Bayesian Survival Analysis With Missing Data Using Integrated Nested Laplace Approximation

Muhammad Irfan bin Abdul Jalal

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics, Statistics & Physics
Newcastle University
Newcastle upon Tyne
United Kingdom

February, 2020

Abstract

Bayesian survival analysis has benefitted from the introduction of Markov Chain Monte Carlo (MCMC) since the 1990s. However, MCMC has high computational cost and requires tuning and convergence checking. These hamper its usefulness. Integrated Nested Laplace Approximation (INLA) is a convenient alternative to MCMC due to its efficiency and straightforward execution to obtain the posterior distributions of relevant survival parameters such as the regression coefficients and Weibull shape parameter. This has been demonstrated in parametric and semi-parametric proportional hazard and piecewise constant hazard models. Nevertheless, it has not been possible until now to use INLA when covariate data are missing since it neither can integrate out missing covariate data nor is it satisfactory to use other subsidiary methods such as multiple imputation to overcome this difficulty.

We therefore investigate the application of INLA to piecewise exponential constant hazard models when covariate information is missing. We extend and modify the INLA within MCMC method to circumvent the missing covariate data problem in survival cases, assuming that the data are missing at random. We use both hierarchical and autoregressive priors for the baseline log-hazard and covariate effects and compare the results with those obtained by MCMC.

The methods are applied to three different data sets; Catheter-related kidney infection, Scotland-Newcastle Lymphoma Group Non Hodgkin Lymphoma, and Malaysian Hospital Universiti Sains Malaysia Advanced Lung Cancer data sets. The priors are constructed based on the information obtained from the meta-analysis of results from previous studies. The results obtained demonstrate that the developed methods are suitable for various survival data sets with reasonable numbers of missing covariate values, making this proposed method a convenient alternative to standard MCMC algorithms for survival analysis.

I dedicate this thesis to GOD

Acknowledgements

My first word goes to GOD WHO has taken me this far. To my wife Mazian Ismail for her enduring and endless support. For Maisarah, my daughter who always made me laugh. To my kind and helpful supervisor, Dr Malcolm Farrow, I am totally grateful to know you as a person who taught me the meanings of research and friendship. To my co-supervisor, Dr Daniel Henderson, your honesty and kindness are two things that I always cherish. To my sponsor, MARA, thousands of thanks for all the financial support and unwavering support for the past 4 years whilst I was (and am) still struggling with my doctoral studies. To my friends and colleagues, Juliana Iworikumo Consul, Safwan Ibrahim and Wael Al-Taei, many thanks for the beautiful friendship that we have forged together since the last 4 years. May you all be successful in your future undertaking.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Thesis Outline

1.1 Motivation

This research is concerned with the Bayesian analysis of survival data. The focus is on the development of novel techniques to circumvent the problems with missing covariate data in survival models when posterior inference is obtained using Integrated Nested Laplace Approximation (INLA).

Classical analyses (such as Cox proportional hazard model with unspecified baseline hazard function) are the usual choice among applied researchers when modelling continuous time to event data. The reliance of these techniques on partial likelihood for parameter inference has several limitations since they restrict the utilization of information from the full likelihood and the complete estimation of unknown elements in the model. Survival analysis using Bayesian methodology offers a better alternative and provides a complete and accurate posterior inference for the parameters of the models.

Various techniques have been proposed in the Bayesian literature for the modelling of time-to-event data to accommodate variations in dataset characteristics such as within-group correlation, the effects of non-linear covariates and spatiotemporal differences. A general overview of Bayesian survival analysis is given by Ibrahim et al. (2001). Some research into Bayesian survival analysis has investigated the use of nonparametric and semiparametric forms for the baseline hazard function. For example, Kalbfleisch (1978) introduced piecewise constant hazard models which assume that the baseline log hazard is constant over certain time intervals. Another active area has involved relating survival to spatial location. Banerjee et al. (2003) used a parametric Weibull baseline hazard with spatial component in their geostatistical model. Guo and Carlin (2004) investigated

joint analyses of survival and longitudinal data. They demonstrated that longitudinal and survival data are conditionally independent and postulated that a latent bivariate Gaussian process was present in the model.

Markov chain Monte Carlo (MCMC) methods have been the backbone of Bayesian modelling for the past three decades (Ashby, 2006). The two most widely used MCMC algorithms, Metropolis-Hastings (Hastings, 1970) and Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990), have provided approximations to previously intractable posterior distributions in high dimensionality problems ("curse of dimensionality"). Despite the popularity of MCMC and a large number computational modalities such as WinBUGS (Lunn et al., 2000) and R (R Core Team, 2018) that support MCMC execution for broad classes of models, there are two main drawbacks:

1. high computational time (hours or even days for the computation of posterior marginal distributions),

2. manual "tuning" by the users for convergence diagnosis and judgment on the accuracy of the approximation, especially when modelling spatial components or semi-parametric (smooth) effects.

In order to address these difficulties, the INLA method has been proposed (Rue et al., 2009). In recent years, Martins et al. (2013) introduced new features with INLA which includes additional distributions for survival analysis that can be fitted by INLA. Jiang et al. (2014) gave examples of INLA utility for geostatistical survival models for environmental risk assessment using retrospective cohort data collected from Pickering Nuclear Generating Station, Canada. Martins and Rue (2014) further extended INLA flexibility to accommodate a class of near-Gaussian latent models such as those whose latent field has a near-Gaussian distribution particularly those with a unique and unimodal mode, finite first two moments and distributions with density that has full support on the real line. In Muff et al. (2015), INLA was also extended and shown to be an efficient Bayesian analytical approach for measurement error (ME) models.

In survival analysis, many of the Cox-type models in the literature may be viewed as latent Gaussian models which are a large class of models with jointly-Gaussian latent variables. Therefore, complex constituents of the models (spatiotemporal effects, an assortment of frailty effects) can be included without difficulties since they can be regarded as trivial changes to the Gaussian components of such a model. For Bayesian inferential purposes, a quicker and more exact approximation to the posterior marginal distributions can be made possible using INLA. The method uses a brilliant application of Laplace

approximations and advanced numerical methods. These result in a significant reduction in computational time (compared to standard MCMC methods). Besides, no significant manual tuning or interaction is required from the users. This enables INLA to be used like a"black box". An R package is available (R-INLA) from `http://www.r-inla.org` to aid the execution of INLA computation.

In medicine and public health, INLA methodology has been used for the modelling of spatial and spatio-temporal models (Blangiardo et al., 2013), for a multi-state model for pancreatic cancer progression (Alvaro-Meca et al., 2013), stomach cancer incidence in Southern Portugal (Papoila et al., 2014) and others. In the field of ecological modelling, Carson and Flemming (2014) further highlighted the usefulness of INLA. The authors demonstrated that a more efficient technique of hierarchical spatio-temporal modelling was developed when INLA was combined with a stochastic partial differential equations (SPDE) approach for acoustic telemetry data, specifically for detecting or tagging of Sable Island grey seals (Halichoerus grypus) at sea.

The nascent idea about INLA being a fast approximation method for posterior inference in Bayesian statistics was expounded in a much earlier article, Rue (2001). The author demonstrated that posterior marginal distributions for models that exhibit the properties of Gaussian Markov Random Fields (GMRF), which is also known as conditional autoregression (Besag and Kooperberg, 1995), may be efficiently computed since they are straightforwardly amenable to numerical integration for sparse matrices. This finding was further followed by an attempt to extend it to models that possess hidden Gaussian Markov random field properties (Rue et al., 2004). The ideas contained in these papers were taken one step forward by Rue and Martino (2007). The contributors developed a deterministic substitute to MCMC, which can be regarded as an INLA prototype, utilizing the sparsity of the precision matrix Q in GMRF. This eliminates the need for inversion of the precision matrix Q for computing marginal variances. This eventually culminates in the seminal paper by Rue et al. (2009) which cemented the birth of INLA as a fast deterministic approach for computing posterior distributions.

In their original contribution, Rue et al. (2009) had demonstrated that INLA is an efficient technique for Bayesian inference for a subclass of structured additive regression models called latent Gaussian models (LGM). In comparison to MCMC, INLA is superior in terms of accuracy and speed in the computation of posterior marginals in LGM. This is due to the additive errors of MCMC which implies that longer computational time is required just to increase the precision of the estimate by a single digit. Apart from that, the probabilities in the tail region are much harder to compute due to the additive errors

of Monte Carlo estimates in MCMC. INLA is able to avoid these limitations since it has only a relative error nature which permits the accurate approximation of tail probabilities.

Nevertheless, the robustness of the INLA scheme suffers from some limitations. Firstly, the dependence on a Gaussian distribution for the latent field impairs its application to other models that are not within the LGM class. For instance, a survival model with gamma frailty (Ibrahim et al., 2001) whose random effects follow a log gamma distribution is not amenable to using INLA methodology. Martins and Rue (2014) attempted to provide an alternative strategy to this problem by embedding the initial standard model within a LGM framework to create a more flexible model. They then corrected the independent non-Gaussian components of the latent field for asymmetry and kurtosis using near-Gaussian distributions. However, this strategy is hampered by the unavailability of clear diagnostics to identify which non-Gaussian distributions really have near-Gaussian properties. Therefore, further strategies are required to extend the work of Martins and Rue (2014) to permit a wider class of non-Gaussian models which can be computed using the INLA scheme.

Secondly, the modest number of hyperparameters ($m < 6$) that can be accommodated by INLA severely limits the application of this scheme. According to Rue et al. (2009), even a moderate number of hyperparameters ($6 \leq m \leq 12$) may prove to be quite taxing for INLA, resulting in an exponential rise in computing time. The low dimensional hyperparameters and Markov random field are both underlying requirements for the latent Gaussian field and INLA computation. Perhaps these are quite unrealistic in many biological and medical applications (Wikle and Holan, 2009). More works are thus required to expand the utility of INLA in models with high-dimensional hyperparameters ($m > 12$).

Thirdly, the impact of missing covariates on the accuracy and flexibility of INLA computation has not been investigated in the context of survival analysis. In particular, work is required for dealing with non-Gaussian, especially categorical covariates, possibly using further latent Gaussian variables.

1.2 Research objectives

The primary objectives of this research are:

1. Compare the performance of different computational approaches (MCMC vs INLA) in Bayesian inference for survival analysis models.

2. Investigate the construction of appropriate forms for the prior distributions of the log hazard.

3. Assess the performance of INLA in missing covariates values and develop approaches to improve the flexibility of INLA in this setting.

4. Apply the formulated approaches to practical problems using real medical datasets and evaluate the strengths and limitations of each approach

Therefore, this research is concerned with answering these specific research questions:

- How can we fit piecewise constant hazard models using INLA and what is its efficiency in obtaining posterior quantities of interest relative to MCMC approach?
- How can we resolve INLA's deficiency in dealing with missing covariate information in the most appropriate and efficient way?
- How can prior distributions for the log hazard be meaningfully constructed based on the information from previous research?
- What are the advantages of INLA in the construction of prognostic indices for different real medical datasets compared to MCMC?

1.3 Thesis outline

Chapter 2 introduces the fundamental ideas of survival analysis and Bayesian inference. We give a brief introduction to survival analyses which is then followed by discussion of the types of censoring and truncation. We then explain two commonly-used survival models: the proportional hazard (PH) and accelerated failure time (AFT) models. Parametric survival models are then introduced, with special attention given to deriving probability density and hazard functions, survival time distributions and likelihood functions for survival times that follow Weibull distributions. Frailty is subsequently introduced and the derivations of the marginal survival time distributions and hazard function are illustrated. In section 2.7, the piecewise constant hazard model is introduced and the methods used for selecting the cut points for time intervals and the inclusion of time-varying covariate effects are discussed in detail. In 2.8, the fundamental aspects of Bayesian inference in a survival analysis context are succintly expounded and this is then followed by a brief dicussion of Markov Chain Monte Carlo (MCMC) which includes illustrations of different

variants of the Metropolis Hastings and Gibbs sampling algorithms. We then discuss the data augmentation approach as a means to handle missing covariate values. The computation of posterior quantities of interest is then illustrated using a catheter-related urinary tract infection data set and this is performed using RJAGS.

In Chapter 3, we introduce latent Gaussian models (LGMs) and Gaussian Markov random fields (GMRF). Firstly, the relationship between generalised linear models and LGMs is discussed in detail. This is then followed by elaborations on GMRF characteristics and proofs for the sparsity of the precision matrix, Q, are presented. Subsequently, we demonstrate the derivations of the recursive formula for computing marginal variances for each parameter of interest and conditional GMRF distributions in the presence of some linear constraints. We then explicitly describe how a Bayesian hierarchical model can be viewed as a GMRF. The GMRF approximation for non-Gaussian likelihood is then illustrated.

In Chapter 4, the main ideas of the integrated nested Laplace approximation (INLA) will be propounded in detail. The ideas behind the use of a Laplace approximation for computing a posterior distribution are first described. We then present an outline of INLA methodology which includes the use of the central composite design (CCD) approach to tackle the problems with high numbers of hyperparameters. Next, we demonstrate INLA implementation using two simple examples in 4.4. In 4.5, we show how piecewise constant hazard models can be fitted using INLA. We then show how the missing covariate data problem can be addressed with a new computational approach which is the INLA within MCMC (INLA-MCMC) method. This is a novel approach which is introduced in this thesis.

In Chapter 5, we introduce the three data sets used for application purposes: the kidney infection data set, Scotland and Newcastle Lymphoma Group (SNLG) Non Hodgkin Lymphoma (NHL) data set and a Malaysian lung cancer data set. For each data set, we give the operational definitions and the characteristics of each covariate and explain how the data were gathered.

In Chapter 6, we will give background information on how prior distributions can be constructed for shape parameters in Weibull survival models, baseline log hazard, coefficients of linear predictors and frailty variances. The constructions of two types of prior for piecewise constant hazard models (hierarchical priors and autoregressive priors) is then illustrated for the SNLG-NHL example.

In Chapter 7, the performance of the INLA-MCMC approach for dealing with missing covariate information will be assesed and compared with standard MCMC by applying

the methods to the example data sets. In Chapter 8, we shall give conclusions and some ideas for future works.

Chapter 2

Bayesian Survival Analysis

2.1 Introduction

In this chapter, two main ideas shall be discussed: the fundamental approaches and results in survival analysis and Bayesian Inference. In the first part, which consists of subsections 2.2 to 2.7, we will introduce the main ideas in survival analysis, truncation and censoring, the two main models used in survival analysis: Proportional Hazard (PH) and Accelerated Failure Time (AFT) models, parametric survival models, the effects of frailty on survival time distributions and piecewise constant hazard models.

In the second part, the primary ideas and standard results in Bayesian inference will be elucidated. This will include a brief introduction to Bayesian ideas where basic theory and notation will be introduced. We will subsequently discuss survival analysis within the Bayesian context, the use of the Markov Chain Monte Carlo (MCMC) approach for computing posterior distributions and the strategies for dealing with missing data using the data augmentation approach (Tanner and Wong, 1987). We will finally demonstrate the application of Bayesian survival analysis via the RJags (Plummer, 2016) package using a catheter-associated urinary tract infection dataset.

2.2 Introduction to survival analysis

Survival analysis is a statistical technique for the analysis of time-to-event data. This data has the form of times from a specific origin until a certain event of interest occurs (Collett, 2015). As an example, in a clinical trial, the time origin is usually at the time when the patients were initially enrolled into the study and the events of interest are

usually death, cancer progression or remission and other clinical events of interest.

For the analysis of survival data, the hazard function is usually employed to specify the survival models. To aid discussion later on, the following notations are introduced; $F(t)$ is the cumulative distribution function (cdf) of the survival time, $f(t)$ is the probability density function (pdf) of the survival time and the survival function, $S(t)$ is defined as $S(t) = \Pr(T > t)$. The pdf and cdf of survival times are given as

$$f(t) = \frac{dF(t)}{dt} \qquad \text{and} \qquad F(t) = \Pr(T \leq t).$$

Hence, the survival function can be rewritten as

$$S(t) = 1 - F(t).$$

The corresponding hazard function, $h(t)$, is given by the following limit of a conditional probability

$$h(t) = \lim_{\delta t \to 0} \frac{1}{\delta t} \Pr(t < T < t + \delta t | T > t).$$

Hence, the hazard function $h(t)$ can be rewritten as

$$h(t) = \lim_{\delta t \to 0} \frac{S(t) - S(t + \delta t)}{S(t)\delta t} = \frac{f(t)}{S(t)}.$$

There are two additional features that should be considered in survival analysis: censoring and truncation. These will be discussed in the next section.

2.3 Censoring and truncation

2.3.1 Types of censoring and truncation

The primary outcome in survival analysis is the time until the event of interest occurs. Unfortunately, this is often incompletely observed and, as a result, censored observations (observations with incomplete information on the event times) ensue. Several types of censoring have been established and they are

Right censoring (the most common) - The censoring time for the subjects (C_r) is less than the true survival time (T). Hence, it can be deduced that T is larger than C_r.

Left censoring - The true survival time, T, is known to be less than the left censoring time (C_l). For instance, the event occurs before the commencement of the study ($T < C_l$).

Interval censoring - The true survival time is not precisely known but is known to lie between two censoring limits ($C_l \leq T \leq C_r$).

Truncation differs from censoring. Truncation is about the entry of subjects that have the event times within the observational period or window of the studies. Truncated observations occur when only subjects whose event times are within the observation window (T_l, T_r) are observed. Two types of truncated observations have been described:

Left truncation (also known as delayed entry) - occurs when T_r is limitless ($T_r \to \infty$). Therefore, only those with observation times more than T_l are observed.

Right truncation - $T_l = 0$. The observation window is hence $(0, T_r)$. Hence, only individuals with event time less than a prespecified threshold of T_r are going to be included in the study.

2.3.2 Independent and non-informative censoring

Independent censoring is one of the most critical assumptions in survival analysis. In some cases it could be that the event that a patient is censored at time c is not independent of the patient's remaining lifetime. For example, patients who are ill and who have shorter expected remaining lifetime may be more likely to drop out of a study and be censored. When the independent censoring assumption is met, the actual survival time of a study participant (patient) is independent of the censoring mechanism that causes his or her observed survival time to be censored at time c (in this case, $c < T$). In other words, subjects who are censored at time c are still representative of other subjects who are still at risk at that time point. As a result, the censoring and survival times can be treated independently.

Suppose that each patient has a survival time T and a censoring time C. In some cases $C \to \infty$. If $C > T$, then the patient is not censored and the event indicator $\delta = 1$. If $C \leq T$ then the patient is censored and $\delta = 0$.

Let the parameters of the lifetime distribution be θ_T and those of the censoring process be θ_C. The independence condition is:

$$\Pr(\boldsymbol{T < t \mid C < s, \theta_T, \theta_C}) = \Pr(\boldsymbol{T < t \mid C \geq s, \theta_T, \theta_C}) \qquad \text{for all } t, s > 0.$$

If this holds then the joint density of T and C given θ_T and θ_C is

$$f_T(\boldsymbol{t} \mid \boldsymbol{\theta_T})\ f_C(\boldsymbol{c} \mid \boldsymbol{\theta_C}).$$

Hence, if we make an uncensored observation at time t then the likelihood contribution is

$$L_1 = f_T(\boldsymbol{t} \mid \boldsymbol{\theta_T}) \int_t^\infty f_C(\boldsymbol{c} \mid \boldsymbol{\theta_C}) dc$$

since we observe that $C > t$.

Similarly, the likelihood contribution of a censored observation, censored at time c is

$$L_0 = f_C(\boldsymbol{c} \mid \boldsymbol{\theta_C}) \int_c^\infty f_T(\boldsymbol{t} \mid \boldsymbol{\theta_T}) dt.$$

If, in addition, $\boldsymbol{\theta_T}$ and $\boldsymbol{\theta_C}$ contain no elements in common, then in terms of the likelihood function of $\boldsymbol{\theta}_T$,

$$L_1 \propto f_T(\boldsymbol{t} \mid \theta_T), L_0 \propto \int_c^\infty f_T(\boldsymbol{t} \mid \boldsymbol{\theta_C}) \mathrm{dt} = S_T(\boldsymbol{c} \mid \boldsymbol{\theta_t}) = 1 - F_T(\boldsymbol{c} \mid \boldsymbol{\theta_C}) = \Pr(\boldsymbol{T} > \boldsymbol{c} \mid \boldsymbol{\theta}_T) \tag{2.1}$$

and the censoring is said to be non-informative.

There are many scenarios where the assumption of independent censoring is not tenable. In the context of clinical trials, the violation of this assumption occurs when handling drop-out cases. The reason for drop-outs occurring in clinical trials can be two-fold: firstly, drop-out cases may reflect the study participants whose health is deteriorating more quickly, resulting in failing to come for follow-up visits or discontinuation of the drugs under investigation secondary to worsening physical condition. In a more extreme scenario, the study participants might even have passed away undocumented and hence are considered to be lost to follow-up and finally recorded as censored cases. As a result, the study participants who are censored at time t are no longer representative of the study participants who are still at risk at time t. This serious violation of independent censoring will distort the findings obtained from survival analysis (e.g. inflated hazard rate, obtaining survival time that is shorter than actual etc). However, for the purpose of our future analysis, we shall assume that independent censoring holds. When censoring occurs as a result of the end of the study, this is often a reasonable assumption.

2.4 Linking the effects of explanatory variables and lifetime distributions

There are two major approaches for modelling the effects of explanatory variables on the survival distribution: proportional hazard models and accelerated failure time models.

2.4.1 Proportional hazards (PH) model

Following the seminal paper by Cox (1972), the Cox proportional hazard model has been at the forefront of survival data analysis based on the semi-parametric approach. Since Cox's proportional hazards model has been the most popular regression approach for time-to-event problems used in many applications, only a brief account will be given on this topic. For detailed discussion, the original paper by Cox (1972) is an excellent exposition on this modelling approach.

According to Cox (1972), $h(t)$ is a function of a baseline hazard function, $h_0(t)$, covariates $z_1, z_2, z_3, \ldots, z_p$ and regression coefficients $\beta_1, \beta_2, \beta_3, \ldots, \beta_p$. In a proportional hazard model, this can be written as follows:

$$h(t|z_1, z_2, z_3, \ldots, z_p) = h_0(t) \exp(z_1\beta_1 + z_2\beta_2 + z_3\beta_3 + \ldots + z_p\beta_p).$$

Thus, in a proportional hazard model, it is assumed that changing the value of a covariate has a scaling effect on a hazard function of fixed form. Various types of proportional hazard models can be produced with different kinds of distributional assumptions for $h_0(t)$. For instance, if $h_0(t)$ is assumed to be constant over time, then an exponential regression model will be obtained.

Suppose that there are P covariates (p= 1,...,P) and n subjects ($i = 1, \ldots, n$). The vector of the explanatory variables (covariates) for subject i is thus given by $\boldsymbol{X}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,p})^T$. Each explanatory variable can be one of these: discrete, continuous, categorical or even mixed distribution covariates. For two subjects, which are denoted as subject i and subject j, that have different values for the explanatory variables, their respective hazards can be linked by the following equation:

$$h_i(t) = \lambda_{i,j} h_j(t).$$

We can regard subject j as the one having the baseline hazard h_0 function. As a result,

the hazard function for individual i can be said to be a multiple of the hazard of individual j. Hence we can further express the equation above in the following relationship:

$$h_i(t) = \lambda_i h_0(t).$$

Hence, λ_i is the hazard multiplier. Since the hazard cannot be negative, λ_i should thus be constrained to be greater than zero. To ensure this condition is satisfied, the covariates can be linked to the hazard multiplier through a logarithmic link function:

$$\log \lambda_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_p x_{p,i}.$$

If we set $\log \lambda_i = \eta_i$, then η_i represents the model's linear component. This quantity specifies whether an individual has a higher (if η_i is unusually large) or lower (if η_i is small) than average hazard of experiencing the event of interest. Hence, η_i is also known as the prognostic index, or risk score (Collett, 2015), for the i^{th} subject.

