing quality of life data. *Statistics in Medicine*, 9,819–827.

Tanner, M. A. and Wong, W. H., (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82,528–550.

Taulbee, J. D., (1979). A general m odel for the hazard rate with covariables. *Biometrics*, 35,439–450.

Thisted, R. A., (1988). *Elements of Statistical Computing - Numerical Computation.* Chapman and Hall, New York.

Thompson, S. G. and Pocock, S. J., (1991). Can meta-analyses be trusted? *The Lancet*, 338,1127–1130.

Tiao, G. C. and Box, G. E. P., (1967). Bayesian analysis of a three-component hierarchical design model. *Biometrika*, 54,109–125.

Tierney, L. and Kadane, J. B., (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81,82–86.

Tierney, L., Kass, R. E., and Kadane, J. B., (1989). Fully exponential Laplace approximations of expectations and variances of non-positive functions. *Journal of the American Statistical Association*, 84,710–716.

Tierney, L., (1990). *XLISP-STAT - An Object-Oriented Environment for Statistical Computing and Dynamic Graphics.* John Wiley, New York.

Tsiatis, A. A., (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, 68,311–315.

Tuma, N. B., Hannan, M. T., and Groeneveld, L. P., (1979). Dynamic analysis of event histories. *American Journal of Sociology*, 84,820–854.

Van Druten, J. A. M., Reintjes, A. G. M., Jager, J. C., Heisterkamp, S. H., Poos, M. J. J. C., Coutinho, R. A., Dijkgraaf, M. G. W., and Ruitenberg, E. J., (1990). HIV infection dynamics and intervention experiments in linked risk groups. *Statistics in Medicine*, 9,721–736.

Voelkel, J. G. and Crowley, J., (1984). Nonparametric inference for a class of semi-markov processes with censored observations. *The Annals of Statistics*, 12,142–160.

Wahba, G., (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society (B)*, 40,364–372.

WATCOM, (1989). *Maple Reference Manual.* WATCOM Publications Limited, Waterloo, Ontario, Canada.

Wegman, E. J. and Wright, I. W., (1983). Splines in statistics. *Journal of the American Statistical Association*, 78,382,351–363.

West, M. and Harrison, P. J., (1989). *Bayesian Forecasting and dynamic models.* Springer series in statistics. Springer-Verlag, New York.

West, M. and Mortera, J., (1987). Bayesian models and methods for binary times series. In *Probability and Bayesian Statistics*, pages 487–495. Plenum, New York.

West, M., Harrison, P. J., and Migon, H. S., (1985). Dynamic generalised linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80,73–97.

Zelen, M. and Parker, R. A., (1986). Case-control studies and Bayesian inference. *Statistics in Medicine*, 5,261–269.