2.4.2 Accelerated failure time (AFT) models

The Accelerated Failure Time (AFT) approach uses a slightly different methodology compared to the PH models. In the AFT modelling paradigm, the covariates act multiplicatively on the time scale. The effects of the covariates can thus be interpreted as an acceleration factor: a rate at which an individual's survival progresses on the time axis. This approach has a highly instinctive appeal since AFT models can be regarded as a way of measuring the speed of the progression of the event of interest (such as disease progression based on some measured covariate values), making it attractive for application in many scientific fields especially in medical research.

As a motivating example, let us consider a group of patients enrolled in a randomised controlled trial. These patients were randomly allocated to either the standard or novel treatment (denoted by S and N). The event of interest is death due to some fatal illness, such as cancer. Within the context of an AFT model, the effect of the novel treatment is assumed to either "accelerate" or "decelerate" the passing of time relative to the standard treatment. To make things simple, we only consider a single covariate, which is the treatment group, in this example. Based on this assumption, it can be said that the probability of a patient who is allotted to the new treatment surviving beyond t is equivalent to the probability of a patient who is randomised to the standard treatment surviving beyond time t/ϕ. This ϕ is known as the acceleration factor and it is a positive constant.

To illustrate further, let S_s(t) and S_n(t) be the survival functions for patients who are on the standard and novel treatments, respectively. Under the AFT approach, the relationship between S_s(t) and S_n(t) can be written in the following fashion:

$$S_n(t) = S_s(t/\phi).$$

If ϕ lies between 0 and 1 (i.e. $0 < \phi < 1$), then the survival proportion of patients who received the standard treatment is greater than the survival proportion of patients who were on the novel treatment at time t (i.e. the novel treatment accelerates the progression of the fatal disease). However, if ϕ is greater than 1, it means that the survival proportion of patients receiving the standard treatment is less than the survival proportion of patients who are on the novel treatment at time t (ie the novel treatment "decelerates" the progression of the fatal disease). Hence, the novel treatment is better than the standard treatment in prolonging a patient's life if the acceleration factor $\phi > 0$. However, the interpretation should be reversed if the event of interest is, for instance, cancer remission. If ϕ lies between 0 and 1, it means that the proportion of subjects receiving the novel treatment who have achieved disease remission at time t is greater than the proportion of recipients of the standard treatment (i.e. the proportion of recipients of the novel treatment who still have active disease is less than the proportion of recipients of the novel treatment with active disease at time t). On the other hand, the opposite can be concluded if ϕ is more than 1.

Based on the following relationship between survival, hazard and probability density functions:

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

the hazard and survival functions for patients in the standard and novel treatment groups can be written as

$$f_n(t) = \phi^{-1} f_s(t/\phi) \qquad \text{and} \qquad h_n(t) = \phi^{-1} h_s(t/\phi).$$

If we define a new variable X as an indicator whether a patient is on the standard or novel treatment (0 = standard, 1 = novel), the hazard function can thus be written as follows:

$$h_i(t) = \phi^{x_i} h_0(t/\phi^{x_i}).$$

If we set $x_i = 0$ in the equation above for patients on the standard treatment, we will

obtain the baseline hazard function, h_0(t). Therefore, we can regard patients who receive the standard treatment as having a baseline hazard function. It is hence obvious that the hazard function of a patient who is on the novel treatment will be:

$$h_N = \phi^{-1} h_0(t/\phi).$$

Since ϕ should always be positive, we can reparametrise ϕ by writing: $\phi = e^{\alpha}$. Hence, the hazard function of the AFT model will have the following final form:

$$h_i(t) = e^{-\alpha y_i} h_0(t/e^{\alpha y_i}).$$

The AFT model can also be further generalised into models that include p covariates. We can generalise the final form of the AFT hazard function for a single covariate case into the following:

$$S_i(t) = S_0(t \exp\{-\eta_i\})$$

$$h_i(t) = \exp\{-\eta_i\} h_0(t \exp\{-\eta_i\})$$

$$\eta_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_p x_p.$$

Thus, S_0(t) can be considered as a survival function for subjects with zero values for all p covariates.

2.5 Parametric survival models

In an accelerated failure time model, we might specify the baseline survivor function as $S_0(t) = S(t;\theta_0) = 1 - F(t;\theta_0)$ where F(t;θ) is the distribution function for a particular family of survival distribution indexed by a parameter θ. Similarly, in a proportional hazard model, we might specify the baseline hazard function as

$$h_0(t) = \frac{f(t;\theta_0)}{1 - F(t;\theta_0)}$$

where f(t;θ) and F(t;θ) are respectively the pdf and distribution function for a particular family of survival distributions indexed by a parameter θ. A survival model specified in this way is known as a parametric survival model.

However we might prefer to allow more flexibility in the form taken by the baseline survival or hazard function. A model where the form is not specified is known as a

semi-parametric survival model. It is called semi-parametric rather than nonparametric because the dependence on covariates is still specified parametrically.

2.5.1 Weibull distribution

2.5.1.1 Probability density, hazard and survival distributions

In this section, we shall borrow the notation used by Collett (2015). The Weibull distribution was named after a Swedish mathematician Waloddi Weibull, who thoroughly elucidated it (Weibull, 1951). It was originally developed and used for reliability testing in the industrial setting. However, it has now occupied a critical role in parametric survival analysis due to its versatility and flexibility in adopting the characteristics of other distributions by changing its shape parameter, α. The probability density function of a 3-parameter Weibull distribution is given by

$$f(t) = \left(\frac{\alpha}{\eta}\right)\left(\frac{t-\zeta}{\eta}\right)^{\alpha-1} \exp\left\{ -\left(\frac{t-\zeta}{\eta}\right)^{\alpha}\right\}$$

where α is the shape parameter, η is the scale parameter and ζ is the location ($\alpha > 0, \eta > 0, -\infty < \zeta < \infty$ and $0 \leq t < \infty$). If we fix $\zeta = 0$, the Weibull pdf will reduce to the 2-parameter Weibull pdf:

$$f(t) = \left(\frac{\alpha}{\eta}\right)\left(\frac{t}{\eta}\right)^{\alpha-1} \exp\left\{ -\left(\frac{t}{\eta}\right)^{\alpha}\right\}.$$

In medical statistics, the Weibull distribution is often parametrised differently. The shape parameter α is retained in the usual form, and the scale parameter η is parametrised as

$$\lambda = \frac{1}{\eta^{\alpha}}$$

which results in

$$f(t) = \alpha\eta^{-\alpha}t^{\alpha-1}\exp\left\{ -\left(\frac{1}{\eta^{\alpha}}\right)t^{\alpha}\right\}.$$

Substituting $\eta^{-\alpha} = \lambda$ gives

$$f(t) = \alpha\lambda t^{\alpha-1}\exp\left\{ -\lambda t^{\alpha}\right\}, \qquad \lambda > 0, \quad \alpha > 0.$$

Hence, the cumulative distribution function of a Weibull distribution is

$$F(t) = 1 - \exp\Big\{ - \lambda t^{\alpha} \Big\}.$$

From these equations, we can easily deduce that, if we fix the shape parameter $\alpha = 1$ (i.e. the failure rate is constant across time), the Weibull distribution reduces to the exponential distribution

$$f(t) = \lambda \exp\Big\{ - \lambda t \Big\}.$$

The hazard function is

$$h_i(t) = \frac{f_i(t)}{s_i(t)} = \lambda_i \alpha t^{\alpha - 1}.$$

If $\alpha > 1$, the hazard rate will be monotonically increasing and, when $\alpha < 1$, the hazard is monotonically decreasing.

The survival function for Weibull-distributed survival times is

$$S(t) = \exp\Big\{ - \int_0^t \lambda \alpha u^{\alpha - 1} du \Big\} = \exp(- \lambda t^{\alpha}).$$

Through the scale parameter, λ, we may include the covariate effects as

$$\eta_i = g(\lambda_i)$$

where g should be a monotonic function that can be differentiated. Usually g(λ)= log(λ). Commonly, we use the following form to incorporate the linear predictor of a patient:

$$\eta_i = \boldsymbol{X}_i^T \, \boldsymbol{\beta}$$

where $\boldsymbol{X}_i^T$ represents the covariate values of patient i and $\boldsymbol{\beta}$ is a vector of regression coefficients, $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$. We can relate λ_i with η_i as follows:

$$\exp \, (\boldsymbol{X}^T \, \boldsymbol{\beta}) = \exp \, (\eta_i) = g^{-1}(\eta_i) = \lambda_i$$

For a subject with baseline covariate values (i.e. subjects with typical covariate pattern of baseline risk), $h_i(t)$ is reduced to the baseline hazard function

$$h_0(t) = \lambda_0 \alpha t^{\alpha - 1}.$$

Therefore, $h_i(t)$ can be written as

$$h_i(t) = \lambda_i^* \lambda_0 \alpha t^{\alpha-1},$$

where $\lambda_i = \lambda_i^* \lambda_0$,

2.5.1.2 Likelihood function for Weibull survival times

In deriving the likelihood function for Weibull survival times, independent censoring is assumed (for further details, refer to Section 2.2). Firstly, we shall define T_i and C_i corresponding to survival time and censoring time, respectively. Then, the follow-up time t_i experienced by the patient will take the minimum of either T_i or C_i (i.e it is a realisation of a random variable $\psi_i = \min(T_i, C_i)$). We thus observe the paired quantity (ψ_i, δ_i), where δ_i is an event indicator with value $\delta_i = 1$ if the event is observed and $\delta_i = 0$ if the observation is censored. From (2.1), the likelihood contribution is $f_T(t_i;\theta)^{\delta_i} S_t(t_i;\theta)^{1-\delta_i}$ so

$$L(\boldsymbol{\theta}) = \prod_{i=1} \{f_{T_i}(t_i;\boldsymbol{\theta})\}^{\delta_i} \{S_{T_i}(t_i;\boldsymbol{\theta})\}^{1-\delta_i}$$

where θ contains α and the coefficients in the linear predictor.

Since the hazard function is given by

$$h_i(t) = \frac{f_i(t)}{S_i(t)}$$

the likelihood can be simplified to

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \{h_i(t)\}^{\delta_i}\ S_i(t).$$

For a Weibull distribution, there are two parameters that govern its scale and shape (λ and α, respectively). Thus, $\boldsymbol{\theta} = (\lambda, \alpha)^T$. The likelihood function is given by

$$\begin{aligned} L(\lambda, \alpha \mid R) &= \prod_{i=1}^{n} \Big\{\lambda_i \alpha t_i^{\alpha-1}\Big\}^{\delta_i} \exp\Big\{-\lambda_i\ t_i^{\alpha}\Big\}, \\ &= \prod_{i \in E} \Big\{\lambda_i \alpha t_i^{\alpha-1}\}^{\delta_i} \prod_{i \in E \cup C} \exp\Big\{-\lambda_i\ t_i^{\alpha}\Big\}, \end{aligned}$$

$$= \left[\prod_{i\in E} \lambda_i\right] \alpha^{n_D} \left[\prod_{i\in E} t_i^{\alpha-1}\right] \exp\Big\{ - \sum_{i\in E\cup C} \lambda_i \; t_i^{\alpha}\Big\},$$

where R is a collection of $\boldsymbol{X}, n, S, T$ and δ ($R = \{\boldsymbol{X}, n, S, T, \delta\}$), E is the set of subjects experiencing the event and C is the set of subjects with censored survival times and n_D is the number of subjects with the event of interest. For the collection of quantities in R:

- $\boldsymbol{X}$ is the n $\times$ $(p+1)$ design matrix with the i^{th} row representing the covariate values for the i^{th} individual. In other words, the i^{th} row of $\boldsymbol{X}$ is $(1, x_{i,1}, x_{i,2}, x_{i,3}, \ldots, x_{i,p})$
- n is the total number of subjects;
- P is the number of covariates in the model;
- $\boldsymbol{T} = (t_1, t_2, \ldots, t_n)^T$, where t_i is the censoring or event time for individual i;
- $\delta = (\delta_1, \delta_2, \ldots, \delta_n)^T$ is the event indicator.

For easy algebraic manipulation, we take logs on both sides of the equation resulting in

$$\ell(\lambda, \alpha) = \log(L(\lambda, \alpha \mid R)) = \sum_{i\in E} \{\log\{\lambda_i\} + (\alpha - 1)\log\{t_i\}\} + n_D \log\{\alpha\} - \sum_{i=1}^{n} \lambda_i \; t_i^{\alpha}.$$

However, we have previously shown that the linear predictors and the scale parameter λ_i can be linked by

$$\lambda_i = \exp(\boldsymbol{X}^T \boldsymbol{\beta}).$$

The log likelihood function can hence be rewritten as

$$\ell(\boldsymbol{\beta},\ \alpha) = \log(L(\boldsymbol{\beta},\ \alpha \mid R)) = \sum_{i\in E} \left[\beta_0 + \sum_{p=1}^{P} \beta_p x_{i,p} + (\alpha - 1)\log(t_i)\right] +$$

$$n_D \log(\alpha) - \left[\sum_{i=1}^{n} \exp\Big\{\beta_0 + \sum_{p=1}^{P} \beta_p x_{i,p}\Big\} \; t_i^{\alpha}\right].$$

2.6 Random Effects in survival analysis (Frailties)

2.6.1 Frailty: a succinct introduction

The concept of frailty originated in the innovative work of Greenwood and Yule (1920) who conceptualised the idea that a negative binomial distribution of counts could be seen as a mixture of Poisson random variables for "accident proness" and data on recurrent bouts of disease attacks (Hougaard, 2013). The term frailty was firstly coined by Vaupel et al. (1979) to denote the random effects of unknown or unmeasured covariates in univariate survival data. However, frailty was applied to multivariate survival data a little earlier, based on work by Clayton (1978) who assigned a gamma distribution to the frailty term.

When we investigate the risk factors for certain conditions, we always assume that the individuals of interest have homogenous levels of risk factors (except for the ones that are measured and included in the model) for a certain event such as death. However, this is far from the truth since it is overwhelmingly implausible to consider, measure and include all of the covariates associated with the survival time of subjects. This is particularly true due to our limited knowledge with respect to the factors that are mechanistically and prognostically relevant in influencing the time to the event of our interest. In fact, despite taking into account all of the prognostic factors that are determinants of the differences in survival times, there may still be heterogeneity in survival times among those who possess similar covariate values. As a result, the non-observable variations in an individual's risk factors cannot be appropriately neglected. As a remedy, we require a random effect term, known as frailty in the survival analysis context, to account for this survival heterogeneity during the modelling process.

According to Aalen (1994), there are three possible sources of this unobserved heterogeneity. The first one is attributed to the differences in biological "fitness" from the start of the period of observation. This can be explained by the fact that certain subjects or patients may have weaker organ functions compared to other cohort members or have a genetic predisposition that renders them more susceptible to the disease of interest. Secondly, the unobserved heterogeneity may also be caused by the stresses (both biological and psychological stressors) experienced by those subjects in their lifetime. This therefore portrays frailty as a dynamic process which changes as time progresses. The final cause is due to whether the event of interest occurs at an early or late phase. For example, a

protracted diagnosis of cancer may cause the cancer cells to be ubiquitously metastasised all over the bodily systems. Consequently, such a person will be in an advanced stage of cancer and this may cause him or her to have a higher frailty of death compared to his or her cohort.

Before we proceed further, it is imperative to distinguish between the frailty term used in a medical context and the frailty term used in biostatistics and demography. In the former, frailty means the susceptibility of an individual to mortality or morbidity associated with the disease of interest (Wienke, 2011). On the other hand, the statistical frailty means random effects accounting for the unobserved heterogeneity between subjects (Wienke, 2011). It is thus paramount not to confuse between these two types of frailty.

Another aspect that should be emphasised is that the frailty term only accounts for unobserved heterogeneity between subjects, which is indeed a simplified concept in addressing the differences in survival times among individuals. For example, the concepts of adaptation and debilitation are not captured by the frailty term (Wienke, 2011). In the former, individuals are considered to be successfully adapting themselves to the stress (i.e. the event of interest) after a period of exposure to the stress. Debilitation, on the other hand, refers to the worsening of individuals' stress response or physical health after being exposed to the stress for a period of time. Hence, this is a fertile ground for future research.

In the next section, we shall illustrate the utility of the frailty term in, firstly, univariate survival data, followed by multivariate survival data. These are then followed by the formal derivations of conditional and marginal cumulative hazard, hazard, survival, probability density functions in the presence of a frailty term in the survival model.

2.6.2 The motivation behind the inclusion of a frailty term

The frailty term in the univariate survival model is useful for making the model more flexible. For instance, we can incorporate a frailty term with a proportional hazard model. Contingent upon the type of distribution assigned to the frailty term, the assumption of hazard proportionality can be relaxed and thus the model can now accommodate the non-proportionality of hazard. Apart from that, the inclusion of a frailty term may also rectify a survival model with an overdispersed error. We can also assign each subject his or her own frailty and since frailty modifies the baseline hazard function multiplicatively, each individual will thus have their own hazard. To obtain the population-level hazard,

we can simply integrate out the frailty term and thus averaging the frailty over individuals in the population (Wienke, 2011).

Neglecting the frailty term may have undesirable consequences. According to Henderson and Oman (1999), forsaking the frailty term from the model may cause misspecification of a proportional hazard model to the correct marginal distribution. Henderson and Oman (1999) also found that estimates of regression coefficients obtained might be biased towards 0, depending upon the type of the frailty distribution and the level of the unobserved heterogeneity operating on the individual subjects. For more examples of frailty in univariate survival models, Hougaard (1995) provides extensive examples of this.

On the other hand, frailty is more frequently used in multivariate survival models for the purpose of modelling dependencies between observations (Hougaard, 2013). There are four common scenarios where frailty is useful. The first one is when there is strong correlation between two or more individuals, for instance twins, siblings or matched pairs. The second is when modelling the effects of an intervention on related organs in an individual or components in a machine. Thirdly, frailty is useful in modelling the dependencies of recurrent or multiple events, for example when investigating the time to cerebrovascular accident (i.e. stroke), heart attack or seizure associated with epilepsy in the same individual. The final one is when an individual experiences multiple treatments in a designed experiment and the time to an event of interest for each treatment regime is recorded. For further information, refer to Hougaard (2013).

2.6.3 The derivations of conditional and marginal quantities of interest

Before we proceed, we shall assume that the frailty term acts on the baseline hazard function multiplicatively. Borrowing the notation used in Collett (2015) and Fung (2017), we denote a frailty term as z_i. In a proportional hazard model, the frailty term can be incorporated as

$$h_i(t \mid Z_i = z_i) = \ z_i \exp(\boldsymbol{\beta}^T \ \boldsymbol{x}_i) \ h_o(t).$$

The equation above gives the hazard function at t given the value of the frailty term z_i for individual i. We can take log on both sides of the equation to obtain

$$\log h_i(t \mid Z_i) = \log(z_i) + \boldsymbol{\beta}^T \boldsymbol{x_i} + \log h_0(t).$$

There is one particular assumption that we made here. In contrast to the effects of other covariates on survival time (i.e that they may change over time, hence possibly requiring time-varying covariate effects), the effect of the frailty term is considered constant over the time period. It means that the random risk experienced by the individuals will remain constant as time progresses. This may or may not be true and hence there is room for future research to develop a survival model that is not strictly dependent on this assumption.

From here, we will proceed to deriving other quantities. To obtain the conditional cumulative hazard, $H_i(t \mid Z_i)$, we integrate the hazard with respect to t, giving

$$H_i(t \mid Z_i) = \int_0^\infty h_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x_i})\ Z_i dt = H_0(t) \exp(\boldsymbol{\beta}^T \boldsymbol{x_i})\ Z_i.$$

Next, based on the relation

$$S(t) = \exp\Big\{ - \mathrm{H(t)} \Big\},$$

we derive the conditional survival function as

$$\begin{aligned} S_i(t \mid z_i) &= \exp\Big\{ - \mathrm{H}_i(\mathrm{t} \mid Z_i) \Big\} \\ &= \exp\Big\{ - \mathrm{H}_0(\mathrm{t}) \lambda_i Z_i \Big\} \end{aligned}$$

where $\lambda_i = \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)$. Then, by simple algebra, the conditional survival function is obtained as

$$\begin{aligned} S_i(t) \mid Z_i &= \exp\Big\{ - H_0(t) \Big\}^{\lambda_i Z_i} \\ &= \{S_0(t)\}^{\lambda_i Z_i} \end{aligned}$$

where $S_0(\mathrm{t})$ represents the baseline survival function. We can obtain the marginal survival function by multiplying the baseline survival function with $f_Z(z_i)$ and integrating out the z_i term to obtain the marginal survival function:

$$S_i(t) = \int_0^\infty S_0(t)^{\lambda_i z_i} f_Z(z_i) dz_i = \mathrm{E}_Z[S_0(t)^{\lambda_i Z_i}]$$

where $f_z(z_i)$ is the pdf of Z.

Since $f_i(t) = -\frac{d}{dt}S(t)$, we can now derive the marginal survival time pdf:

$$f_i(t) = \int_0^\infty \lambda_i z_i S_0(t)^{\lambda_i z_i - 1} S_0'(t) f_Z(z_i) dz_i$$

$$= f_0(t) \times \mathrm{E}_Z\Big\{\lambda_i Z_i S_0(t)^{\lambda_i Z_i - 1}\Big\}$$

and hazard function

$$h_i(t) = \frac{f_i(t)}{S_i(t)}$$

$$= h_0(t) S_0(t) \frac{\mathrm{E}_Z\big\{\lambda_i Z_i S_0(t)^{\lambda_i z_i - 1}\big\}}{\mathrm{E}_Z\big\{S_0(t)^{\lambda_i Z_i}\big\}}$$

$$= h_0(t) \frac{\mathrm{E}_Z\big\{\lambda_i Z_i S_0(t)^{\lambda_i z_i}\big\}}{\mathrm{E}_Z\big\{S_0(t)^{\lambda_i Z_i}\big\}}. \tag{2.2}$$

As a simple example, we shall now elucidate gamma-distributed frailties with a constant baseline hazard. In this example, we assume that each subject possesses a separate value for the frailty term, z_i, and these values are iid with $Z_i \sim$ Ga (g,h_*). We also assume $S_0(t) = \exp\{-t\}$, so the conditional lifetimes have an exponential distribution. Using (2.2), we obtain the hazard function as

$$h_i(t) = h_0(t) \frac{E_Z\{\lambda_i Z_i \exp(-\lambda_i Z_i t)\}}{E_Z\{\exp(-\lambda_i Z_i t)\}}.$$

Since

$$E_Z\{\exp\,(-\lambda_i Z_i t)\} = \int_0^\infty f_Z(z_i)[S_0(t_i)]^{\lambda_i Z_i} dZ_i$$

we obtain

$$E_Z\{\exp\,(-\lambda_i Z_i t)\} = \frac{h_*^g}{(\lambda_i t + h_*)^g}$$

and

$$E_Z\{\lambda_i Z_i \exp\,(-\lambda_i Z_i t)\} = \frac{\lambda_i g h_*^g}{(h_* + \lambda_i t)^{g+1}}.$$

Hence, the marginal hazard function is given by

$$h_i(t) = h_0 \frac{g\lambda_i}{\lambda_i t + h_*}$$

and the hazard ratio of an individual i to individual j is

$$\frac{h_i(t)}{h_j(t)} = \frac{\lambda_i(\lambda_j t + h_*)}{\lambda_j(\lambda_i t + h_*)}$$

which is not constant with respect to t unless $\lambda_i = \lambda_j$. The mean and variance of the gamma distribution of z_i are given by

$$m = \frac{g}{h_*} \quad \text{and} \quad \upsilon = \frac{g}{h_*^2} = \frac{m}{h_*}.$$

With $m = 1$, we have

$$h_* = \frac{1}{\upsilon}.$$

Thus, we can rewrite the hazard ratio as

$$\frac{h_i(t)}{h_j(t)} = \frac{\lambda_i(\upsilon\lambda_j t + 1)}{\lambda_j(\upsilon\lambda_i t + 1)}. \tag{2.3}$$

As we can see, (2.3) indicates the non-proportionality of hazards since the hazard ratio changes with t.

2.7 Piecewise constant hazard model

2.7.1 Theory and basic notation

A flexible approach to relax the parametric assumption about the form of the baseline hazard is a piecewise constant hazard model (Prairie and Ostle, 1961; Colvert and Boardman, 1976; Shaked, 1979; West, 1982; West et al., 1985; Gamerman and West, 1987; Gamerman, 1991; Ibrahim et al., 2001; Kim et al., 2007; Wilson and Farrow, 2017). We shall describe the piecewise-constant proportional hazard (PCPH) models first before elaborating on the alternative modelling approach which is needed when the proportional hazard assumption is violated. In PCPH models, we assume that the hazard is variable over different time intervals, albeit it is still proportional. In this case, the time axis is divided into disjoint intervals by specifying several cut-off points.

We choose some fixed time points $\tau_0,\ldots,\tau_j$, with $\tau_0 = 0$, $\tau_j > \tau_{j-1}$ and, typically, $\tau_j \to \infty$. Then the j^{th} time interval is $I_j = [\tau_{j-1}, \tau_j)$. Next, we assume that the hazard is constant within each time-interval, but variable between time-intervals. The hazard

function can thus be represented as follows:

$$h_i(t) = \lambda_{ij}, \qquad \text{for} \;\; \tau_{j-1} \leq t < \tau_j$$

where $\tau_0 = 0$ and $\tau_j \to \infty$

The covariates for individual i can be linked to λ_{ij} by the relationship

$$\lambda_{ij} = \boldsymbol{\lambda}_{0j}\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}_{0j} + \boldsymbol{x_i^T}\boldsymbol{\beta}).$$

The integrated hazard is thus given by

$$\begin{aligned} H_i(t) &= \int_0^t h_i(u)\ \mathrm{du} \\ &= \sum_{k=1}^{j-1} \int_{\tau_{k-1}}^{\tau_k} h_{ik}(t)\ \mathrm{d}t + \int_{\tau_{j-1}}^{\mathrm{t}} h_{ij}(t)\ \mathrm{d}t \\ &= \sum_{k=1}^{j-1} \int_{\tau_{k-1}}^{\tau_k} \lambda_{ik}\ \mathrm{d}t + \int_{\tau_{j-1}}^{\mathrm{t}} \lambda_{ij}\ \mathrm{d}t \\ &= \sum_{k=1}^{j-1} \lambda_{ik}(\tau_k - \tau_{k-1}) + \lambda_{ij}(t - \tau_{j-1}), \quad t \in I_j. \end{aligned}$$

Since $S_i(t) = \exp\{-H(t)\}$, the survivor function for the piecewise constant hazard model is given

$$S_i(t) = \exp\left\{ - \left[\sum_{k=1}^{j-1} \lambda_{ik}(\tau_k - \tau_{k-1}) + \lambda_{ij}(t - \tau_{j-1}) \right] \right\}, \quad t \in I_j.$$

Hence, the probability density function of survival times is given by

$$f_i(t) = h_i(t)\ S_i(t) = \left\{ \prod_{k:\tau_k<t} \Big[\exp\{-\lambda_{ik}\ (\tau_k - \tau_{k-1})\} \Big] \right\} \lambda_{ij}\ \Big[\exp\{-\lambda_{ij}\ (t - \tau_{j-1})\} \Big]$$

for $t \in I_j$. Therefore, the conditional density and survivor function given that $T \geq \tau_{j-1}$ are

$$f_i(t \mid T \geq \tau_{j-1}) = \lambda_{ij}\ \exp\{-\lambda_{ij}\ (t - \tau_{j-1})\}$$

and

$$S_i(t \mid T \geq \tau_{j-1}) = \exp\{-\lambda_{ij}\ (t - \tau_{j-1})\}, \qquad \text{for } \tau_{j-1} \leq t < \tau_j.$$

If we consider the likelihood for the event time, t, we shall realise that the likelihood can be factorised temporally. Based on Gamerman (1991), the likelihood for the event time can be written as

$$\mathrm{L}(\boldsymbol{t} \mid D_0, \boldsymbol{\lambda}) = \prod_{i=1}^{n} \prod_{j=1}^{r} \mathrm{L}_j(t_{ij} \mid \boldsymbol{\lambda}_j, D_{j-1})$$

where D_0 is an information set containing the event status for all subjects across all j intervals, D_{j-1} is the event status for subjects who survive into the previous interval I_{j-1} and t_{ij} represents one of these three occurrences for each subject i in interval j:

- the i^{th} subject experienced the event in the interval j. Hence, $\tau_{j-1} \leq t_i < \tau_j$. So, $t_{ij} = t_i$ and $\delta_{ij} = 1$
- the i^{th} subject was censored in the interval j. Therefore, $\tau_{j-1} \leq t_i < \tau_j$. So, $t_{ij} = t_i$ and $\delta_{ij} = 0$
- the i^{th} subject was still alive at the end of the interval j. Therefore, $\tau_j \leq t_i$. So, $t_{ij} = \tau_j$ and $\delta_{ij} = 0$
- the i^{th} subject died or was censored before the beginning of interval j. Then $t_{ij} = \tau_{j-1}$ and $\delta_{ij} = 0$

where δ_{ij} indicates whether the subject experienced the event of interest in I_j. Each $L_j(t_{ij} \mid D_0, \lambda)$ is thus given by:

$$\mathrm{L}_j(t_{ij} \mid \boldsymbol{\lambda}_j, D_{j-1}) = \Big\{\lambda_{ij}\Big\}^{\delta_{ij}} \exp\{-\lambda_{ij}(t_{ij} - \tau_{j-1})\} \tag{2.4}$$

where $\delta_{ij} = 1$ if the subject dies in I_j and $\delta_{ij} = 0$ otherwise.

2.7.2 Methods of selecting the cut points for the time-intervals in a piecewise constant hazard model

One of the requirements for the piecewise constant hazard model is that the time axis should be partitioned. Several authors have recommended that the time intervals are

between the start and end of the event times. However, Kalbfleisch and Prentice (1973) proposed that the time intervals should be independently chosen without looking at or being guided by the observed data. West and Berliner (1992) recommended that the time intervals should be shorter in the beginning of the follow-up time for cancer datasets since there is a higher number of deaths in the beginning than in the later time period. To simplify things, we can attempt to choose the cut points so that a similar number of events occur in every time interval. Nevertheless, caution is recommended since this approach will not be suitable for all cases because events may occur in unequal proportions in each time interval.

As an example, suppose we choose to have ten time intervals to model a cancer dataset. Hence, we can aim to have about 10% of events occurring in each time interval. As an approximation, we may assume that the event times are distributed exponentially. If $\tau_1 < \tau_2 < \ldots < \tau_{J-1}$ are the cut points and the exponential survival function with parameter λ is given by $\mathrm{S(t)} = \exp\{-\lambda t\}$, the probability that a subject survives until time τ_j is given by

$$\exp\{-\lambda\tau_j\} = 1 - 0.1j.$$

This can be rearranged giving

$$-\lambda\tau_j = \log(1 - 0.1) \qquad \text{and} \qquad \tau_j = -\frac{1}{\lambda}\log(1 - 0.1).$$

The mean of the exponential distribution is $\nu = \frac{1}{\lambda}$. Let $\xi = \frac{1}{J}$. Then the final form of the equation will be

$$\tau_j = -\nu\log(1 - j\xi).$$

This requires a prior evaluation of the mean lifetime. For instance, by applying the equation above and based upon the evaluation of our prior beliefs for the mean survival time for a typical patient with Non-Hodgkin Lymphoma (which is approximately about 3 years), we chose the cut points in Table 2.1 to obtain 10 survival time intervals.

Table 2.1: Nine cut points for constructing 10 survival time intervals for piecewise constant hazard model for Scotland Newcastle Lymphoma Group (SNLG) data set

j	1	2	3	4	5	6	7	8	9
τ_j	0.316	0.669	1.070	1.532	2.079	2.749	3.612	4.828	6.908

On the other hand, for a typical advanced non-small cell lung cancer (NSCLC) patient, we assume the mean survival time is 13 months based on our prior belief. In this case, we chose smaller number of time intervals since ten survival time intervals will result in a time interval without event (i.e the time interval will only contain all censored observations). The cut points are given in Table 2.2.

Table 2.2: Six cut points for constructing seven survival time intervals for piecewise constant hazard model for advanced non-small cell lung cancer data set

j	1	2	3	4	5	6
τ_j	2.004	4.374	7.275	11.015	16.286	25.297

Each time interval is now assumed to have a separate baseline hazard. Those who survive in one interval will enter the subsequent interval with a different baseline hazard.

2.7.3 Time-varying covariate effects

Despite the proportional hazard model being the most popular approach used for fitting time-to-event data, it has several pitfalls. The most conspicuous drawback is its overdependence on the proportional hazard assumption since this means that the hazard ratio does not change over time. This assumption is sometimes unrealistic since we can imagine real-life scenarios where this assumption is not tenable. For instance, let us consider a clinical trial comparing the efficacy of bone marrow transplant (BMT) against the standard chemotherapy regimen for acute myeloid leukaemia (AML) patients. In the earlier part of the trial, AML patients who have undergone the curative BMT will have a higher hazard of dying immediately after the BMT than the recipients of the standard chemotherapy regimen. This observation can be attributed to the susceptibility of AML patients for bacterial or viral infections secondary to immunosuppressive drugs that they received to prevent BMT rejection. However, if they survived these early and serious complications of BMT, their hazard of dying will start to decline at the later period of follow up before it settles at a lower level than in those who received the standard chemotherapy regimen due to the curative nature of BMT.

Another example to illustrate the non-proportionality of hazard is the hazard of dying for non-Hodgkin lymphoma (NHL) patients who received the Rituximab, Cyclophosphamide, Hydroxydaunorubicin, Vincristine and Prednisone (R-CHOP) regimen which is higher than that for NHL patients who received best supportive care (BSC) at the beginning of the follow-up period in a clinical trial due to the side effects of the intensive

chemotherapy regimes that they received to induce Non-Hodgkin Lymphoma remission. However, as the time progresses, the R-CHOP NHL recipients will have lower hazard of dying than those NHL patients who are treated with BSC since the R-CHOP recipients are no longer on chemotherapy and the disease activity in this group of NHL patients is now in complete remission. Consequently, the proportional hazard assumption is too idealistic to be indiscriminately applied to the modelling of all survival data.

To relax the proportional hazard assumption, we can make the coefficients of covariates vary over time and this can be represented as follows:

$$\lambda_i(t) = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}(t))$$

where $\boldsymbol{\beta}(t) = (\beta_{1j}, \ldots, \beta_{pj})^T$, for t $\in I_j$.

Besides, other alternatives have been proposed to remedy this. The most straightforward solution is to incorporate interaction terms between follow-up time and covariates in the regression equation. One of the approaches that uses this methodology is the flexible parametric models. These employ restricted cubic splines for the modelling of transformed survival functions (Royston and Parmar, 2002). However, the simple inclusion of interaction terms between the spline function and time-dependent covariates is a rather crude and less flexible technique since the pattern of changes in the function of the regression coefficients is restrictively specified. Hence other more versatile and complicated functions such as quadratic function cannot be fully implemented via the flexible parametric modelling approach. Alternatively, the nonparametric smoothing spline approach (He et al., 2010) and the piecewise constant hazard model (Kalbfleisch, 1978; Gamerman, 1991) could also be used. Nevertheless, they also have their own disadvantages, for instance the arbitrary selections of change points for the piecewise constant hazard model and the lack of convergence of the MCMC sampler when greater degrees of the spline basis function are used (He et al., 2010).

2.8 Bayesian inference: theory and notation

The goal of Bayesian inference is to quantify uncertainty about parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ using observed data $\mathbf{y}$. Suppose $\mathbf{y}$ is modelled by some probability density function

$f_y(\boldsymbol{y} \mid \boldsymbol{\theta})$. Then the likelihood is defined as

$$\mathrm{L}(\boldsymbol{\theta} \mid \mathbf{y}) = \prod_{i=1}^{n} f_i(y_i \mid \boldsymbol{\theta}).$$

if $y_1, \ldots, y_n$ are conditionally independent given $\boldsymbol{\theta}$. The likelihood represents the probability or probability density of the data $\mathbf{y}$ as a function of the parameters $\boldsymbol{\theta}$. Prior beliefs about $\boldsymbol{\theta}$ are represented by the density $\pi(\boldsymbol{\theta})$ and Bayes' Theorem provides a way of updating these beliefs based upon observed data. The posterior density is therefore

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{\pi(\theta)L(\theta \mid \boldsymbol{y})}{\int\limits_{\theta} \pi(\theta)L(\theta \mid \boldsymbol{y})\mathrm{d}\theta} \tag{2.5}$$

which represents the updated beliefs about $\boldsymbol{\theta}$ after observing the data $\mathbf{y}$. Since the denominator of (2.5) is not a function of $\boldsymbol{\theta}$, it can be regarded as a constant of proportionality and hence

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{\theta}) \times \mathrm{L}(\boldsymbol{\theta} \mid \boldsymbol{y})$$

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}.$$

2.9 A brief introduction to MCMC and RJAGS

2.9.1 Fundamental elements of Markov Chain Monte Carlo (MCMC) in Bayesian settings

Sometimes, it can be difficult to evaluate the integral in the denominator in (2.4) or other integrals such as posterior expectations. Hence, numerical methods, such as Markov Chain Monte Carlo (MCMC), are required. MCMC is an approach used to draw samples from a target distribution by simulating from a specially constructed Markov chain with stationary distribution equal to the target distribution $\pi(.)$, Thus, providing that the chain has converged, any value sampled will be from the density of interest $\pi(.)$, here the joint posterior density. Additionally, for a multidimensional chain, samples of each component will be drawn from the marginal distribution of the respective component. Let us assume that the distribution of interest is the posterior distribution, with density $\pi(\theta \mid D)$ (known as the target distribution), where $D = (y_1, y_2, \ldots, y_n)$. Here, we discuss two algorithms to construct these chains, specifically the Metropolis-Hastings algorithm

and the Gibbs sampler.

2.9.2 The Metropolis-Hastings (MH) algorithm

Metropolis et al. (1953) introduced the algorithm which was generalised by Hastings (1970), leading to the name Metropolis-Hastings. Central to Metropolis-Hastings is the idea of a proposal density, denoted q(.|.), which is some (arbitrary) transition kernel. It can be advantageous to have a proposal density which is easy to sample from. The Metropolis-Hastings algorithm is shown as Algorithm 1.

Step 2 generates a new value of the chain from the proposal density $\mathrm{q}(\theta^* \mid \theta)$, which in step 4 is either accepted (the chain moves) or rejected (the chain remains where it was). Note that $\pi(\theta \mid D)$ enters the acceptance probability as a ratio, and hence it is only necessary to know $\pi(\theta \mid D)$ up to a constant of proportionality. Therefore, by Bayes theorem, A (in the acceptance probability of step 3) can be expressed as

$$A = \frac{\pi(\theta^*)\ \mathrm{L}(\theta^* \mid D)\ q(\theta \mid \theta^*)}{\pi(\theta)\ \mathrm{L}(\theta \mid D)\ q(\theta^* \mid \theta)}.$$

Given that we have complete freedom in the choice of the proposal density q(.|.), the natural question is, 'What choices of q(.|.) might be good, or indeed useful?' In particular, a good choice of q(.|.) will lead to a chain which converges rapidly and mixes well; that is, it moves often and well around the support of $\pi(\theta$ |D). We present the first algorithm for the Metropolis-Hastings (MH) procedure as follows:

Algorithm 1: The Metropolis-Hastings Algorithm

---

1. Initialise the iteration counter i =1 and initialise the chain with $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)^T$ where $\theta^{(0)}$ is chosen from somewhere in the support of $\pi(\theta \mid D)$.

2. Propose a new value θ^* using the transition kernel $\mathrm{q}(\theta^* \mid \theta^{(i-1)})$

3. Evaluate the acceptance probability min(1,A), where,

$$A = \frac{\pi(\theta^* \mid D)\ \mathrm{q}(\theta \mid \theta^*)}{\pi(\theta \mid D)\ \mathrm{q}(\theta^* \mid \theta)}$$

4. Set $\theta^{(i)} = \theta^*$ with probability min(1,A), otherwise set $\theta^{(i)} = \theta^{(i-1)}$.

5. Set counter to i+1 and return to step 2.

2.9.2.1 Symmetric proposal

Suppose that we use a proposal distribution which is symmetric. That is

$$\mathrm{q}(\theta^* \mid \theta) = \mathrm{q}(\theta \mid \theta^*), \qquad \forall \theta, \theta^*$$

and the proposal distribution can be, for example a Gaussian distribution

$$\theta^* \mid \theta \sim N(\theta, \sigma^2).$$

In this instance, A simplifies to

$$A = \frac{\pi(\theta^* \mid D)}{\pi(\theta \mid D)}.$$

That is, the acceptance probability does not depend on the proposal density.

2.9.2.2 Random walk Metropolis

It is possible to use a random walk as the proposal distribution q(.|.) in step 2 of Algorithm 1. In this instance, q(.|.) takes the form

$$\theta^* = \theta^{(t-1)} + \omega_i$$

where $\omega_1, \ldots, \omega_p$ are independent and in many cases identically distributed. Typically ω has a Gaussian distribution with zero mean vector. In this instance the Metropolis-Hastings algorithm is known as a random walk sampler (or random walk Metropolis).

The variance of the random variates ω will determine the mixing of the chain. Too low a variance and the chain will explore the space slowly, but many proposed values will be accepted. Too large a variance and few proposed values will be accepted. Reflecting the correlation within θ in the covariance structure of ω is an important aspect in ensuring the chain efficiently explores the space.

If the target distribution is Gaussian, Roberts and Rosenthal (2001) suggest that the optimal acceptance probability is 0.234. Sherlock and Roberts (2009) extend this result to ellipitically symmetric targets and subsequently Sherlock (2013) proposes a general set of sufficient conditions for which the optimal acceptance probability is 0.234. Roberts

et al. (1997) and Roberts and Rosenthal (2001) recommend that the variance of ω should be given by

$$\frac{2.38^2 \text{Var}(\theta \mid D)}{p}.$$

where p is the number of dimension and Var $(\theta \mid D)$ is the variance matrix of the target distribution $\pi(\theta \mid \text{D})$. Typically though, $\text{Var}(\theta \mid D)$ will not be available and hence an estimate from one or more pilot runs should be used.

We note that, for large p, sampling θ^* from a multivariate Gaussian distribution may be expensive. In these instances, an alternative approach is to take the components of $\omega = (\omega_1, \ldots, \omega_p)^T$ as independent and identically distributed (univariate) Gaussian random variates. For example, $\omega_i \sim \text{N}(0, \sigma_i^2)$, where

$$\sigma_i^2 = \frac{2.38^2 \ \text{Var}(\theta_i \mid D)}{p}.$$

2.9.2.3 Independence sampler

An independence sampler is a special case of a Markov chain. As the name suggests, an independence sampler (or independent chain) proposes a new value θ^* independently of the current value θ. Hence, $\text{q}(\theta^* \mid \theta) = \text{g}(\theta^*)$ for some density g(.). Whilst the form of such a proposal may appear to disagree with the Markovian structure of the chain, both θ and θ^* feature in the acceptance probability, which means that a proposal still depends upon the current state, and thus, the Markov property is preserved. Using such a proposal distribution leads to an acceptance probability $\min(1, A)$, where

$$A = \frac{\pi(\theta^* \mid D)}{\pi(\theta \mid D)} \Bigg/ \frac{\text{g}(\theta^*)}{\text{g}(\theta)}.$$

Clearly, we can increase the acceptance probability by making g(.) and $\pi(. \mid D)$ as similar as possible. It is worth noting that, in the context of an independent sampler (and in contradiction to the above on random walk Metropolis), the higher the acceptance probability, the better the sampling is. Tierney (1994) suggests the avoidance of densities g(.) with thin tails.

2.9.3 Gibbs sampling

2.9.3.1 General Algorithm

The Gibbs sampler (or generically Gibbs sampling) originated in statistical physics before it came into Bayesian statistics via image processing (Geman and Geman, 1984). It was introduced by Geman and Geman (1984) before being generalised and brought to the interest of the wider statistical community by Gelfand and Smith (1990). In essence, the Gibbs sampler is an MCMC scheme in which the full conditional distributions (FCDs) are used to form the transitional kernel.

Before we proceed further, we shall firstly define the FCD. Let θ_c be a component of a parameter space $\boldsymbol{\theta}$. Hence, the FCD of θ_c is the distribution of θ_c given other components of $\boldsymbol{\theta}$ and data $\mathbf{x}$. Mathematically, this is written as

$$\pi(\theta_c \mid \{\theta_q : q \neq c\}, \boldsymbol{x}).$$

For a Gibbs sampler, we assume that the FCDs for all components of $\boldsymbol{\theta}$ are available and can be sampled from. The Gibbs sampler is then given by Algorithm 2:

Algorithm 2: The Gibbs sampler

---

1. Initialise the iteration counter i =1 and initialise the chain with $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)^T$

2. Gain a new value $\theta^{(i)}=\left(\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_p^{(i)}\right)^T$ from $\theta^{(i-1)}$ using successive generation from the full conditional distributions:

$$\theta_1^{(i)} \sim \pi\left(\theta_1 \mid \theta_2^{(i-1)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right)$$

$$\theta_2^{(i)} \sim \pi\left(\theta_2 \mid \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right)$$

.

.

.

$$\theta_p^{(i)} \sim \pi\left(\theta_p \mid \theta_2^{(i)}, \theta_3^{(i)}, \ldots, \theta_{p-1}^{(i)}, D\right)$$

3. Set iteration counter to i+1 and return to step 2.

The chain approaches its equilibrium state as the number of iterations increases, and once the chain has converged, a value of θ^i is a sample from $\pi(\theta \mid D)$. Thus the Gibbs sampler is a way to sample from $\pi(\theta \mid D)$ when direct sampling is costly, complicated or indeed impossible, but sampling from $\pi(\theta_i \mid \theta_{-i}, D)$ is possible. Algorithm 2 is known as a fixed sweep Gibbs sampler. Whilst other versions of the Gibbs sampler are available, such as the random scan Gibbs sampler, the fixed sweep is simple to implement, and thus appealing. For details of other versions of the Gibbs sampler see, for example, Chapter 5 of Gamerman and Lopes (2006).

2.9.3.2 Blocking

Given that the components of $\boldsymbol{\theta}$ can take the form of scalars, vectors or matrices, it can be useful to block certain components that are strongly correlated in the posterior together in multidimensional problems. Such a strategy is known as a block update and makes use of multivariate sampling techniques. Blocking is a strategy used to improve the convergence (and indeed mixing) of the chain, although it can come at a higher computational cost. As discussed in Gamerman and Lopes (2006) it is not the case that the larger the block update, the faster the convergence. Indeed for highly multidimensional problems a large block update is likely to be highly detrimental. Instead, components of $\boldsymbol{\theta}$ should be blocked together such that the correlations between the blocks is low. Any conditionally independent components should be updated on their own (a single-block update).

2.9.3.3 Component-wise transitions (Metropolis within Gibbs Algorithm)

In practice, the construction of a suitable proposal density in a Metropolis-Hastings scheme could be difficult. However, for many problems of interest, it may be possible to sample from the full conditional distributions for a subset of $\boldsymbol{\theta}$. Let the full conditional distribution for the i^{th} component of $\boldsymbol{\theta}$ be denoted by

$$\pi(\theta_i \mid \theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p, D) = \pi(\theta_i \mid \theta_{-i}, D), \qquad i = 1, \ldots, p.$$

The algorithm for component-wise transitions is given by Algorithm 3 below. Note that Algorithm 3 is actually just a special case of Algorithm 1 (Dellaportas and Roberts, 2003). Besides, Algorithm 2 is a special case of Algorithm 3.

Algorithm 3: The Metropolis-Hastings Algorithm: Component-wise Transitions

1. Initialise the iteration counter i =1 and initialise the chain with $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}\right)^T$.

2. Obtain a new value $\theta^{(i)} = \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_p^{(i)}$ from $\theta^{(i-1)}$ using successive generation from distributions:

   $\theta_1^{(i)} \sim \pi\left(\theta_1 \mid \theta_2^{(i-1)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right)$ using a Metropolis-Hastings step with proposal $q_1\left(\theta_1^* \mid \theta_1^{(i-1)}\right)$
   $\theta_2^{(i)} \sim \pi\left(\theta_2 \mid \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_p^{(i-1)}, D\right)$ using a Metropolis-Hastings step with proposal $q_2\left(\theta_2^* \mid \theta_2^{(i-1)}\right)$
   .
   .
   .
   $\theta_p^{(i)} \sim \pi\left(\theta_p \mid \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{p-1}^{(i-1)}, D\right)$ using a Metropolis-Hastings step with proposal $q_p\left(\theta_p^* \mid \theta_p^{(i-1)}\right)$

3. Set i+1 and return to step 2

If the full conditional distribution for the i$^{\text{th}}$ component of θ is available to sample from directly, the resulting acceptance probability is 1. This method is also referred to as *Metropolis-within-Gibbs*. If the full conditional distributions are completely known and can be sampled from for all components of θ, we obtain the Gibbs sampler.

2.9.3.4 Analysing MCMC output

As mentioned above, a Markov chain Monte Carlo scheme will only give samples from the target distribution provided convergence has been reached. It is therefore important to monitor convergence carefully and ensure convergence truly has been reached. As the number of iterations increases the distribution of the chain, $\theta^{(i)} \mid D$, tends to the posterior distribution $\theta \mid D$, and convergence is reached. Samples obtained before convergence, when the distribution of the chain is not the posterior are discarded. This number of

iterations is known as the burn-in period. Viewing the trajectory of the chain via a trace plot can be used to check convergence informally. In this instance we are looking for the chain to display the same qualitative behaviour after some initial burn-in period. Gelfand and Smith (1990) (amongst others) suggest a number of informal checks for convergence. More formal checks for ensuring convergence has been reached have been proposed by, for example, Heidelberger and Welch (1983) and Raftery and Lewis (1992).

Samples of the MCMC scheme will be dependent, meaning successive draws are autocorrelated. Autocorrelation at different lag times can be observed via an autocorrelation (ACF) plot (Geyer, 2011). If samples are highly correlated then the chains may be thinned. This involves retaining the sampled values from iterations $m, 2m, 3m, \ldots$, where m is an integer and $m > 1$. Thinning retains only n/m sampled values, resulting in an increase in Monte Carlo variance. Nevertheless, the increase in Monte Carlo variance might be small if there is a positive autocorrelation. Thinning is useful for evaluating convergence, as a remedy to storage space problems and to reduce overall computational cost if expensive computations are to be done on the sampled values (Geyer, 1992).

Once a chain has converged, the (suitably thinned) output can be analysed. It is effortless to compute estimates of summary statistics (or standard statistical measures) such as marginal means and variances. Joint and marginal distributions can be viewed through the use of density plots (or histograms).

2.10 Missing Data and Data Augmentation Approach

2.10.1 Missing Covariate Values

The analysis of survival data with missing covariate values will have serious consequences. The posterior distributions may possess dissimilar properties if the analyses based on complete cases only (i.e cases without missing covariate values) are compared with those performed on all cases (i.e cases with missing covariate values are also included in the analyses).

The general framework of missing data mechanisms is given by Little and Rubin (2002) and they are as follows:

Missing completely at random (MCAR): Missingness is independent of the observed and unobserved data. For example, we can divide the full data in a matrix $\mathbf{x}$ into

two components: the observed component which is denoted by $\boldsymbol{x}_{observed}$ and $\boldsymbol{x}_{missing}$ which represents the missing component. Using this missing data framework, the conditional distribution of the missing data indicator can be written as:

$$f(\boldsymbol{m}_i \mid \boldsymbol{x}_{observed}, \boldsymbol{x}_{missing}, \phi) = f(\boldsymbol{m}_i \mid \phi)$$

where $\boldsymbol{m}_i$ is the missing data indicator, $\boldsymbol{m}_i = (m_{i,1}, \ldots, m_{i,J})^T$ where $\boldsymbol{m}_{ij} = 0$ if $\boldsymbol{x}_{ij}$ is observed and $\boldsymbol{m}_{ij} = 1$ if $\mathbf{x}$ is missing, ϕ is the parameter for the missing data mechanism and x_{ij} is the value of the j^{th} covariate for the i^{th} individual.

Missing at random (MAR): Missingness is conditionally independent of the unobserved component given the observed data. The conditional distribution of the missingness indicator can be written as:

$$f(\boldsymbol{m}_i \mid \boldsymbol{x}_{observed}, \boldsymbol{x}_{missing}, \phi) = f(\boldsymbol{m}_i \mid \boldsymbol{x}_{observed}, \phi)$$

Since our inference about parameters of interest is only conditional upon $\boldsymbol{x}_{observed}$, MAR has an ignorable missing data mechanism provided that ϕ and θ are independent in the prior, where θ is the collection of parameters of the data model.

Missing not at random (MNAR): Missingness is not conditionally independent of $\boldsymbol{x}_{missing}$ given $\boldsymbol{x}_{observed}$. This is a non-ignorable missing data mechanism that occurs when the failure of observing the data values is dependent upon the missing data.

In survival analysis, the missing covariate data can be represented using the missing data framework as follows. Let $\boldsymbol{X}^*$ represent the covariate matrix and $x_{i,j}$ is the value of covariate j for the i^{th} subject. The survival time is represented by vector $\boldsymbol{T}$, where $\boldsymbol{T} = (t_1, t_2, \ldots, t_n)^T$. Either $\boldsymbol{X}^*$ or $\boldsymbol{T}$ is observed entirely (i.e. all subjects experienced the event of interest (e.g death) by the end of the study, no missing values for all covariates) or incompletely observed (i.e. several subjects did not experience the event of interest (e.g death) and hence were censored by the end of a clinical trial, missing values for some of the covariates). The missing indicator matrix is represented by $\boldsymbol{M}$ whose elements $m_{i,j}$ denote the missing data indicator for covariate j for the i^{th} patient, where $m_{i,j} = 0$ if the value for covariate j is fully observed or $m_{i,j} = 1$ if the value is missing. The model parameters are represented by θ and ϕ designates the parameter for missing data mechanism. Hence, the joint distribution of a collection of $\boldsymbol{X}^*$, $\boldsymbol{M}, \boldsymbol{T}$ conditional upon

the θ and ϕ can be written as:

$$f(\boldsymbol{X}^*, \boldsymbol{M}, \boldsymbol{T} \mid \boldsymbol{\theta}, \phi) = f(\boldsymbol{X}^*, \boldsymbol{T} \mid \boldsymbol{\theta}) f(\boldsymbol{M} \mid \boldsymbol{X}^*, \boldsymbol{T}, \phi).$$

If the covariate values are considered to be MCAR or MAR, the missing data models are thus ignorable. In this research work, the missing data will be treated as MAR and we shall assume $\boldsymbol{\theta}$ and ϕ are independent *a priori*, and hence $\pi(\boldsymbol{\theta}, \phi) = \pi(\boldsymbol{\theta})\pi(\phi)$). Our approach for missing covariate data will hence be a fully Bayesian one and this involves specifying the prior distributions for all parameters as well as specifying the distributions for the missing data. There are several examples for this approach and one of them is that demonstrated by Ibrahim et al. (2001) for forward conditioning and another is reverse conditioning (Zhao, 2010).

2.10.2 Data augmentation

Certain models possess rather complicated likelihood functions which, if handled in a direct fashion, would not be amenable to tractable algebraic manipulation. It is thus occasionally feasible to make the likelihood more straightforward when extra variables, known as auxiliary variables, are introduced. These variables could be considered as latent observations, which, if they were actual observations, would result in the likelihood being more manageable. These auxiliary variables could therefore be regarded as missing data. This is known as the data augmentation approach (Tanner and Wong, 1987). MCMC methods such as the Gibbs sampler are appropriate for this technique in handling missing data in general.

We illustrate the use of the data augmentation approach for handling missing survival data in censored observations. This approach makes the likelihood simpler since we assume that all quantities are observed. In this approach, an auxiliary variable (denoted by $\boldsymbol{T}_{missing}$) was introduced to represent the parameter by which the survival times for the censored observations is generated (i.e. missing survival times). If we decompose $\boldsymbol{T}$ (all survival times) into $\boldsymbol{T}_{observed}$ and $\boldsymbol{T}_{missing}$, we can then compute the posterior distribution of model parameters (denoted by $\boldsymbol{\theta}$) given the observed survival time $\boldsymbol{T}$ by integrating out $\boldsymbol{T}_{missing}$ as follows:

$$\Pr(\boldsymbol{\theta} \mid \boldsymbol{T}) = \int \Pr(\boldsymbol{\theta}, \boldsymbol{T}_{missing} \mid \boldsymbol{T}) d\boldsymbol{T}_{missing}$$

According to Bayes theorem, Pr $(\boldsymbol{\theta}, \boldsymbol{T}_{missing} \mid \boldsymbol{T}_{observed})$ is given by:

$$\Pr(\boldsymbol{\theta}, \boldsymbol{T}_{missing} \mid \boldsymbol{T}_{observed}) \propto \mathrm{L}(\boldsymbol{\theta}, \boldsymbol{T}_{missing}) \Pr(\boldsymbol{\theta}, \boldsymbol{T}_{missing}).$$

The likelihood can then be written as

$$f_{T,I}(\boldsymbol{T}_{observed}, \boldsymbol{I} \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \int f_T(\boldsymbol{T}_{observed}, \boldsymbol{T}_{missing} \mid \boldsymbol{\theta}) f(\boldsymbol{I} \mid \boldsymbol{T}_{observed}, \boldsymbol{T}_{missing}, \boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{T}_{missing}$$

This is the likelihood of the data conditioned upon the augmented parameters and I is the inclusion indicator where $I_i = 1$ if $T_{observed}$ and $I_i = 0$ if $T_{missing}$ in individuals with censored survival times. Besides, $f(\boldsymbol{I} \mid \boldsymbol{T}_{observed}, \boldsymbol{T}_{missing}, \boldsymbol{\vartheta})$ is known as the missing data mechanism. If we assume that the priors for $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ are independent, then $\Pr(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = \Pr(\boldsymbol{\vartheta}) \Pr(\boldsymbol{\theta})$. Hence, in this case the missing data mechanism is ignorable.

We can also use data augmentation approach to solve the problem with missing covariate data. This is implemented by iteratively sampling the conditional distributions of the missing covariate information, $\boldsymbol{X}_{unobserved}$ given the parameters $\boldsymbol{\theta}$ and observed data $\boldsymbol{X}_{observed}$ and subsequently the conditional distributions of $\boldsymbol{\theta}$ given the augmented data, $\boldsymbol{X}_{augmented} = (\boldsymbol{X}_{observed}, \boldsymbol{X}_{unobserved})$. In this case, $\boldsymbol{X}_{augmented}$ can be considered as an augmented data set containing both the observed values and the sampled values for the missing covariate information, $\boldsymbol{X}_{unobserved}$. For each iteration k, the sampling scheme is as follows:

$$\boldsymbol{X}^{(k)}_{unobserved} \sim \pi(\boldsymbol{X}_{unobserved} \mid \boldsymbol{X}_{observed}, \boldsymbol{\theta}^{(k-1)})$$

$$\boldsymbol{\theta}^{(k)} \sim \pi(\boldsymbol{\theta} \mid \boldsymbol{X}_{observed}, \boldsymbol{X}^{(k)}_{unobserved})$$

This approach is highly beneficial in the sense that it enables the sampling of $\boldsymbol{\theta}$ when they could not be sampled by conditioning them on $\boldsymbol{X}_{observed}$ only. Besides, we may regard that this algorithm is a special case of Gibbs sampling procedure whereby we sample the joint posterior π $(\boldsymbol{\theta}, \boldsymbol{X}_{unobserved} \mid \boldsymbol{X}_{observed})$ from their FCDs. We can then use the sampled values of $\boldsymbol{\theta}$ as if they are drawn from the marginal posterior, π $(\boldsymbol{\theta} \mid \boldsymbol{X}_{observed})$ and these can be used to compute the posterior mean and variance of $\boldsymbol{\theta}$.

2.11 Bayesian survival analysis computation using RJAGS

In this short section, the fitting of an exponential survival model using a package called Just Another Gibbs Sampler (JAGS) (Plummer, 2016) which is implemented in an R environment as RJAGS (R Core Team, 2018) is used as an example. The kidney infection dataset (Nahman et al., 1992) will be used as an example and this dataset consists of survival time (time to kidney infection) and types of catheter insertion (either inserted through surgery or under the patient's skin). Let the model specification be given as follows. We have

$$T_i \sim \exp(\lambda_i),$$

where for patient i ($i = 1, \ldots, 119$), T_i is the time to the first catheter exit-site infection (in days). The pdf of the survival time and the corresponding survival function are thus given by

$$f(t_i \mid \lambda_i) = \lambda \exp(-\lambda_i t_i)$$

and

$$S(t_i \mid \lambda_i) = \exp(-\lambda_i t_i)$$

There is only one covariate, catheter placement status for each patient i (denoted by x_{1i}, and coded as -1 for surgically-inserted catheters and 1 for catheters inserted under the patient's skin), in this example. Hence, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, where β_0 is the intercept and β_1 is the coefficient for catheter placement (covariate trt). Therefore, $\lambda_i = \exp(\beta_0 + x_{1i}\beta_1)$. The likelihood can then be written as

$$L(\boldsymbol{\beta} \mid D) = \prod_{i=1}^{119} \lambda_i^{\delta_i} \exp\left\{ -\sum_{i=1}^{119} \lambda_i t_i \right\}$$

where $\lambda_i = \exp\{\beta_0 + \beta_1 x_{1i}\}$.

For this case, we assume normal priors for $\boldsymbol{\beta}$ with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. For illustration, let

$$\beta_0 \sim N(0, 20)$$

$$\beta_1 \sim N(0, 20)$$

Hence, the posterior distribution for $\boldsymbol{\beta}$ can be written as

$$\pi(\boldsymbol{\beta} \mid D) \propto L(\boldsymbol{\beta} \mid D)\pi(\boldsymbol{\beta} \mid \boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$$

The RJAGS code for this model is as follows:

```
model
{
   for(i in 1:119)
       {is.censored[i]~dinterval(t[i],t.cen[i])
       t[i]~dexp(lambda[i])
       lambda[i]=exp(beta0 + beta.trt * x[i])
       }
   beta0~dnorm(0.0,0.05)
   beta.trt~dnorm(0.0,0.05)
}
```

In `rjags`, dnorm (m, p) denotes a normal distribution with mean m and precision p. The data is stored in a data frame named `data`. The data frame contains the following

- `data$status`: the censoring or event status (1 if the event occurs and 0 if censoring occurs)
- `data$time`: the censoring or event time
- `x`: the covariate. in this case is the type of catheter placement received by the patients (1:surgically, 2:percutaneously)

In this analysis, the dataset used by `rjags` contains the following vectors

- `t`: For this vector, the element i represents the time to event if the event was observed for case i. If case i, on the other hand, is censored, then the element i of `t` is `NA`. We can construct it as follows:

  ```
  t=data$t
  is.na(t)=data$status==0
  ```

- `is.censored`: This is a difference in terms of rjags code for survival analysis when compared to other standard R survival and R-INLA packages. The survival indicator should be coded as follows: 1 if the time is censored and 0 if the event time is observed. This can be easily performed using the following

```
is.censored=1-data$status
```

- `t.cen`: Element i is the censoring time for case i if this case was censored. If the event was observed in case i, this censoring time should be set at a time greater than the observed event time. This can be done as follows:

```
t.cen=data$t+data$status
```

- `trt`: A covariate.

```
x=data$trt
```

As we can see from the model specification, a special distribution was allocated to `is.censored`.

```
is.censored[i]~dinterval[t[i],t.cen[i])
```

The missing values of `t` should be initialised. The reason for this is that the specification of `is.censored` means that these cases are the censored ones. Hence, the unobserved event time must be greater than the observed censoring time. If `rjags` is permitted to select the initial values randomly, this may result in some of the chosen initial values for the unobserved event times being smaller than the censoring time and this is inconsistent, causing errors in `rjags` command execution. Therefore, we have to initialise the missing values of `t` at greater values than the corresponding values of `t.cen`. For instance, the two sets of initial values for two parallel chains can be constructed as follows:

```
tinits=data$t+5
is.na(tinits1)=data$status==1
tinits2=tinits1+5
```

We can now form the data and initial values list and run the MCMC algorithm in the following fashion:

```
#1 kidneydata=list(t=t,tcen=t.cen,x=x,is.censored=is.censored)
#2 kidneyinits=list(list(t=tinits1),list(t=tinits2))
#3 kidneyjags=jags.model("exsurv.txt",data=kidneydata,inits=kidneyinits,
n.chains=2)
#4 update(kidneyjags,5000)
#5 kidneysamples=coda.samples(kidneyjags,c("beta0,"beta.trt"),100000)
```

The command in the 4^{th} line is for burn in and the last line command is for collecting the samples from the posterior distributions for each model parameter. Using this code, rjags will treat the censored event times as auxiliary variables and sample them. However, for this exponential case, the likelihood could, of course, be easily and exactly written and posterior sampling could be done without resorting to data augmentation.

This example is discussed in Section 7.2.1.2, where results are given.

2.12 Summary

In this chapter, we discussed the fundamentals of survival analysis in 2.1 - 2.3. In 2.4, the methods of relating the explanatory covariates with the the lifetime distribution were elucidated. In 2.5.1, the parametric Weibull survival models were introduced and further elaborated. The motivations behind including the frailty effects in survival analysis were explained in 2.6, together with the derivations of conditional and marginal survival and hazard functions. The piecewise constant hazard model was introduced in 2.7 and the theory of Bayesian inference was explored in 2.8. A brief introduction to MCMC was elaborated in detail in 2.9, followed by a discussion on missing data and data augmentation in the context of Bayesian survival analysis in 2.10. Finally, the implementation of an MCMC scheme in RJAGS was demonstrated in 2.11.

Chapter 3

Latent Gaussian Models & Gaussian Markov Random Fields

3.1 Introduction

The purpose of this chapter is to present the basic ideas and known results for latent Gaussian models and Gaussian Markov random field (GMRF). We shall firstly discuss the fundamentals of generalised linear models (GLM) and their extension, the latent Gaussian models. We shall then elaborate on the standard results of Gaussian Markov Random Fields (GMRF) and their relationship with Bayesian hierarchical models. The paramount ideas discussed in this chapter will be implemented in later chapters and hence this chapter is indispensable when elucidating the contents in subsequent chapters.

3.2 Generalised Linear Models and Latent Gaussian Models (LGMs)

Latent Gaussian models (LGMs) are popular in statistical modelling due to their simplicity and flexibility. LGMs treat observations, denoted here as $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, as depending on covariates (patient's clinical characteristics, spatial locations etc). In LGMs, observations are considered to be conditionally independent given a set of latent variables and these observations have distributions based on the data types (e.g. Bernoulli for binary data or Poisson for count data). These latent variables are then assigned a Gaussian process (GP) prior, which can be defined as an ensemble of a finite number

of random variables which follows a multivariate normal distribution (Rasmussen, 2004). This implies marginal normality due to the properties of marginal distributions of multivariate Gaussian distributions. Then, the covariance structure of the latent variables is parameterised by a set of hyperparameters which specifies the covariance function. Hence, LGMs can be visualised to be hierarchical in nature, with hyperparameters conditioning the latent variables, which in turn condition the observations.

LGMs can also be regarded as a type of Bayesian structured additive models with a structured additive predictor, η_i. Before we proceed further, the much simpler generalised linear models (GLM) shall be first described. There are two constituents in a GLM: the conditional distribution of the outcome variable (Y) should be one of the members of the exponential family and a link function should be specified to relate the mean of Y with the linear combination of predictors, $X_1, X_2, \ldots, X_n$(Dobson and Barnett, 2008; Faraway, 2016). For the distribution of Y to be considered as belonging to the exponential family of distribution, its density should be expressible in the following form:

$$f(y \mid \theta, \phi) = \exp\Big[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \theta)\Big]$$

where ϕ is the dispersion parameter and θ constitutes the canonical parameter. As an example, let Y be a Gaussian-distributed random variable. Hence the pdf of Y is

$$f(y \mid \theta, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big[-\frac{(y-\mu)^2}{2\sigma^2}\Big].$$

The pdf can then be rewritten in the following form:

$$f(\boldsymbol{y} \mid \boldsymbol{\theta\phi}) = \exp\Big[\frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2}\Big\{\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\Big\}\Big].$$

We can then relate the mean of this distribution, E[Y]=μ to the linear predictors, $X_1, X_2, \ldots, X_n$ using the link function g(.):

$$g(.) = \eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \tag{3.1}$$

Some other types of link function are presented in Table 3.1. In the modelling approach used in this thesis, however, we shall not restrict the distribution of Y to belong to the exponential family.

We can relax the assumption of a known form of linear function in the linear predictor. The framework for this model can be conceptualised as follows. The mean μ_i of this

Table 3.1: The link functions for selected members of the exponential family of distributions used in the GLM (adapted from Faraway (2016))

Distributional Family	Link Function	Variance Function
Normal	$\eta = \mu$	1
Poisson	$\eta = \log(\mu)$	μ
Binomial	$\eta = \log(\mu/(1-\mu))$	$\mu/(1-\mu)$
Gamma	$\eta = \mu^{-1}$	μ^2
Inverse Gaussian	$\eta = \mu^{-2}$	μ^3

observation variable is connected to the structured additive predictor variables, η_i through a link function, g(.), so that $g(\mu_i) = \eta_i$. This can be represented as

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{j=1}^{n_f} f^{(j)}(u_{ij}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i. \tag{3.2}$$

In (3.2), $\beta_1, \ldots, \beta_{n_\beta}$ constitute the linear effects of covariates $\mathbf{z}$, $f^{(.)}, \ldots, f^{n_f}$ are the unknown functions of covariates $\mathbf{u}$ which can be: i) non-linear effects of continuous covariates, ii) time trends and seasonal effects, iii) frailty effects, iv) spatial random effects, v) i.i.d random intercepts and slopes (Martino et al., 2011). The structured or unstructured random errors are $\epsilon_1, \ldots, \epsilon_n$. Then, a Gaussian prior is allotted to each $\beta_0, \ldots, \beta_{n_\beta}$ and $f^{(j)}$ with precision matrix $\mathrm{Q}(\boldsymbol{\theta}_1)$ where $\boldsymbol{\theta}_1$ is a vector of hyperparameters of the precision matrix for the latent field $\mathbf{x}$. In other word

$$\mathbf{x} \mid \boldsymbol{\theta}_1 \sim N\ (\mathbf{0},\ \mathbf{Q^{-1}}(\boldsymbol{\theta_1})).$$

Let $\mathbf{x}$ be the vector of all the Gaussian-distributed quantities such as β_0, β_1,...,β_{np}, ϵ_1,...,ϵ_n, and $f^{(0)}$,...,$f^{(n_f)}$ whilst $\boldsymbol{\theta}_1$ represents the vector of hyperparameters which are not necessarily Gaussian.

The multivariate distribution of the observed $\mathbf{y}$ (i.e. $\mathbf{y} = \{y_i : i \in \mathcal{I}\}$) is also dependent on another set of hyperparameters $\boldsymbol{\theta}_2$ (i.e. conditional on another set of hyperparameters

that should not be confused with $\boldsymbol{\theta}_1$). Hence, the likelihood is given by

$$\pi(\mathbf{y} \mid \mathbf{x}, \theta_{\mathbf{2}}) = \prod_{i=1}^{n} \pi(y_i \mid x_i, \theta_{\mathbf{2}}).$$

Therefore, the joint posterior density of the latent field $\mathbf{x}$ and hyperparameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ given the observed data $\mathbf{y}$ is

$$\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\ \pi(\mathbf{x} \mid \boldsymbol{\theta})\ \prod_{i=1}^{n} \pi(y_i \mid x_i, \boldsymbol{\theta}).$$

Exact inference for the parameters of a LGM is usually intractable analytically since the likelihood of the observations, $\prod_{i=1}^{n} \pi(y_i \mid x_i, \boldsymbol{\theta})$, is usually not the conjugate of GP priors for the latent variables and neither can the hyperparameters be easily integrated out. Attempts using deterministic approaches such as Laplace Approximation (LA) (Tierney and Kadane, 1986) or refined Expectation Propagation (EP) (Cseke, 2011) have been suggested to address such a problem. Rue et al. (2009) proposed the Integrated Nested Laplace Approximation (INLA) as an approximate inference method for LGM. Basically, INLA uses numerical quadrature based on a gridding scheme that can be applied to problems with a relatively small set of hyperparameters and this method can be used to find the posterior marginal distributions of the latent Gaussian field, $\pi(x_i \mid \boldsymbol{y})$, and the hyperparameters, $\pi(\theta_k \mid \boldsymbol{y})$.

In the following subsections, further discussions will be made on Gaussian Markov Random Fields (GMRF) which are a special case of Gaussian Random Fields. The Markov assumption enables INLA to work efficiently but it is not strictly necessary since a departure from this assumption is not an absolute hindrance for INLA to function satisfactorily (Eidsvik et al., 2009). We will present a brief overview of INLA methodologies as described in Rue et al. (2009) in Chapter 4.

3.3 Gaussian Markov Random Field (GMRF)

3.3.1 Introduction

The efficiency of INLA is dependent on the sparsity of the precision matrix. Therefore, the sparse property of Gaussian Markov random fields (GMRFs) is essential to ensure INLA efficiency since GMRFs allow fast and efficient computation of posterior summaries using

numerical methods specific for sparse matrices (e.g. Cholesky factorisation). Therefore, in this section, we give a discussion of GMRFs which will facilitate the understanding of INLA methodology in subsequent chapters.

A Gaussian Random Field (GRF) is simply a finite-dimensional random vector, $\mathbf{X}$= $(x_1,x_2,\ldots,x_n)^T$, that follows a multivariate normal distribution. A Gaussian Markov Random Field (GMRF), on the other hand, is a special case of GRF in which we have an adjacency relationship between elements of the vector $\mathbf{X}$, such that two elements x_i, x_j may or may not be neighbours, and an assumption of conditional independence between non-adjacent elements given the elements in between.

The simple but powerful concept of conditional independence can be seen from the following example. Let $\mathbf{X} = (X_1, X_2, X_3)^T$ be a random vector. If X_1 and X_2 are conditionally independent given X_3, then once we know the value of X_3, knowing X_2 gives us no further information about X_1. To get a clearer picture, let the joint density of $\mathbf{x}$ be

$$\pi(\boldsymbol{x}) = \pi(x_1 \mid x_2, x_3)\pi(x_1 \mid x_3)\pi(x_3).$$

If X_1 and X_2 are conditionally independent given X_3,then

$$\pi(\boldsymbol{x}) = \pi(x_1 \mid x_3)\pi(x_2 \mid x_3)\pi(x_3).$$

To illustrate further, let us consider a first-order autoregressive (AR(1)) process with standard Gaussian errors which can be expressed as follows:

$$X_t = \phi X_{t-1} + \epsilon_t,\ \epsilon_t \sim N(0,1).$$

The conditional independence assumption is not explicitly shown here. However, by expressing the equation as:

$$X_t \mid X_1, \ldots, X_{t-1} \sim N(\phi X_{t-1}, 1)$$

with t = 2,. . . , n, we see that X_t and X_s (with $1 \leq$ s$<$ t $\leq$ n) are conditionally independent given $\{x_{s+1},\ldots,\ x_{t-1}\}$ provided t - s $>$1. If $\mid \phi \mid < 1$ and the marginal distribution of X_1 is Gaussian with mean 0 and variance $\sigma^2 = 1\ /\ (1 - \phi^2)$, then the process is stationary and this marginal distribution is the stationary distribution for the process. Hence the joint pdf of $\mathbf{X}$ can be expressed as follows:

$$\pi(\boldsymbol{x}) = \pi(x_1)\pi(x_2 \mid x_1)\ldots\pi(x_n \mid x_{n-1})$$

$$= \frac{1}{(2\pi)^{n/2}} \mid \boldsymbol{Q} \mid^{1/2} \exp\Big(-\frac{1}{2} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} \Big)$$

where, in this case, the precision matrix $\mathbf{Q}$ is tridiagonal:

$$\boldsymbol{Q} = \begin{pmatrix} 1 & -\phi & & & & & & & \\ -\phi & 1+\phi^2 & -\phi & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & \ddots & \ddots & \ddots & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & & -\phi & 1+\phi^2 & -\phi \\ & & & & & & & -\phi & 1 \end{pmatrix}$$

From the matrix $\mathbf{Q}$ above, we can clearly see that the zero entries occur outside the diagonal and first off-diagonal elements. This tridiagonal feature of this matrix is due to the conditional independence assumption: X_i and X_j are conditionally independent if $\mid i - j \mid > 1$ given the rest of X (i.e. $\boldsymbol{X}_{-ij}$).

3.3.2 Fundamental characteristics of GMRFs

Consider a random vector $\mathbf{X}$ which has a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. An undirected graph $\mathcal{G}$ can be defined as a tuple of ν and ξ; where ν represents the set of nodes from 1 to n in the undirected graph and ξ are the edges. We define that there is an absence of an edge between node i and j if and only if $X_i \perp X_j \mid \boldsymbol{X}_{-ij}$ ($\perp$ denotes conditional independence, $\boldsymbol{X}_{-ij}$ represents other nodes than node i or node j). Hence, $\mathbf{X}$ has a GMRF property with respect to $\mathcal{G}$. Let Q $= \Sigma^{-1}$ and i $\neq$ j, then

$$X_i \perp X_j \mid X_{-ij} \qquad \Longleftrightarrow \qquad Q_{ij} = 0$$

where Q_{ij} is the i, j element of Q.

The following results are also true if $\mathbf{X}$ is a GMRF with regard to $\mathcal{G}$ and has mean $\boldsymbol{\mu}$ $= (\mu_1, \ldots, \mu_r)^T$ and precision matrix, $\mathbf{Q}$. The conditional expectation of X_i given $\boldsymbol{X}_{-i}$ is

$$\mathrm{E}(X_i \mid \boldsymbol{X}_{-i}) \;=\; \mu_i - \frac{1}{Q_{ii}} \sum_{j:j\sim i} Q_{ij}(x_j - \mu_i).$$

The conditional precision of X_i given $\boldsymbol{X}_{-i}$ is

$$P \;\; (X_i \mid \boldsymbol{X}_{-i}) \;=\; Q_{ii}.$$

The conditional correlation of X_i and X_j given $\boldsymbol{X}_{-ij}$

$$\mathrm{Corr}\;(X_i, X_j \mid \boldsymbol{X}_{-ij}) \;=\; \frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \qquad i \neq j$$

where j $\sim$ i indicates that node i and node j are neighbours. The proofs are given in section 3.3.3. Hence, the diagonal elements of precision matrix, $\mathbf{Q}$ are the conditional precisions of X_i given $\boldsymbol{X}_{-i}$ and the elements which are located off-diagonally are the conditional covariance between X_i and X_j given $\boldsymbol{X}_{-ij}$ (Rue and Held, 2005). The following Markov properties of GMRF also hold:

The pairwise Markov Property:

$$x_i \perp x_j \mid \mathbf{x_{-ij}}, \qquad \text{if} \quad \{i,j\} \notin \xi \quad \text{and} \quad i \neq j$$

The local Markov Property:

$$x_i \perp x_{-(i,ne(i))} \mid \boldsymbol{x}_{ne(i)}, \qquad i \in \mathcal{V}$$

The global Markov Property:

$$\mathbf{x_A} \perp \mathbf{x_B} \mid \mathbf{x_C}$$

where $\boldsymbol{x}_{ne(i)}$ are the elements in the neighbourhood of i, A,B and C are disjoint non-empty sets in which C separates A and B, $\mathcal{V}$ is the set of nodes in the graph and ξ is the set of edges $\{i, j\}$.

3.3.3 The proof for the properties of the precision matrix Q

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ be distributed as a Gaussian Markov random field. Let $\delta_{ij} = \delta_{ji} = 1$ if X_i is a neighbour of X_j and $\delta_{ij} = \delta_{ji} = 0$ otherwise. The joint pdf is proportional to

$$e^{-\frac{1}{2}S} \tag{3.3}$$

where S = $(X - \mu)^T$ Q $(X - \mu)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$. We can also write the joint pdf in the form

$$f_1(x_1) \prod_{i=2}^{n} f_i(x_i \mid x_1, \ldots, x_{i-1})$$

Because this is a linear Gaussian process, we can write

$$S = \sum_{i=1}^{n} \frac{e_i^2}{\nu_i} \quad \text{where} \quad f_1(x_1) \propto e^{-\frac{e_1^2}{2\nu_1}}$$

and

$$f_i(x_i \mid x_1, \ldots, x_{i-1}) \propto e^{-\frac{e_i^2}{2\nu_i}}$$

where ν_1 is the marginal variance of X_1 and ν_i $(i > 1)$ is the conditional variance of $X_i \mid x_1, \ldots, x_{i-1}$. The last element, x_n, only appears in $f_n(x_n \mid x_1, \ldots, x_{n-1})$ and therefore in e_n. Using the Markov property that X_n is conditionally independent of $\overline{N}_n$ given N_n, where $N_n = \Big\{x_i : \ i < n \text{ and } \delta_{in} = 1\Big\}$ and $\overline{N}_n\Big\{x_i : \ i < n \text{ and } \delta_{in} = 0\Big\}$, we can write

$$e_n = x_n - \mu_n - \sum_{i=1}^{n-1} \delta_{in}\phi_{in}(x_i - \mu_i) \tag{3.4}$$

for some coefficients ϕ_{in}.

Sparsity

Let the i, j element of Q be q_{ij}. From (3.3) and (3.4), we see that $q_{ni} = q_{in} = \delta_{in}\ \phi_{in}/\sqrt{\nu_n}$, for i = 1,. . . ,n-1 and $q_{nm} = \nu_n$, and so the last row and column of Q only contain nonzero elements corresponding to the neighbours of X_n. Of course, by rearranging the order of

$X_1,\ldots,X_n$, we can choose any to be the last.

Conditional Expectation

Now,

$$\begin{aligned}\mathrm{E}(X_n \mid \boldsymbol{X}_{-n}) &= \mu_n + \sum_{i=1}^{n-1} \delta_{in}\phi_{in}(x_i - \mu_i) \\ &= \mu_n - V_n \sum_{i=1}^{n-1} q_{in}(x_i - \mu_i) \\ &= \mu_n - \frac{1}{q_{nn}} \sum_{i=1}^{n-1} q_{in}(x_i - \mu_i).\end{aligned}$$

Conditional Precision

We have

$$\mathrm{Var}(X_n \mid \boldsymbol{X}_{-n}) = v_n = q_{nn}^{-1}.$$

Hence, the conditional precision is

$$P(X_n \mid \boldsymbol{X}_{-n}) = q_{nn}.$$

Conditional Correlation

The last two elements X_{n-1} and X_n, only appear in $f_{n,n-1}(X_n, X_{n-1} \mid X_1,\ldots,X_{n-2})$ and $f_{n-1}(X_{n-1} \mid X_1,\ldots,X_{n-2})$ and hence in the joint density

$$f_{n,n-1}(X_n, X_{n-1} \mid X_1,\ldots,X_{n-2}) \propto \exp\big\{-\frac{1}{2}\widetilde{s}\big\}.$$

Now

$$\widetilde{S} = \frac{1}{1-p^2}\Big[\frac{\widetilde{e}_n^2}{\widetilde{v}_n} + \frac{e_{n-1}^2}{v_{n-1}} - 2\rho\frac{\widetilde{e}_n^2 e_{n-1}}{\sqrt{\widetilde{v}_n v_{n-1}}}\Big]$$

where $\widetilde{v}_n$ is the conditional variance of X_n given $X_1,\ldots,X_{n-2}$, p is the conditional correlation of X_n and X_{n-1} given $X_1,\ldots,X_{n-2}$

$$e_{n-1} = x_{n-1} - \mu_{n-1} - \sum_{i=1}^{n-2} \delta_{i,n-1}\phi_{i,n-1}(x_i - \mu_i),$$

and

$$\widetilde{e}_{n-1} = x_{n-1} - \mu_{n-1} - \sum_{i=1}^{n-2} \delta_{i,n-2}\widetilde{\phi}_{i,n}(x_i - \mu_i),$$

for some coefficients $\widetilde{\phi}_{1,n}, \ldots, \widetilde{\phi}_{n-2,n}$. We can also write $\widetilde{S}$ as

$$\widetilde{s} = (x_n - \mu_n)^2 q_{nn} + 2(x_n - \mu_n)(x_{n-1} - \mu_{n-1})q_{n,n-1} + (x_{n-1} - \mu_{n-1})^2 q_{n-1,n-1} + R$$

where R contains terms not involving x_n or x_{n-1}. Thus

$$q_{nn}^{-1} = (1-p^2)\widetilde{v}_n, q_{n-1,n-1}^{-1} = (1-p^2)v_{n-1}$$

and

$$q_{n,n-1}^{-1} = -p^{-1}(1-p^2)\sqrt{\widetilde{v}_n v_{n-1}}.$$

Combining these, we obtain

$$p = -\frac{q_{n,n-1}}{\sqrt{q_{n,n}\ q_{n-1,n-1}}}.$$

3.3.4 Deriving the recursive formula for marginal variance computation

We can rearrange the order of the elements of $\mathbf{X}$ and reorder the rows and columns of $\mathbf{Q}$ accordingly to make $\mathbf{Q}$ sparse. By considering what happens if we choose any element to be the last element, we see that, for all i =1,. . . ,n and j = 1,. . . ,n, $q_{ij} = 0$ if $\delta_{ij} = 0$. In other words, $\mathbf{Q}$ for any GMRF will have the following property: $\boldsymbol{Q}_{ij} = 0$ for i $\neq$ j if x_i and x_j are conditionally independent given the other elements $\{x_k$: k$\neq$ i and k $\neq$ j$\}$. This results in the sparsity of the structure of $\mathbf{Q}$, making it convenient for fast computation via Cholesky factorisation of $\mathbf{Q}$

$$\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^T$$

where $\mathbf{L}$ is a lower triangular matrix. The beauty of this approach is that $\mathbf{L}$ also inherits the sparsity of $\mathbf{Q}$ since it also has global Markov characteristics. Consequently, the computing time becomes more efficient since we do not have to compute and store the zero terms. Hence, only the non-zero terms in $\mathbf{L}$ are computed and stored during the procedures for posterior computation. This is thus the rationale behind the fast computation of posterior summaries in the case of a GMRF.

To illustrate this further, let us consider a GMRF with a considerable size of 10000

to 100000 nodes. By constructing a band matrix $\mathbf{Q}$ via permuting the indices of observation x_i, the $\mathbf{L}$ matrix will be acquired. In this case, we classically use a band-Cholesky factorisation method on $\mathbf{Q}$ to obtain $\mathbf{L}$, which is a lower band matrix that possesses similar bandwidth to $\mathbf{Q}$ (Rue and Held, 2005). As an example, let us consider an MCMC procedure for obtaining posterior samples from a complicated hierarchical model that possesses a GMRF structure with mean value of zero and precision matrix, $\mathbf{Q}$. We can do this by working out the Cholesky factor $\mathbf{L}$ for $\mathbf{Q}$ and thus subsequently obtaining $\mathbf{z}$ which is a vector of independent standard Gaussian variables. The relationship between $\mathbf{L}$, $\mathbf{x}$ and $\mathbf{z}$ is given by

$$\mathbf{L^T x} = \mathbf{z}. \tag{3.5}$$

Specifically, (3.5) can be written (element-wise) as

$$L_{ii}x_i = z_i - \sum_{r=i+1}^{n} L_{ri}x_r. \tag{3.6}$$

To obtain the marginal variance $\sum_{ij}$, we multiply both sides of (3.6) by x_j (j$\geq$ i) and divide by L_{ii} which results in

$$x_i x_j = z_i x_j / L_{ii} - 1/L_{ii} \sum_{r=i+1}^{n} L_{ri}x_r x_j. \tag{3.7}$$

Since $L_{ii}x_j = z_j$ and taking expectation, (3.7) can be written as

$$\Sigma_{ij} = \delta_{ij}/L_{ii}^2 - 1/L_{ii} \sum_{r=i+1}^{n} L_{ri}\Sigma_{rj}, \quad j \geq i, \quad i = n, \ldots, 1 \tag{3.8}$$

with $\delta_{ij} = z_i z_j$. In this case, $\delta_{ij} = 1$ if i = j, and 0 if i $\neq$ j. In the case where we are only concerned with obtaining the marginal variance, only Σ_{ij} whose L_{ij} is not 0 should be determined using the recursive formula (3.8). As a result, the computing costs to perform this algorithm are efficiently reduced to $O(n)$, $O(n^{3/2})$ and $O(n^2)$ for temporal, spatial and spatiotemporal GMRFs, respectively (Rue and Held, 2005).

3.3.5 The derivation of GMRF distribution conditional on a linear constraint

To derive the distribution of a GMRF with an additional linear constraint such as $\mathbf{Ax} = \mathbf{e}$ (i.e from $\pi(\mathbf{x} \mid \mathbf{Ax}=\mathbf{e})$ and $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$, we can start with the joint Gaussian distribution of $\mathbf{x}$ and $\mathbf{Ax}=\mathbf{e}$. In this case, $\mathbf{A}$ is a k $\times$ n matrix and of rank k, and $\mathbf{e}$ is a k $\times$ 1 matrix. The mean and covariance of this distribution are

$$\mathrm{E}\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{Ax} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{A\mu} \end{pmatrix} \quad \text{and} \quad \mathrm{Cov}\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{Ax} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}^{-1} & \boldsymbol{Q}^{-1}\boldsymbol{A}^T \\ \boldsymbol{AQ}^{-1} & \boldsymbol{AQ}^{-1}\boldsymbol{A}^T. \end{pmatrix}$$

Using the formula for computing conditional expectation ($\mu^* = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$) and conditional covariance ($\Sigma^* = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$), $\widetilde{\boldsymbol{\mu}} = \mathrm{E}(\boldsymbol{x} \mid \boldsymbol{Ax})$ can be derived as

$$\begin{aligned} \widetilde{\boldsymbol{\mu}} &= \boldsymbol{\mu_x} + \boldsymbol{Q^{-1}A^T}(\boldsymbol{AQ^{-1}A^T})^{-1}(\boldsymbol{e} - \boldsymbol{A\mu}) \\ &= \boldsymbol{\mu_x} - \boldsymbol{Q^{-1}A^T}(\boldsymbol{AQ^{-1}A^T})^{-1}(\boldsymbol{A\mu} - \boldsymbol{e}) \end{aligned} \tag{3.9}$$

and $\widetilde{\boldsymbol{\Sigma}} = \mathrm{Cov}(\boldsymbol{x} \mid \boldsymbol{Ax})$ as

$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{Q^{-1}} - \boldsymbol{Q^{-1}A^T}(\boldsymbol{AQ^{-1}A^T})^{-1}\boldsymbol{AQ^{-1}}. \tag{3.10}$$

To take full advantage of the sparsity of Q matrix, we can then use a geostatistical method called conditioning by Kriging (Ren et al., 2005). Using this method, samples from the unconstrained GMRF, $\mathbf{x} \sim \mathrm{N}(\hat{\boldsymbol{\mu}}, \boldsymbol{Q}^{-1})$, are initially obtained. We can then obtain samples from $\pi(\boldsymbol{x} \mid \boldsymbol{Ax} = \boldsymbol{e})$ (denoted by $\widetilde{\boldsymbol{x}}$) by

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{Q^{-1}A^T}(\boldsymbol{AQ^{-1}A^T})^{-1}(\boldsymbol{Ax} - \boldsymbol{e}). \tag{3.11}$$

It is evident that $\widetilde{\boldsymbol{x}}$ has the correct conditional distribution if we compare its mean and covariance with (3.9) and (3.10). Using algorithm 2.6 as elaborated in Rue and Held (2005), samples from $\pi(\boldsymbol{x} \mid \boldsymbol{Ax} = \boldsymbol{e})$ can thus be efficiently acquired.

3.3.6 GMRF and Bayesian hierarchical models

GMRFs are commonly used for Bayesian hierarchical models since they allow convenient specification of stochastic dependence between unknown parameters. A common GMRF

framework for this model type has a three-stage layout: hyperparameters $\boldsymbol{\theta}$ specifying a GMRF $\mathbf{X}$, the field $\mathbf{X}$ is linked to data $\mathbf{Y}$ which are usually assumed to be conditionally independent with one another given the field $\mathbf{X}$, and finally the priors for the hyperparameters $\boldsymbol{\theta}$ (in this case, $\boldsymbol{\theta}$ represents a collection of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$). Using the most trivial example, each Y_i is only dependent on the corresponding i^{th} element in $\mathbf{x}$, such that both $\mathbf{y}$ and $\mathbf{x}$ possess the same number of dimensions. These three stages can be written as

$$\begin{aligned} \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}) \quad \text{hyperparameter stage} \\ \boldsymbol{X} \mid \boldsymbol{\theta_1} &\sim N(\mathbf{0}, \boldsymbol{Q^{-1}}(\boldsymbol{\theta_1})) \quad \text{latent field stage} \\ \boldsymbol{Y} \mid \boldsymbol{x},\ \boldsymbol{\theta_2} &\sim Ga(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_2) \quad \text{data stage} \end{aligned}$$

with data-stage density

$$\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_2) = \prod_{i=1}^{n} \pi(y_i \mid x_i, \boldsymbol{\theta}_2).$$

The posterior distribution can thus be written as

$$\pi(\boldsymbol{x}, \boldsymbol{\theta_1}, \boldsymbol{\theta_2} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2)\pi(\boldsymbol{x} \mid \boldsymbol{\theta}_1) \prod_{i=1}^{n} \pi(y_i \mid \boldsymbol{x}, \boldsymbol{\theta_2})$$

or more succintly

$$\pi(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x} \mid \boldsymbol{\theta}) \prod_{i=1}^{n} \pi(y_i \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

As an example, Y_i is a Poisson random variable with mean μ_i. Suppose that we assume that there is also information on covariate z_i available for each y_i. Hence, we can assume that Y_i now has a mean exp $(\beta_0 + \boldsymbol{z_i^T}\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a vector of unknown regression parameters that has a multivariate Gaussian distribution with some mean and precision matrix as its prior. We can also add independent normal random effects, ν_i with zero mean to accommodate the random effects from unmeasured extra-Poisson variation and hence Y_i will have a mean of exp $(\beta_0 + \nu_i + \boldsymbol{z_i^T}\boldsymbol{\beta})$. This will also result in a GMRF that still has a sparse structure which simplifies the computations of the posterior distribution for the quantities of interest in the model. The exploitation of this feature will be crucial when Bayesian posterior computation is performed using INLA methodology.

3.3.7 GMRF approximation for non-normal likelihood

In certain models such as binomial models with logit or probit link functions, we can employ a data augmentation strategy to obtain full conditional distributions (FCD) that are still GMRF despite the non-Gaussian likelihood. However, there are cases where a data augmentation strategy is not feasible to obtain FCDs that still possess a GMRF characteristics. Fortunately, there is still a method to circumvent this difficulty.

In this scenario, the likelihood can be approximated using a Taylor series expansion up to the second order. To illustrate this method, we firstly define the GMRF based on canonical paramters $\mathbf{b}$ and $\mathbf{Q}$ as

$$\pi(\boldsymbol{x}) \propto \exp\Big(-\frac{1}{2}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{x}\Big) \tag{3.12}$$

where $\mathbf{x} \sim \mathrm{N}_p(\mathbf{b}, \mathbf{Q})$, with p as the number of dimensions, mean $\boldsymbol{\mu} = \boldsymbol{Q}^{-1}\boldsymbol{b}$ and precision $\mathbf{Q}$. Suppose that $\mathbf{x}$ represents the posterior density of GMRF latent field and it has the following form

$$\pi(\boldsymbol{x} \mid \boldsymbol{y}) \propto \exp\Big\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu}) + \sum_{i=1}^{n} g_i(x_i)\Big\}. \tag{3.13}$$

In this case, $g_i(x_i) = \log\{\pi(y_i \mid x_i, \boldsymbol{\theta})\}$, where $\boldsymbol{\theta}$ is the hyperparameter. As an example, $\log\{\pi(y_i \mid x_i, \boldsymbol{\theta})\}$ might be the Poisson log likelihood. By expanding $g_i x_i$ using a Taylor series up to the second order around an initial value, $\boldsymbol{\mu}^{(0)}$, we can derive its approximation as follows:

$$g_i(x_i) \approx g_i(\mu_i^{(0)}) + g_i^{'}(\mu_i^{(0)})(x_i - \mu_i^{(0)}) + \frac{1}{2}g_i^{''}(\mu_i^{(0)})(x_i - \mu_i^{(0)})^2.$$

By expanding the quadratic term and simple rearrangement, we obtain

$$g_i(x_i) \approx g_i(\mu_i^{(0)}) - g^{'}(\mu_i^{(0)})(\mu_i^{(0)}) + \frac{1}{2}g_i^{''}(\mu_i^{(0)})(\mu_i^{(0)})^2 + \Big\{g_i^{'}(\mu_i^{(0)}) - g_i^{''}(\mu_i^{(0)})\mu_i^{(0)}\Big\}x_i +$$

$$\Big\{\frac{1}{2}g_i^{''}(\mu_i^{(0)})\Big\}x_i^2.$$

If we set $g_i(\mu_i^{(0)}) - g^{'}(\mu_i^{(0)})(\mu_i^{(0)}) + \frac{1}{2}g_i^{''}(\mu_i^{(0)})(\mu_i^{(0)})^2 = a_i$, $g_i^{'}(\mu_i^{(0)}) - g_i^{''}(\mu_i^{(0)})\mu_i^{(0)} = b_i$ and $-g_i^{''}(\mu_i^{(0)}) = c_i$, the equation above can be rewritten as

$$g_i(x_i) = a_i + b_i x_i - \frac{1}{2}c_i x_i^2. \tag{3.14}$$

Then, we substitute (3.14) into (3.13) and this results in

$$\pi(\boldsymbol{x} \mid \boldsymbol{y}) \propto \exp\Big\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu}) + \sum_{i=1}^{n} a_i + b_i x_i - \frac{1}{2}c_i x_i^2)\Big\}.$$

Since a_i is a constant, this can be brought into the proportionality sign. By expanding the quadratic form, removing any constant term (i.e $\boldsymbol{\mu^T Q \mu}$) and simple rearrangement, we shall obtain

$$\begin{aligned}\pi(\boldsymbol{x} \mid \boldsymbol{y}) &\propto \exp\Big\{ -\frac{1}{2}\Big(\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{x}^T\text{diag}(\boldsymbol{c})\boldsymbol{x}\Big) + \boldsymbol{b^T x} + \boldsymbol{\mu^T Q x}\Big\}, \\ &\propto \exp\Big\{ -\frac{1}{2}\boldsymbol{x}^T\Big(\boldsymbol{Q} + \text{diag}(\boldsymbol{c})\Big)\boldsymbol{x} + \Big(\boldsymbol{Q\mu} + \boldsymbol{b}\Big)^T\boldsymbol{x}\Big\}.\end{aligned}$$

This has the same form as (3.12), with $\mathbf{x} \sim \text{N}_W(\boldsymbol{Q\mu} + \boldsymbol{b}, \boldsymbol{Q} + \text{diag}(\boldsymbol{c}))$. The mean, say $\mu^{(1)}$, is obtained by

$$\boldsymbol{\mu}^{(1)} = \Big(\boldsymbol{Q} + \text{diag}(\boldsymbol{c})\Big)^{-1}\Big(\boldsymbol{Q\mu} + \boldsymbol{b}\Big).$$

We can improve the approximation now by expanding about $\mu^{(1)}$ and repeat the whole process for s iterations, where μ^s is now the mean of $\pi(\boldsymbol{x} \mid \boldsymbol{y})$ with precision $\boldsymbol{Q}^m = Q + \text{diag}(\boldsymbol{c}^m)$.

Chapter 4

Integrated Nested Laplace Approximation (INLA)

4.1 Introduction

In this chapter, we will introduce the Integrated Nested Laplace Approximation (INLA) approach as an alternative to MCMC. In 4.2, we will explain the fundamental concepts of the Laplace approximation in solving integrals. In 4.3, an overview of INLA methodology will be given with emphases on the INLA steps for fast computation and exploration of the posterior distribution and the use of the central composite design (CCD) approach when the number of hyperparameters is moderately large. In 4.4, we will illustrate how INLA works using a simple linear regression example. In 4.5, we will show how INLA can be implemented in the survival analysis context, with particular attention given to dynamic Bayesian survival models based on piecewise constant hazard models. Finally, in 4.6, we propose a novel extension to the use of INLA which allows us to incorporate missing data within the INLA framework. This extension greatly extends the applicability of INLA in areas such as the analysis of survival data.

4.2 Laplace approximation

The Laplace approximation of an integral is the main building block for INLA methodology. To illustrate this, consider that we want to solve this integral:

$$\int_{-\infty}^{\infty} f(x)\mathrm{dx} \;=\; \int_{-\infty}^{\infty} \exp\,[\log\{\mathrm{f(x)}\}]\;\mathrm{dx}$$

where $f(x)$, for instance, is proportional the probability density function (pdf) of a random variable X. We can obtain an approximation to the function log f(x) using a Taylor's series expansion up to a second degree Taylor polynomial, assessed at x = x_0. We write

$$\log\,\mathrm{f(x)} \approx \log\,f(x_0) + (x-x_0)\frac{\mathrm{dlog\ f(x)}}{\mathrm{dx}}\Big|_{x=x_0} + \frac{(x-x_0)^2}{2}\frac{\mathrm{d^2log\ f(x)}}{\mathrm{d}x^2}\Big|_{x=x_0}.$$

If we let x_0 be the mode of X (written here as x^*), provided that log [f(x)] is differentiable at the mode, the term $\frac{\mathrm{d}\log f(x)}{\mathrm{d}x}\Big|_{x=x^*}$ will be zero and the approximation will reduce to

$$\log\,\mathrm{f(x)} \approx \log\,f(x^*) + \frac{(x-x^*)^2}{2}\frac{\mathrm{d}^2\log f(x)}{\mathrm{d}x^2}\Big|_{x=x^*}.$$

Hence, the original integrand can thus be approximated by

$$\int \mathrm{f(x)}\;\mathrm{dx} \;\approx\; \int \exp\!\left(\log\,f(x^*) + \frac{(x-x^*)^2}{2}\frac{\mathrm{d^2log\ f(x)}}{\mathrm{d}x^2}\Big|_{x=x^*}\right)\mathrm{d}x.$$

The term exp[log{f(x^*}] can be factorised out, resulting in the final form

$$\int \mathrm{f(x)}\;\mathrm{dx} \;\approx\; \exp\!\left(\log\,f(x^*)\right)\int \exp\!\left(\frac{(x-x^*)^2}{2}\frac{\mathrm{d^2log\ f(x)}}{\mathrm{d}x^2}\Big|_{x=x^*}\right)\mathrm{d}x.$$

In this case, we can associate the integral term with a Gaussian density by substituting $\frac{\mathrm{d}^2\log f(x)}{\mathrm{d}x^2}\Big|_{x=x^*} = -\frac{1}{\sigma^2}$ and which results in

$$\int \mathrm{f(x)}\;\mathrm{dx} \;\approx\; \exp\!\left(\log\,f(x^*)\right)\int \exp\!\left(-\frac{(x-x^*)^2}{2\sigma^{*^2}}\right)\mathrm{d}x.$$

We can hence clearly see here that the integrand has the form of Gaussian density's kernel with mean x^* and variance σ^{*^2}. Therefore, an integrand whose limits are γ and κ can be approximated by

$$\int_{\gamma}^{\kappa} \mathrm{f(x)}\;\mathrm{dx} \approx\; \mathrm{f}(x^*)\sqrt{2\pi\sigma^{2*}}\Big[\Phi\Big(\frac{\kappa-x^*}{\sigma^*}\Big)\;-\;\Phi\Big(\frac{\gamma-x^*}{\sigma^*}\Big)\Big],$$

where $\Phi(.)$ represents the standard Gaussian cdf.

4.3 Overview of INLA methodology

4.3.1 INLA steps in a nutshell

For this subsection, we use $\mathbf{x}$ to represent n variables in the latent Gaussian random field that include structured additive predictors η_i, the intercept β_0, $\{\beta_k\}$ and the unknown functions $\{f^{(j)}\}$. The hyperparameters are represented by $\boldsymbol{\theta}$ and $\mathbf{y}$ is the vector of random observations (refer Eq 3.2). We want to compute the posterior marginal densities $\pi(x_i \mid \boldsymbol{y})$ and $\pi(\theta_j \mid \boldsymbol{y})$ as

$$\pi(x_i \mid \boldsymbol{y}) = \int \pi(x_i \mid \boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta},$$

$$\pi(\theta_j \mid \boldsymbol{y}) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}_{-j}.$$

The INLA approach is dependent on constructing nested approximations $\widetilde{\pi}(x_i \mid \boldsymbol{y})$ and $\widetilde{\pi}(\theta_j \mid \boldsymbol{y})$ using

$$\widetilde{\pi}(x_i \mid \boldsymbol{y}) = \int \widetilde{\pi}(x_i \mid \boldsymbol{\theta}, \boldsymbol{y})\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}$$

where

$$\widetilde{\pi}(\theta_j \mid \boldsymbol{y}) = \int \widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}_{-j}.$$

According to Rue et al. (2009), the posterior marginal for the latent field $\pi(x_i \mid \boldsymbol{y})$ can be computed using the following three steps:

- **Step 1**: Approximate $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$ using a Laplace approximation. The approximation is represented by this notation: $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$. It is derived as

$$\begin{aligned}
\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y}) &= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x} \mid \boldsymbol{y})}{\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \\
&= \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})/\pi(\boldsymbol{y})}{\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \\
&\propto \frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \\
&\propto \left.\frac{\pi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}.
\end{aligned}$$

  The denominator represents the Gaussian approximation to the FCD of $\mathbf{x}$ and $\boldsymbol{x}^*(\boldsymbol{\theta})$ represents the mode of the FCD of $\mathbf{x}$ given $\boldsymbol{\theta}$.

This is performed by locating the mode $\boldsymbol{\theta}^*$ for $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ via the optimization of log $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$. A quasi-Newtonian method is employed to produce an approximation to the second derivative of log $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ by using the difference between successive gradient vectors. Next, the negative Hessian matrix, $\mathbf{H}$, is computed at modal $\boldsymbol{\theta}^*$ using finite differencing. Then we take the inverse of the Hessian matrix to obtain our covariance matrix $\boldsymbol{\Sigma}$ for $\boldsymbol{\theta}$. Subsequently, log $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ is explored by firstly parameterising $\boldsymbol{\theta}$ into a standardised variable $\mathbf{z}$ using a **z-parametrisation**. These standardised variables, $\mathbf{z}$ are computed by firstly obtaining the spectral decomposition of our covariance matrix, $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T.$$

Then $\boldsymbol{\theta}$ is parametrised as

$$\boldsymbol{\theta}(\boldsymbol{z}) = \boldsymbol{\theta}^* + \boldsymbol{V}\boldsymbol{\Lambda}^{\mathbf{1/2}}\boldsymbol{z}$$

where $\boldsymbol{\theta}^*$ is the modal configuration. As a simple example, if $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ is a normal density, $\mathbf{z}$ will then be a standard multivariate Gaussian variable, $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$. This z-parametrisation is required since the most efficient grid of evaluation points would not be parallel to the untransformed axes. The z-parameterisation will thus rectify the scale and rotation of a density which will make numerical integration easier and more efficient (Smith et al., 1987). Subsequently, log $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ is explored in a regular grid to locate the bulk of the probability mass.

To illustrate this, let us consider the case where the number of dimensions is 2. In Figure 4.1, the subfigure on the left shows the new z-axes for z-values obtained via the previous z-parametrisation and the modal position for a 2-dimensional case. To locate the probability mass of log $[\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})]$, we initially move along the positive direction of z_1 using a prespecified length of step of, for instance, 1 (denoted by δ_z =1), provided

$$\log[\widetilde{\pi}(\boldsymbol{\theta}(\mathbf{0}) \mid \boldsymbol{y})] - \log[\widetilde{\pi}(\boldsymbol{\theta}(\boldsymbol{z}) \mid \boldsymbol{y})] < \delta_\pi$$

where δ_π can be prespecified and take any value deemed appropriate. However, we can use different values for step length, δ_z, such as 0.5 or 0.25 if more stringent accuracy is needed for $\mathbf{x}$. For δ_π, Rue et al. (2009) chose $\delta_\pi = 2.5$ in his example. We then change direction and move along the negative direction of z_1 and the same process is repeated for z_2.

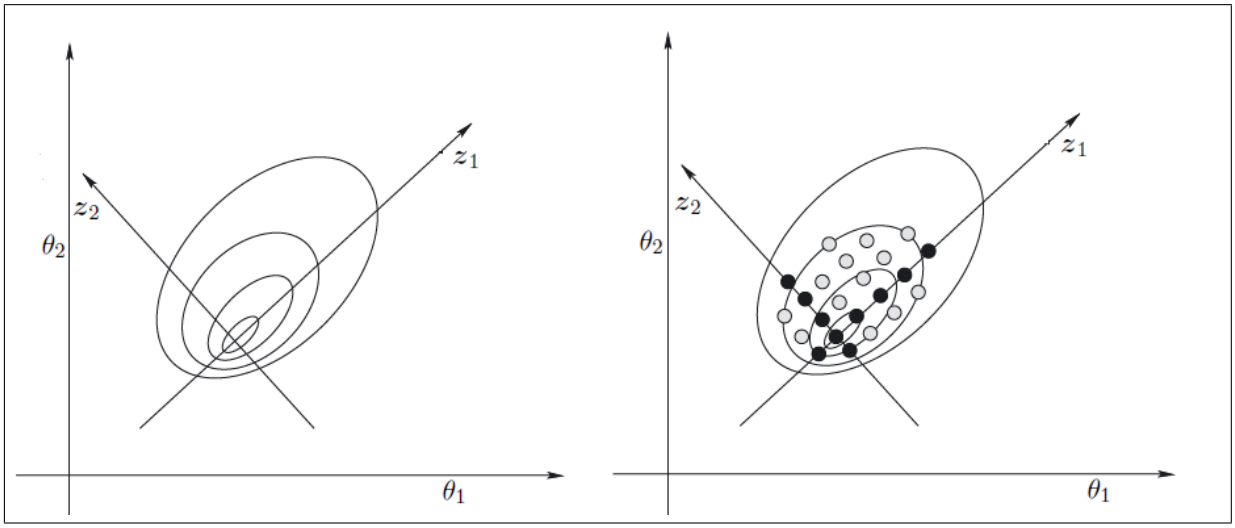

Figure 4.1: Log $[\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y}]$ exploration (Figure from Rue et al., 2009)

This results in the black dots seen in the subfigure on the right 4.1. To obtain the intermediate values, we can take combinations of z_1 and z_2 values and these are represented by the grey dots in the same subfigure and these values are retained as long as the condition for π_δ above holds. Since the points are placed in regular grids, all area weights, Δ_k, are equal. The posterior marginal of θ_j, $\widetilde{\pi}(\theta_j \mid \boldsymbol{y})$, can then be straightforwardly acquired from $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ by creating an interpolant with respect to $\log \widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$ using the point obtained previously. We can then use this interpolant to compute the posterior marginals, $\widetilde{\pi}(\theta_j \mid \boldsymbol{y})$, via numerical integration. Besides, these integration points (area weights) can also be used as area weights in step 3 of the INLA for approximating $\widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})$.

- **Step 2**: $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ is approximated by any one of these three methods: Gaussian approximation, Laplace and simplified Laplace approximation. According to Rue et al. (2009), the Gaussian approximation is the easiest to compute but the least accurate due to errors in the location and skewness. The Laplace approximation gives accurate results and this is computed using the relationship

$$\widetilde{\pi}_{LA}(x_i \mid \boldsymbol{\theta}, \boldsymbol{y}) \propto \left. \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_{GG}(\boldsymbol{x_{-i}} \mid x_i, \boldsymbol{\theta}, \boldsymbol{y})} \right|_{\boldsymbol{x}_{-i} = \boldsymbol{x}^*_{-i}(x_i, \boldsymbol{\theta})}$$

where $\boldsymbol{x}_{-i} = \boldsymbol{x}^*_{-i}(x_i, \boldsymbol{\theta})$ is the modal configuration for the full conditional distribution of $\boldsymbol{x}_{-i}$ given x_i and $\boldsymbol{\theta}$, and $\widetilde{\pi}_{GG}(\boldsymbol{x}_{-i} \mid x_i, \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation for the full conditional distribution $\widetilde{\pi}(\boldsymbol{x}_{-i} \mid x_i, \boldsymbol{\theta}, \boldsymbol{y})$. However, note that

$\widetilde{\pi}_{GG}(\boldsymbol{x}_{-i} \mid x_i, \theta, y)$ is not the same as the Gaussian approximation for $\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$ [i.e. $\widetilde{\pi}_G(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$]. Unfortunately this approach is quite expensive since $\widetilde{\pi}_{GG}$ has to be reevaluated for each x_i and $\boldsymbol{\theta}$ since its precision matrix, $\mathbf{Q}$, is dependent upon both x_i and $\boldsymbol{\theta}$. There are several modifications that had been proposed by Rue et al. (2009), which are beyond the scope of our discussion, to improve the computational efficiency of this approximation strategy but the Laplace approximation is still considered to have the highest computational cost among those three approximation strategies (Rue et al., 2009).

The simplified Laplace approximation is the best approximation in terms of optimal computing efficiency due to the much reduced computational cost of the simplified Laplace approximation. This is because the simplified Laplace approach approximates $\pi(x_i \mid \boldsymbol{\theta}, \boldsymbol{y})$ via a Taylor's series expansion around the mode up to the third order which will produce the linear and cubic correction terms for the standardized Gaussian approximation (Rue et al., 2009). The multivariate skew-normal distribution is subsequently matched to the linear and cubic correction terms in the Taylor's series expansion of $\pi(x_i \mid \boldsymbol{\theta}, \boldsymbol{y})$ to rectify the errors in location and skewness of the Gaussian approximation (Azzalini and Capitano, 1999; Rue et al., 2017). This is considered to be the optimal choice despite a small reduction in computing efficiency.

- **Step 3**: $\widetilde{\pi}(x_i \mid \boldsymbol{y})$ is obtained by combining the previous two approximations and integrating out $\boldsymbol{\theta}$ using a numerical integration technique called weighted finite sum. This is represented as follows:

$$\begin{aligned} \widetilde{\pi}(x_i \mid \boldsymbol{y}) &= \int \widetilde{\pi}(x_i \mid \boldsymbol{\theta}, \boldsymbol{y}) \widetilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta} \\ &\approx \sum_{k=1}^{K} \widetilde{\pi}(x_i \mid \boldsymbol{\theta}_k, \boldsymbol{y}) \widetilde{\pi}(\boldsymbol{\theta}_k \mid \boldsymbol{y}) \Delta_k. \end{aligned}$$

Here, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k$ are the points selected in step 1 within the $\boldsymbol{\theta}$ space and $\Delta_i, \ldots, \Delta_k$ are the weights of integration. Using similar steps, the posterior marginal densities of $\widetilde{\pi}(\theta_j \mid \boldsymbol{y})$, j = 1,..., M can be computed. These two posterior marginal distributions can be summarised using means, quantiles and variances. The deviance information criterion (DIC) (Spiegelhalter et al., 2002, 2014) can also be produced for model comparisons and assessing the model complexity.

To ensure the efficiency of INLA, additional conditions are required for the latent Gaussian model. Firstly, the latent Gaussian field $\mathbf{x}$ is required to have a conditional

independence property, making it a latent Gaussian Markov random field that posesses a sparse precision matrix $\mathbf{Q}$. Besides, the number, m, of hyperparameters (i.e. the unknown parameters that are not included in the latent field $\boldsymbol{x}$, for instance the shape parameter, α, for the Weibull survival time distribution, the variance σ^2 for normal observations $\boldsymbol{y}$, the overdispersion parameter for a negative binomial distribution etc.) should not be excessively large (ideally $m \leq 6$). However, Rue et al. (2009) demonstrated that using an integral scheme called Central Composite Design (CCD), this limitation can be easily circumvented and a larger number of hyperparameters can thus be easily accommodated ($m \leq 20$). In this approach, the integration problem is considered as a design problem and CCD is used to choose the optimal selection of points to be laid out when approximating the bulk of probability mass in the higher-dimensional posterior distribution.

4.3.2 INLA and the central composite design approach (CCD)

As mentioned above, INLA is an efficient algorithm if the number of hyperparameters is small ($m \leq 6$). The efficiency of the algorithm starts to drop when m increases to a moderate number ($6 < m \leq 12$), signified by rising computing time. If we employ the integration strategy shown in Figure 4.1, we need evaluation points of $\mathcal{O}(5^m)$. If we restrict our evaluation points to a lesser magnitude per dimension, the computational cost is still unacceptably high. Hence, we require a new approach to overcome this impasse.

One way to do this is to consider this integration problem as an experimental design problem where our task is to find the optimal number of points that should be placed in a space of m dimensions. One particular approach is based on response surface methodology, using a conventional design such as a central composite design (CCD) (Box and Wilson, 1951). A CCD is a factorial or fractional factorial design with some clever modifications: the centre points (central black dot in Figure 4.2) are supplemented by additional design points (black dots on the squares and corners of the cubes) and star points (the black dots on the axes, away from the central points (eg. the points with (α, 0),($-\alpha$,0), (0,α), (0,$-\alpha$) coordinates for $m = 2$)). Using this approach, we can optimise the number of evaluation points for numerical integration.

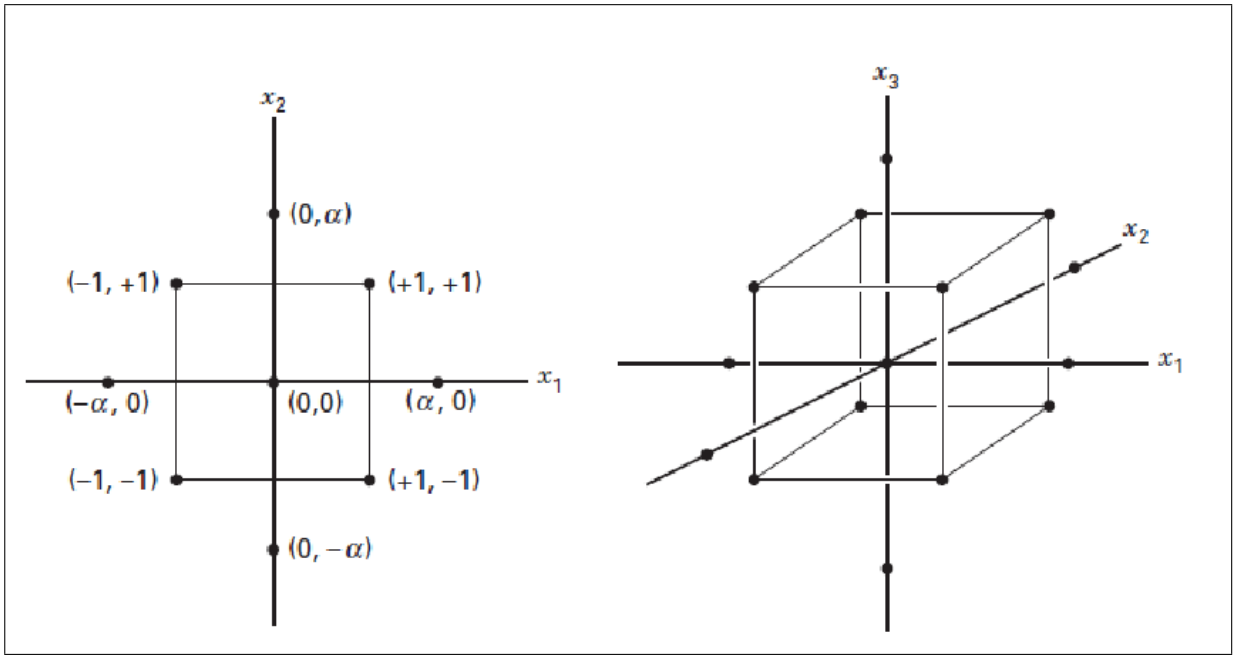

Figure 4.2: Central composite design (CCD) for $m = 2$ (left) and $m = 3$ (right) (Figure from Montgomery (2017))

As an example, for $m = 5$, the usual grid-search method requires 3125 evaluation points if 5 evaluation points per dimension are used ($5^5 = 3125$), or at least 243 evaluation points if only 3 evaluation points are placed in each dimension ($3^5 = 243$). However, if the CCD approach is used, we only require 27 evaluation points for numerical integration. Hence, this approach will be definitely invaluable in preserving the computational efficiency of INLA.

4.4 A Simple Example

4.4.1 Simple linear regression

In this example, we shall carefully describe how the algorithm works. This example is adapted from Blangiardo and Cameletti (2015). In this example, we will use a simple linear regression model with only an intercept. For further information on INLA code used in this example, please refer to appendix A.10.

Firstly, we shall assume a vector $\mathbf{Y}$ to be a realisation of Gaussian random variables that are identically and independently distributed with mean μ and variance σ^2. For simplicity, we assign independent priors for both μ and σ^2 (or $\tau = \frac{1}{\sigma^2}$)

$$\mu \sim \mathcal{N}\ (\mu_0, \tau)$$

$$\tau \sim \text{Ga}\ (a, b).$$

In the original INLA notation (Rue et al., 2009), the observations can be represented by the following

$$Y_i \mid \theta, \tau \sim \mathcal{N}\ (\eta_i, \tau) \quad i = 1, \ldots, \text{n}$$

where $\eta_i = \theta = \mu$ and τ is the precision.

The initial stage of the INLA algorithm is to integrate out θ from the joint distribution of θ and τ so that the marginal posterior distribution of the hyperparameter τ can be obtained

$$\text{p}(\tau \mid \boldsymbol{y}) = \int \text{p}\ (\theta, \tau \mid \boldsymbol{y})\ \text{d}\theta.$$

Since $\text{p}(\theta, \tau \mid \boldsymbol{y}) \propto \text{p}(\boldsymbol{y} \mid \theta, \tau)\ \text{p}(\theta)\ \text{p}(\tau)$, we can bypass the integration using the following relationship

$$\begin{aligned} \text{p}(\tau \mid \boldsymbol{y}) &= \frac{\text{p}(\theta, \tau \mid \boldsymbol{y})}{\text{p}(\theta \mid \tau, \boldsymbol{y})} \\ &\propto \frac{\text{p}(\boldsymbol{y} \mid \theta, \tau)\text{p}(\theta)\text{p}(\tau)}{\text{p}(\theta \mid \tau, \boldsymbol{y})} \end{aligned}$$

and the distributions for each quantity are

$$\boldsymbol{y} \mid \theta, \tau\ \sim \prod_{i=1}^{n} \mathcal{N}\ (\theta, \tau^{-1})$$

$$\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$\tau \sim \text{Ga}(a, b)$$

$$\theta \mid \tau, \boldsymbol{y} \sim \mathcal{N}(\theta_n, \sigma_n^2)$$

where

$$\theta_n = \frac{\tau \sum_{i=1}^{n} y_i + \frac{\mu_0}{\sigma_0^2}}{n\tau + \frac{1}{\sigma^2}},\ \sigma_n^2 = \frac{1}{n\tau + \frac{1}{\sigma_0^2}}.$$

One particular point should be stressed here. Since the observations in this example follow a Gaussian distribution, the quantity $\text{p}(\theta \mid \tau, \boldsymbol{y})$ does not require Laplace approximation. Besides, we do not need to perform Laplace approximation even though the

posterior distribution, $\mathrm{p}(\tau \mid \boldsymbol{y})$, does not have conjugate structure since the terms in the numerator and denominator that are dependent on θ will cancel out. As a result, we may opt for any θ values and the most sensible option is to equate this with the conditional posterior mode, $\theta = \theta_n$. Hence, this results in

$$\mathrm{p}(\tau \mid \boldsymbol{y}) \propto \left.\frac{\mathrm{p}(\boldsymbol{y} \mid \theta, \tau)\mathrm{p}(\theta)\mathrm{p}(\tau)}{\mathrm{p}(\theta \mid \tau, \boldsymbol{y})}\right|_{\theta=\theta_n} = \left.\frac{\mathrm{p}(\boldsymbol{y} \mid \theta, \tau)\mathrm{p}(\theta)\mathrm{p}(\tau)}{\frac{1}{\sqrt{2\pi\sigma_n^2}}}\right|_{\theta=\theta_n}. \tag{4.1}$$

It should be noted that 4.1 is an unnormalised posterior distribution for τ. Hence, the most possible values for τ were selected ($\{\tau^{(z)}\}$)in order to appraise the the unnormalised τ pdf. The value of the density function of each τ is computed using

$$\mathrm{p}(\tau^{(z)} \mid \boldsymbol{y}) \propto 2\pi\sigma_n^2 \underbrace{\mathrm{p}(\boldsymbol{y} \mid \theta = \theta_n, \sigma^2 = \frac{1}{\tau^{(z)}})}_{\prod_{i=1}^{n}\mathcal{N}(\theta_n, \frac{1}{\tau})} \underbrace{\mathrm{p}(\theta = \theta_n)}_{\mathcal{N}(\mu_0, \sigma_0^2)} \underbrace{\mathrm{p}(\tau^{(z)})}_{Ga(a,b)} \tag{4.2}$$

which in this case

$$\theta_n = \frac{\tau^{(z)} + \sum_{i=1}^{n} y_i + \frac{\mu_0}{\sigma_0^2}}{n\tau^{(z)+} + \frac{1}{\sigma_0^2}}$$

and

$$\sigma_n^2 = \frac{1}{n\tau^{(z)} + \frac{1}{\sigma_0^2}}.$$

The quantities for 4.2 can be computed using the R code given in appendix A.10, starting from line #1 until the last line in compute quantities section. The unnormalised τ pdf can be normalised using the codes presented in normalise the tau density (precision) section.

In the next step, we have to analyse $\mathrm{p}(\theta \mid \boldsymbol{y})$ for every τ value within the τ set, $\{\tau^{(j)}\}$, and for every θ within $\{\theta^{(j)}\}$. We can do this by assessing $\mathrm{p}(\theta = \theta^{(\ell)} \mid \boldsymbol{y})$ based upon the Gaussian distribution with mean θ_n and variance σ_n^2. Next, using a finite weighted sum, we integrate out τ from the joint distribution $\mathrm{p}(\theta, \tau \mid \boldsymbol{y})$:

$$\mathrm{p}(\theta = \theta^{(\ell)} \mid \boldsymbol{y}) \propto \sum_j \mathrm{p}(\theta^{(\ell)} \mid \tau^{(j)}, \boldsymbol{y})\mathrm{p}(\tau^{(j)} \mid \boldsymbol{y})\Delta_j$$

with $\Delta_j = \Delta = \frac{1}{\sum_j \mathrm{p}(\tau^{(j)}|\boldsymbol{y})}$. The R codes for this second step can be obtained in appendix A.10 (starting from Select J Grid Points until normalising the density for marginal posterior theta).

4.4.2 Analysis of Piston-ring failures data

4.4.2.1 The data set characteristics and setup

This data was analysed as a simple INLA example. The data, which was originally published in Davies and Goldsmith (1972), was obtained from Hand et al. (1994) which is available for public access via **"http://www.stat.ncsu.edu/research/sas/sicl/data/piston.dat"**. The data are on the number of piston-ring failures in three separate legs for four different types of steam-driven compressors within a building. All compressors have the same design and were positioned in similar fashion. The purpose of analysing this data was originally to assess whether the probability of piston-rings failures was associated with the different types of legs and compressors. Davies and Goldsmith (1972) initially analysed this data using a simple chi-square test, but a Poisson regression model is also a suitable alternative for answering this research question.

Table 4.1: Piston-ring failures data adapted from Hand et al. (1994)

Compressor no	**Leg**			
	North	**Centre**	**South**	**Total**
1	17	17	12	46
2	11	9	13	33
3	11	8	19	38
4	14	7	28	49
Total	53	41	72	166

4.4.3 The relationship between Poisson and multinomial likelihoods

The Poisson-multinomial relationship was shown by McCullagh and Nelder (1989) who demonstrated that the conditional distribution of multiple Poisson random variates given the total is equivalent to a multinomial distribution. To be more specific, consider counts from individual groups or species as independent Poisson random variables, each with mean μ_i. We can subsequently obtain the joint distribution of these independent Poisson random variables which can be conditioned on the sum of the total observed count. Hence, the sum of these independent Poisson random variables is also a Poisson random variable

with a mean obtained by summing up their respective Poisson means (i.e. $\mu_1 + \mu_2 + \ldots + \mu_k$). Hence, the total observations, N, will also be a random variable (since N = $Y_1 + Y_2 + \ldots + Y_k$).

Nevertheless, if we restrict or fix the total of the observations, $N = n$, then individual Y_i will no longer be independent Poisson random variables since the maximum count of Y in group i cannot exceed the total observed count. Besides, the assumption of independence will also be violated since the larger count of Y_i in group i will make the observed counts in other groups smaller. Hence, the distribution of $Y_1, \ldots, Y_k$ given $\sum_{i=1}^{k} Y_i = n$ is no longer Poisson.

To derive it, we first write

$$\begin{aligned}
\Pr(Y_1 = n_1, \ldots, Y_k = n_k) &= \frac{\mu_1^{n_1} \mathrm{e}^{-\mu_1}}{n_1!} \times \ldots \times \frac{\mu_k^{n_k} \mathrm{e}^{-\mu_k}}{n_k!} \\
&= \prod_{i=1}^{k} \frac{\mu_i^{n_i} \mathrm{e}^{-\mu_i}}{n_i!} \\
&= \mathrm{e}^{-\sum_{i=1}^{k} \mu_i} \prod_{i=1}^{k} \frac{\mu_i^{n_i}}{n_i!}
\end{aligned}$$

Since $Y_i, \ldots, Y_k$ are independent, $\sum_{i=1}^{k} Y_i$ has a Poisson distribution with mean $M = \sum_{i=1}^{k} \mu_i$ (Grimmett and Stirzaker, 2001). Hence, $\Pr(\sum_{i=1}^{k} Y_i = n) = \mathrm{e}^{-M} \frac{M^n}{n!}$

So the conditional distribution of $Y_1, \ldots, Y_k$ given $\sum_{i=1}^{k} Y_i$ is given by

$$\begin{aligned}
\Pr(Y_1 = n_1, \ldots, Y_k = n_k \mid \sum_{i=1}^{k} Y_i = n) &= \frac{\Pr(Y_1 = n_1, \ldots, Y_k = n_k \cap \sum_{i=1}^{k} Y_i = n)}{\Pr(\sum_{i=i}^{k} Y_i = n)} \\
&= \frac{\Pr(Y_1 = n_1, \ldots, Y_k = n_k)}{\Pr(\sum_{i=1}^{k} Y_i = n)}
\end{aligned}$$

$$= \frac{\mathrm{e}^{-\sum_{i=1}^{k}\mu_i}\prod_{i=1}^{k}\mu_i^{n_i}/\ n_i!}{\mathrm{e}^{-\sum_{i=1}^{k}\mu_i}\left(\sum_{i=1}^{k}\mu_i\right)^{n}/\ n!}$$

$$= \left(\frac{n!}{\prod_{i=1}^{k} n_i}\right)\prod_{i=1}^{k}\left(\frac{\mu_i}{M}\right)^{n_i}$$

$$= \left(\frac{n!}{\prod_{i=1}^{k} n_i!}\right)\ \prod_{i=1}^{k}(p_i)^{n_i}$$

with $p_i = \frac{\mu_i}{M}$. Hence, the probability function obtained above is a multinomial probability function. As a summary, the conditional distribution (i.e. conditioned on the overall sample size) is a multinomial distribution. For two independent Poisson random variables, X_1 and X_2, the distribution of X_1 conditional upon their sum is a binomial distribution, $X_1 \sim \mathrm{Bin(n,\ } \frac{\mu_1}{[\mu_1+\mu_2]})$.

4.4.3.1 Modelling strategies

For this scenario, the numbers of piston-ring failures are assumed to follow Poisson distributions. Since the mean of Poisson distribution should be positive, the log-link function is used to connect the mean number of piston-ring failures with the covariates. The mean for the number of failures in case i is λ_i. We write

$$\eta_i = \ln(\lambda_i)$$

and

$$\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j$$

where x_{ij} is the number of failures according to the types of leg and compressor.

This is hence regarded as a generalised linear model with Poisson errors and log link. Besides, a zero-sum constraint was also employed using orthogonal contrast. To achieve this, the dataset was modified so that it will have the data structure as presented in Table 4.2.

Table 4.2: The modified piston-ring failures data structure based on orthogonal contrasts

No.	Leg	Failures	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
One	North	17	2	0	-1	-1	-1	-2	-2	-2	0	0	0
One	Centre	17	-1	1	-1	-1	-1	1	1	1	-1	-1	-1
One	South	12	-1	-1	-1	-1	-1	1	1	1	1	1	1
Two	North	11	2	0	-1	1	1	-2	2	2	0	0	0
Two	Centre	9	-1	1	-1	1	1	1	-1	-1	-1	1	1
Two	South	13	-1	-1	-1	1	1	1	-1	-1	1	-1	-1
Three	North	11	2	0	1	-1	1	2	-2	2	0	0	0
Three	Centre	8	-1	1	1	-1	1	-1	1	-1	1	-1	1
Three	South	19	-1	-1	1	-1	1	-1	1	-1	-1	1	-1
Four	North	14	2	0	1	1	-1	2	2	-2	0	0	0
Four	Centre	7	-1	1	1	1	-1	-1	-1	1	1	1	-1
Four	South	28	-1	-1	1	1	-1	-1	-1	1	-1	-1	1

In this data structure, X_1 and X_2 represent the main effects for legs and X_3, X_4 and X_5 are the main effects for compressors. The rest of the Xs represent the two-way interaction effects between legs and compressors. Finally, the parameters are given independent normal distributions and the values are given in Table 4.3.

Table 4.3: The prior means and variances for the intercept (α) and regression coefficients of the piston-failure model

Parameter	Mean	Variance
α	2.3	2.2
β_1	0	0.5
β_2	0	1.5
β_3	0	1.5
β_4	0	1.5
β_5	0	1.5
β_6	0	0.25
β_7	0	0.25
β_8	0	0.25
β_9	0	0.75
β_{10}	0	0.75
β_{11}	0	0.75

The R codes for MCMC (using rjags) and INLA (R-INLA) are given in appendix

A.3. The number of iterations for MCMC was fixed at 100,000 with a burn in of 1,000. The posterior means and standard deviations are given in Table 4.4. The results are presented as posterior means and standard deviations (in brackets). Figure 4.3 show marginal posterior densities of the parameters computed using both methods.

Table 4.4: The comparisons of results for regression coefficients obtained via MCMC and INLA

Parameter	MCMC (Posterior mean (SD))	INLA (Posterior mean (SD))
α	2.5131 (0.0851)	2.5158 (0.0832)
β_1	0.0067 (0.0589)	0.0092 (0.0580)
β_2	-0.2934 (0.1054)	-0.2927 (0.1036)
β_3	0.0016 (0.0848)	0.0012 (0.0833)
β_4	-0.0432 (0.0843)	-0.0433 (0.0833)
β_5	-0.1280 (0.0848)	-0.1275 (0.0833)
β_6	-0.0257 (0.0585)	-0.0254 (0.0580)
β_7	-0.0035 (0.0585)	-0.0031 (0.0580)
β_8	-0.0231 (0.0588)	-0.0232 (0.0580)
β_9	-0.2863 (0.1050)	-0.2871 (0.1036)
β_{10}	-0.1583 (0.1050)	-0.1591 (0.1036)
β_{11}	-0.0258 (0.1044)	-0.0255 (0.1036)

As we can see from the results presented in both Table 4.4 and Figure 4.3, the posterior means, standard deviations and densities are closely similar. However, the computing time for INLA is much shorter than MCMC (0.87 seconds vs 16.83 seconds). This indicates R-INLA is more than 19 times more efficient than RJAGS in obtaining posterior distributions.

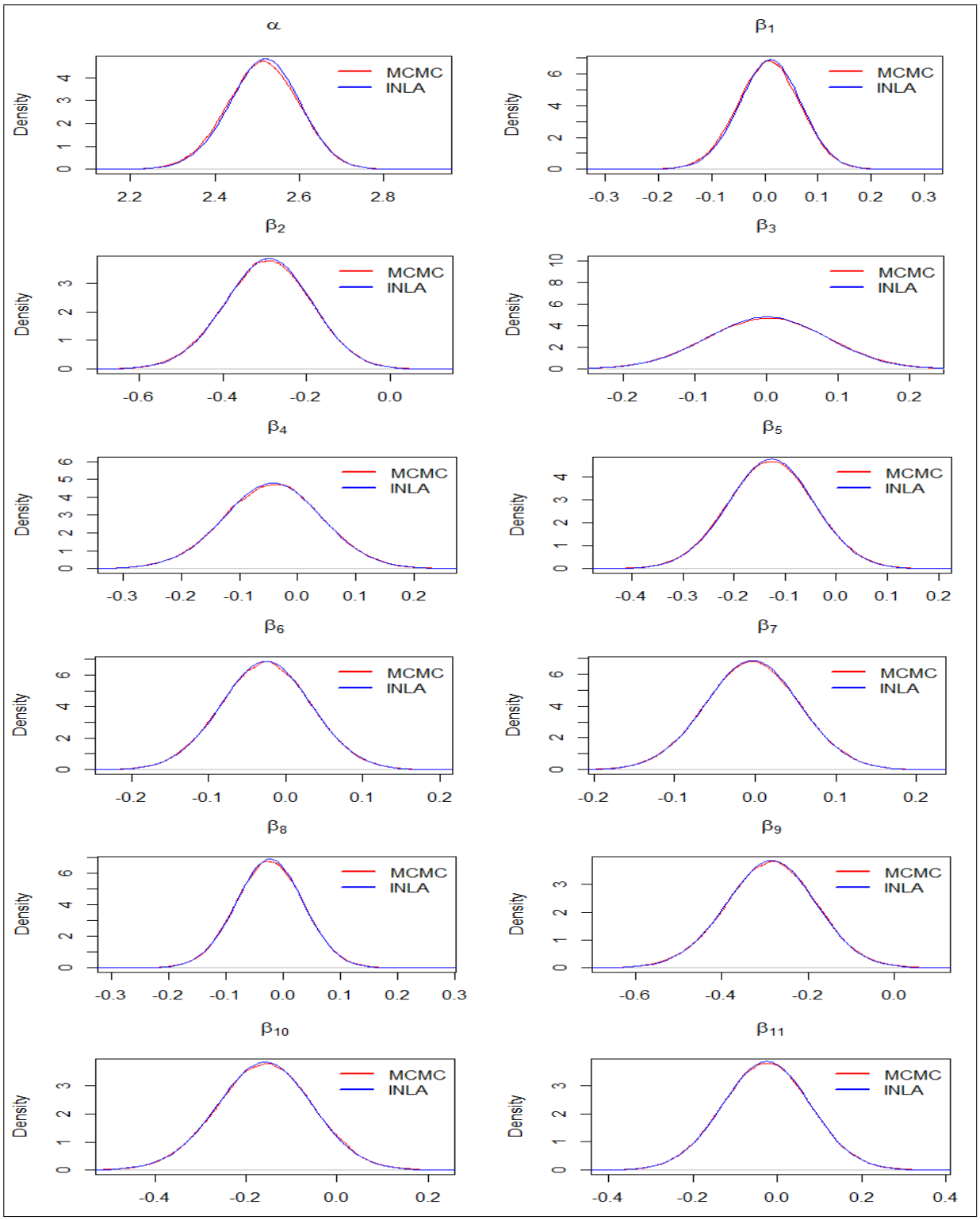

Figure 4.3: Comparisons of posterior densities of α till β_{11} obtained via MCMC (red) and INLA (blue)

4.5 INLA and survival analysis

4.5.1 Weibull lifetimes

In survival analysis, the model under consideration can easily be visualised as a latent Gaussian model (LGM). An example using a Weibull proportional hazard model will be used to elaborate on the methodological aspects of INLA in the context of survival analysis. Let us assume that we have survival times $t_1, t_2, \ldots, t_n$ which are iid observations from a Weibull distribution. Hence, the probability density function (pdf) for these survival times, given the two Weibull parameters $\alpha > 0$ and λ, can be alternatively written as

$$f(t_i \mid \alpha, \lambda_+) = \alpha t_i^{\alpha-1} \exp\{\lambda_+ - \exp\{\lambda_+\} t_i^{\alpha}\}.$$

where $\exp\{\lambda_+\} = \lambda$ (the scale parameter of a Weibull pdf).

For Weibull-distributed survival times, the survival function $S(t) = \Pr(T > t)$ is

$$S(t_i \mid \alpha, \lambda_+) = \exp\{-\exp\{\lambda_+ t_i^{\alpha}\}\}.$$

Hence, the likelihood for α and λ_+ is

$$\begin{aligned}
L(\alpha, \lambda_+ \mid D) &= \prod_{i=1}^{n} f(t_i|\alpha, \lambda_+)^{\delta_i} \;\; S(t_i|\alpha, \lambda_+)^{1-\delta_i} \\
&= \prod_{i=1}^{n} [\alpha t_i^{\alpha-1} \exp\{\lambda_+ - \exp\{\lambda_+\} t_i^{\alpha}\}]^{\delta_i} \; [\exp\{-\exp\{\lambda_+\} t_i^{\alpha}\}]^{1-\delta_i} \\
&= \prod_{i=1}^{n} [\alpha t_i^{\alpha-1} \exp\{\lambda_+\}]^{\delta_i} \; [\exp\{-\exp\{\lambda_+\} t_i^{\alpha}\}] \\
&= \alpha^{\sum \delta_i} \exp\{\lambda_+ \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} ([(\delta_i)(\alpha - 1)\log(t_i)] - \exp(\lambda_+) t_i^{\alpha})\}.
\end{aligned}$$

In the equation above, δ_i represents the survival status of the subject and hence will be recorded as 1 if the event occurs (hence there is a survival time, T_i; the first term on the right hand side exists, the second term is 1) or 0 if the observation is right censored (Censoring time (C_i); second term on the right hand side of the equation exists, first term is 1). Since $\lambda_{+_i} = \eta_i$, the covariates can be brought in to establish a regression model. Hence $\log \lambda_i = \boldsymbol{z}_i^T \boldsymbol{\beta}$; $\boldsymbol{z}_i$ is the $p \times 1$ vector of the covariates and $\boldsymbol{\beta}$ is the vector of regression coefficients. From here, we can allow the predictor to have a structured

additive form (equation 4.1). Then, Gaussian priors are allotted to each of the elements on the right side of the equation:

$$\lambda_{+_i} = \eta_i = \beta_0 + \sum_{j=1}^{n_f} w_{ij} f^{(j)}(u_{ij}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i \tag{4.3}$$

so that $\mathbf{x}$ will have a Gaussian field property and a precision matrix of $\mathbf{Q}(\theta_1)$. This is a latent Gaussian model with a latent field, $\mathbf{x}$, hyperparameters $\theta = \{\theta_1, \theta_2\}$; $\theta_2 = \alpha$. The observed data, D, are the survival status and survival time of each subject (t_i, δ_i) with i = 1,2,3,...,n. It can hence be seen that the likelihood of the observed data is dependent on the latent field $\mathbf{x}$ only via the predictor η_i. As a result, INLA can easily be applied in this setting.

An INLA scheme could also be implemented for a piecewise constant hazard model. Within the piecewise constant hazard model set-up, the log likelihood contribution of the i^{th} individual with survival time t_i (where $t_i \in [\tau_{j-1}, \tau_j)$ is given by

$$\log[f_i(t)] = \log[h_i(t)^{\delta_i} S_i(t)] = \delta_{ij} \log \lambda_{ij} - \left\{ \sum_{k=1}^{j-1} \lambda_{ik}(\tau_k - \tau_{k-1}) + (t - \tau_{j-1})\lambda_{ij} \right\} \tag{4.4}$$

where δ_{ij} is either 0 if the i^{th} individual experiences the event or 0 if censoring occurs. Equation 4.4 can be further rewritten as

$$\log[f_i(t)] = \delta_{ij}\eta_{ij} - (t - \tau_{j-1}) \exp\{\eta_{ij}\} - \sum_{k=1}^{j-1} \exp\{\eta_{ik}\}(\tau_k - \tau_{k-1}) \tag{4.5}$$

where $\lambda_{ij} = \lambda_{0j} \exp\{\boldsymbol{X}_i^T \boldsymbol{\beta}\}$ and λ_{0j} is the baseline hazard function in the j^{th} interval. Since the log likelihood is only dependent on Gaussian Latent field via the predictors $\eta_1, \ldots, \eta_j$(where $\eta_j = \log(\lambda_{ij}) = \log(\lambda_{0j} + \boldsymbol{X}_i^T \boldsymbol{\beta})$), the INLA scheme cannot be directly implemented for such model. Fortunately,the piecewise constant hazard model can still be rewritten into a Poisson regression framework, through which the INLA scheme can be directly executed.

Holford (1976), Holford (1980) and Laird and Olivier (1981) have demonstrated the equivalence of piecewise constant hazard model and Poisson regression model and this framework has been utilised by the previous researchers to fit a variety of survival models using traditional statistical software such as GLIM(Aitkin and Clayton, 1980). To proof

this, let $\delta_{ij} \sim \text{Po}(\mu_{ij})$ where $\mu_{ij} = \lambda_{ij} t_{ij}$. The pdf of δ_{ij} is therefore

$$\text{f}(\delta_{ij}) = \frac{\mu_{ij}^{\delta_{ij}} \exp\{-\mu_{ij}\}}{\delta_{ij}!}.$$

Since δ_{ij} can only be 0 or 1, $\delta_{ij}!$ can be safely disregarded and the log likelihood contribution of the i^{th} observation is therefore given by

$$\begin{aligned}
\ell_i(\lambda) &= \sum_{j=1}^{j(i)} \delta_{ij} \log(\mu_{ij}) - \mu_{ij} \\
&= \sum_{j=1}^{j(i)} \delta_{ij} \log(\lambda_{ij} t_{ij}) - \lambda_{ij} t_{ij} \\
&= \sum_{j=1}^{j(i)} \delta_{ij} \log(\lambda_{ij}) + \delta_{ij} \log(t_{ij}) - \lambda_{ij} t_{ij}. \qquad (4.6)
\end{aligned}$$

Since $\delta_{ij} \log(t_{ij})$ term does not involve any λ_{ij}, it can be disregarded and therefore 4.6 is proportional to 4.4 since the summation of over j intervals equals to the summation of $j(i)$.

Hence, from (4.5), we can clearly observe that the first two terms on the right hand side of the equation, $\delta_{ij}\eta_{ij} - (t - \tau_{ij}) \exp\{\eta_{ij}\}$ has the form of Poisson log likelihood with mean $(t - \tau_{ij}) \exp\{\eta_{ij}\}$ which is observed either 0 or 1, depending on the value of δ_{ij}. The third term, $\sum_{k=1}^{j-1} \exp\{\eta_{ik}\}(\tau_k - \tau_{k-1})$, has a $k-1$ Poisson log likelihood with mean $\exp\{\eta_{ik}\}(\tau_k - \tau_{k-1})$ since individual i survives at the end of each previous $j-1$ interval and therefore δ_{ij} is always 0. Therefore, we can fit the piecewise constant hazard model within the framework of INLA by treating each t_i as "augmented data points" originated from a Poisson distribution.

4.6 INLA and missing data: INLA within MCMC (INLA-MCMC) approach

In this section, we propose to extend the INLA within MCMC (INLA-MCMC) method proposed by Gomez-Rubio and Rue (2018) to survival analysis with missing covariate values. In their paper, Gomez-Rubio and Rue (2018) used a multiple imputation scheme

to fill in the missing covariate values for linear regression models before obtaining the posterior estimates for model parameters. Our approach is different since we directly sample the missing values and then apply this to various survival models. The overview of this proposed method will be given in this section.

Let $\boldsymbol{\theta}$ be a collection of unknown parameters and hyperparameters in a Bayesian hierarchical model. In this case, $\boldsymbol{\theta}$ represents all of the parameters of the latent effects, $\mathbf{x}$, and all the hyperparameters which include hyperparameter of the latent field $\mathbf{x}$, where $\mathbf{x} \mid \boldsymbol{\theta_1} \sim (\mathbf{0}, \boldsymbol{Q^{-1}}(\boldsymbol{\theta_1})]$ and the hyperparameter in the likelihood of the observations or data [i.e. $\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta_2} \sim \prod_{i=1}^{n} \pi\ (y_i \mid \mathbf{x}, \boldsymbol{\theta_2})$]. To ensure that INLA can be used to fit the model, we fix some of the model parameters. Hence, we split $\boldsymbol{\theta}$ into $\boldsymbol{\theta}_\xi$ and $\boldsymbol{\theta}_{-\xi}$, where $\boldsymbol{\theta}_\xi$ is the collection of parameters we fixed and $\boldsymbol{\theta}_{-\xi}$ is the parameters we do not fix. Thus, by conditioning on $\boldsymbol{\theta}_\xi$, INLA can be used to obtain the posterior marginals of all parameters in $\boldsymbol{\theta}_{-\xi}$, $\pi\ (\boldsymbol{\theta}_{-\xi} \mid \boldsymbol{y}, \boldsymbol{\theta}_\xi)$, and the conditional marginal likelihood, $\pi(\mathbf{y} \mid \boldsymbol{\theta}_\xi)$. In R-INLA, $\pi(\mathbf{y} \mid \boldsymbol{\theta}_\xi)$ is automatically stored and therefore can be easily called for and combined with the prior distributions for $\boldsymbol{\theta}_\xi$ to obtain the posterior marginal distributions for each element in $\boldsymbol{\theta}_\xi$.

The Metropolis-Hastings algorithm (Gamerman and Lopes, 2006) can subsequently be used to draw values for $\boldsymbol{\theta}_\xi$ so that its joint posterior distribution can be obtained. At step $i+1$ of the Metropolis Hastings algorithm, new values $\boldsymbol{\theta}_\xi^{(i+1)}$ for $\boldsymbol{\theta}_\xi$ are proposed and the new values are accepted with an acceptance probability

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{y} \mid \boldsymbol{\theta}_\xi^{(i+1)})\ \pi(\boldsymbol{\theta}_\xi^{(i+1)})\ q(\boldsymbol{\theta}_\xi^{(i)} \mid \boldsymbol{\theta}_\xi^{(i+1)})}{\pi(\boldsymbol{y} \mid \boldsymbol{\theta}_\xi^{(i)})\ \pi(\boldsymbol{\theta}_\xi^{(i)})\ q(\boldsymbol{\theta}_\xi^{(i+1)} \mid \boldsymbol{\theta}_\xi^{(i)})}\right\}.$$

If the proposed value is not accepted, then $\boldsymbol{\theta}_\xi^{(i+1)}$ is set to $\boldsymbol{\theta}_\xi^{(i)}$. Here $\pi(\boldsymbol{y} \mid \boldsymbol{\theta}_\xi^{(i)})$ and $\pi(\boldsymbol{y} \mid \boldsymbol{\theta}_\xi^{(i+1)})$ are the marginal likelihoods of the model conditioned upon $\boldsymbol{\theta}_\xi^{(i)}$ and $\boldsymbol{\theta}_\xi^{(i+1)}$, which may be acquired by fitting the model using INLA with values $\boldsymbol{\theta}_\xi$ fixed to $\boldsymbol{\theta}_\xi^{(i)}$ and $\boldsymbol{\theta}_\xi^{(i+1)}$. The prior density of $\boldsymbol{\theta}_\xi$, evaluated at $\boldsymbol{\theta}_\xi^{(i)}$ and $\boldsymbol{\theta}_\xi^{(i+1)}$, are $\pi(\boldsymbol{\theta}_\xi^{(i)})$ and $\pi(\boldsymbol{\theta}_\xi^{(i+1)})$ respectively. Hence, at each step of the Metropolis-Hastings algorithm, only a model conditional upon the proposal needs to be fitted. Besides, the INLA calculation will also provide the posterior marginals $\pi(\boldsymbol{\theta}_{-\xi} \mid \boldsymbol{y}, \boldsymbol{\theta}_\xi^{(i+1)})$ for all the parameters in $\boldsymbol{\theta}_{-\xi}$.

After a suitable number of iterations, the Metropolis-Hastings algorithm will produce samples from $\pi(\boldsymbol{\theta}_\xi \mid \boldsymbol{y})$, from which the posterior marginals of the parameters in $\boldsymbol{\theta}_\xi$ can be derived. Apart from that, we will obtain a family of conditional marginal distributions for all the parameters in $\boldsymbol{\theta}_{-\xi}$. Their posterior marginals can be obtained by combining

all these conditional marginals and integrating out over $\boldsymbol{\theta}_\xi$ using

$$\pi(\theta_{-\xi,i} \mid \boldsymbol{y}) = \int \pi(\theta_{-\xi,i} \mid \boldsymbol{y}, \boldsymbol{\theta}_\xi)\pi(\boldsymbol{\theta}_\xi \mid \boldsymbol{y})d\boldsymbol{\theta}_\xi \simeq \frac{1}{N}\sum_{j=1}^{N} \pi(\boldsymbol{\theta}_{-\xi,i} \mid \boldsymbol{y}, \boldsymbol{\theta}_\xi^{(j)}).$$

Values $\{\ \boldsymbol{\theta}_\xi^{(j)}\ \}_{j=1}^{N}$ represent N samples from $\pi(\boldsymbol{\theta}_\xi \mid \boldsymbol{y})$ obtained with the Metropolis-Hastings algorithm. Thus, the posterior marginal of $\theta_{-\xi,i}$ can be obtained by averaging the conditional marginals obtained at each iteration of the Metropolis-Hastings algorithm.

4.7 Summary

In this chapter, we described the fundamentals of INLA and application of INLA to survival data. We then showed how, by nesting INLA, with an MCMC scheme we can efficiently calculate posterior marginal distribution of parameters in models even when there are missing covariate values. In Chapter 7, we will evaluate the usefulness of this method using some practical examples of survival data.

Chapter 5

Applied Datasets

5.1 Introduction

In Chapter 4, we discussed the application of INLA to survival data. In particular, in Section 4.6, we introduced a new method, INLA within MCMC, which allows the use of INLA with survival datasets when there are missing covariate values. In Chapter 7, we will illustrate the use of this method by applying it to a number of data sets. This will also allow us to evaluate the method's usefulness.

In this chapter, we present details of the data sets used for application purposes. In sections 5.2 and 5.3, the characteristic details of the kidney infection and non Hodgkin Lymphoma datasets will be respectively elucidated. In 5.4, the Malaysian lung cancer dataset will be described.

5.2 Kidney infection dataset

The kidney infection data set arose from a study which was performed to evaluate the time to the first exit-site infection in patients who required catheters for renal insufficiency (Nahman et al., 1992). Forty three (36.1%) patients used catheters which were placed surgically (Group 1) whilst for the other 76 (63.9%) patients, the catheters were placed percutaneously (Group 2). The total number of patients is thus 119. This data set contains no missing covariate information. There are three variables in the dataset:

- **Infection indicator** - 0: no, 1: yes
- **Cathether placement** - 1: surgical placement, 2: percutaneous placement

- **Time** - Times to first exit-site infections (in months).

This dataset is used as a motivational example.

5.3 Non-Hodgkin Lymphoma

5.3.1 Introduction and classification

Non Hodgkin lymphoma (NHL) is a type of blood cancer. Its origin is from lymphocytes (a subtype of white blood cells), primarily B lymphocytes (85-90% of cases) whilst the rest are derived from T-lymphocytes. According to the recent statistics, 12,294 new cases of NHL were reported in 2009 in the UK, with 4452 deaths in 2010 (Shankland et al., 2012). The age-standardised incidence of NHL was reported to have risen by 35% in the period 1988-2007. Fortunately, the survival of NHL patients in England and Wales has considerably improved: 50.8% of NHL cases are expected to survive longer than 10 years, a marked improvement compared to 21.8% 10-year survival rate for those diagnosed in the 1970s (Shankland et al., 2012).

There are diverse classifications for non Hodgkin lymphoma but the 2008 WHO classification for NHL has been adopted based on the universal consensus of experts in this field. See Table 5.1.

5.3.2 Development of NHL

The mechanisms by which lymphoma develops are complex. For B cell lymphoma, the mutations that drive the development of B cell lymphomas occur primarily during the recombination of genes responsible for the production of heavy and light chain antibodies, made possible by enzymes that enable double strand DNA breaks (Shankland et al., 2012). In normal B cell development, these breaks in the DNA strands are easily repaired by DNA repair mechanisms. Nevertheless, these strand breaks may also trigger chromosomal translocations which in turn result in proto-oncogene activation. This culminates in the development of B cell lymphoma.

On the other hand, the pathogenesis of T cell lymphoma is less well understood. Most subtypes are not associated with any well-characterised genetic or biological alterations to the T cell normal development. Recurring chromosomal translocations are rarely seen in the development of T-cell NHL (Shankland et al., 2012). Exposure to human T-cell

Table 5.1: 2008 WHO classification of Non Hodgkin Lymphoma (Shankland et al., 2012)

Diseases	
B-cell lymphoma	
Precursor B cell	Precursor B-cell lymphoblastic leukaemia or lymphoma
Mature B cell	Chronic lymphocytic leukaemia/small lymphocytic lymphoma; lymphoplasmacytic lymphoma; splenic marginal-zone lymphoma; extranodal marginal-zone B-cell lymphoma of mucosa associated lymphoid tissue; nodal marginal-zone B-cell lymphoma; follicular lymphoma; mantle-cell lymphoma; diffuse large B-cell lymphoma; Burkitt's lymphoma
B-cell proliferations of uncertain malignant potential	Lymphomatoid granulomatosis; post-transplantation lymphoproliferative disorders (polymorphic)
T-cell and Natural Killer (NK)-cell lymphoma	
Precursor T cell	Precursor T-cell lymphoblastic leukaemia or lymphoma
Extranodal mature T cell and NK cell	Mycosis fungoides; cutaneous anaplastic large-cell lymphoma; extranodal NK-cell or T-cell lymphoma; enteropathy-type lymphoma; hepatosplenic lymphoma; subcutaneous panniculitis-like lymphoma; primary cutaneous CD8-positive lymphoma; primary cutaneous γ/δ T-cell lymphoma; primary cutaneous CD4-positive lymphoma
Nodal mature T cell and NK cell	Peripheral T-cell lymphoma, not otherwise specified; angioimmunoblastic lymphoma; anaplastic large-cell ALK-positive lymphoma; anaplastic large-cell ALK-negative lymphoma; adult T-cell leukaemia/lymphoma

leukemia virus 1 (HTLV-1) and Epstein Barr Virus (EBV) have been identified to be one of the causes (aetiology) of T-cell lymphoma development (Rizvi et al., 2006). The infections by these viruses are believed to result in the clonal rearrangement of T-cell receptors, with evidence pointing towards clonal integration of HTLV-1 (Ohshima et al., 1997). The prognostic outlook for T-cell lymphoma is poor with a median survival of 8 months following diagnosis (Rizvi et al., 2006).

The pathogenesis of NHL development for both subtypes and potential therapeutic targets are summarised in Figure 5.1.

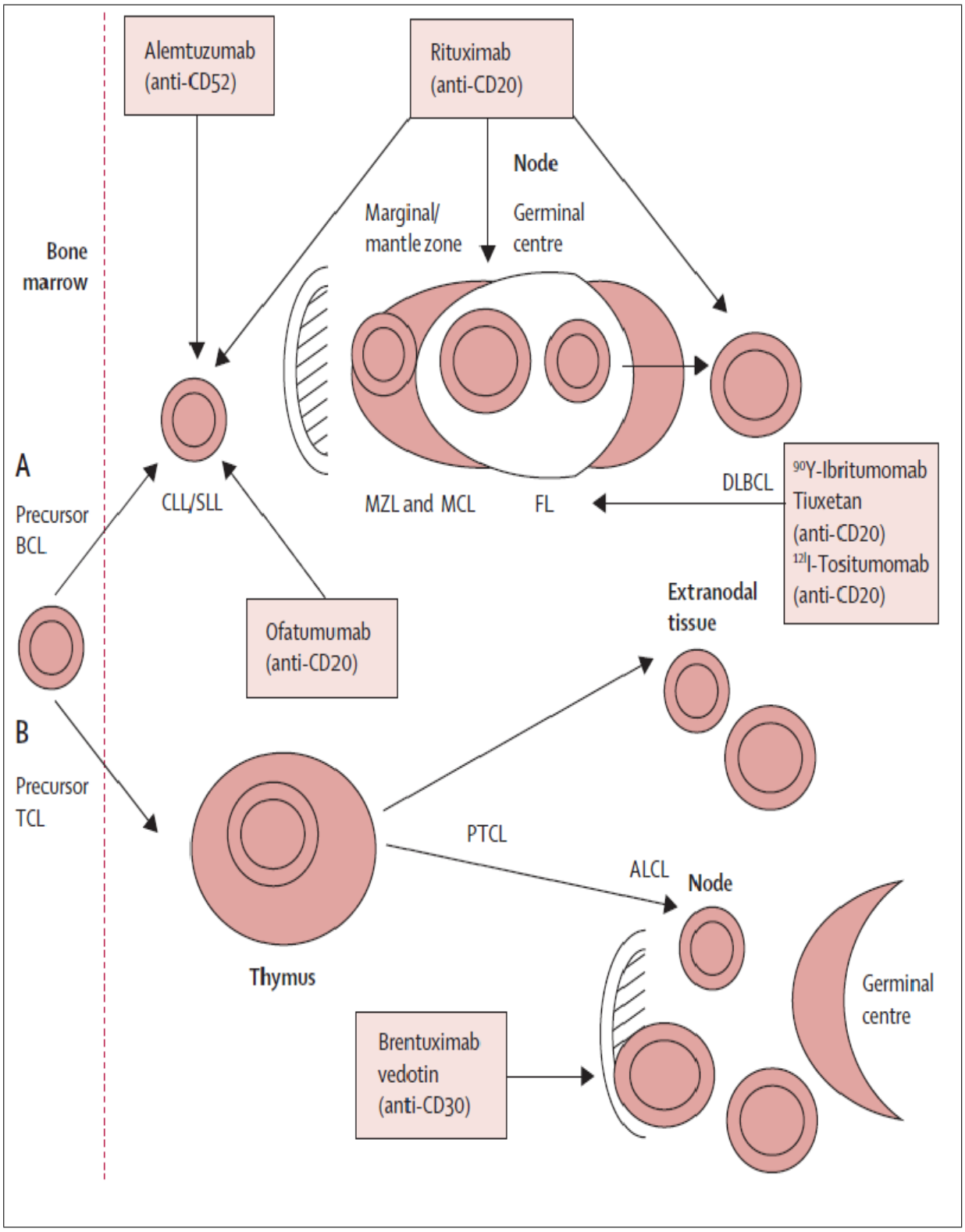

Figure 5.1: Pathogenesis of NHL and therapeutic targets (Shankland et al., 2012)

5.3.3 Method of data collection for SNLG dataset

The *Scotland and Newcastle Lymphoma Group* (SNLG) has maintained a database of Non Hodgkin Lymphoma patients from Northern England and Scotland since 1979. So far, data on 18000 NHL patients have been included in the database. The data collection process was performed using Population Adjusted Clinical Epidemiology (PACE) methodology employed by the Northern Regional Haematology Group (NRHG). In brief, the core methodology of PACE lies in the establishment and maintenance of a regional registry for all NHL subjects that includes relevant clinical profiles on each subject (demographic and clinical profiles; prognostic outcomes etc.) in a geographical area that pertains to the census population of incidence cases (Proctor and Taylor, 2000). Figure 5.2 shows the areas covered in the data collection process.

Figure 5.2: Areas of data collection

5.3.4 Covariates and outcomes

The original SNLG dataset includes all types of haematological malignancy. For the purpose of modelling, only subjects with Diffuse Large B Cell Lymphoma (DLBCL), a subtype of B cell lymphoma, were selected and used. In total, data from 2025 DLBCL patients were included in the dataset, with more than 150 prospective prognosticators and 2 outcome measures (overall survival (OS) and time to first relapse (TFR)) recorded for each NHL subject (Zhao, 2010). The whole data collection process was conducted between January 1990 and January 2004 (14 years). The data set was made available for use by Professor Stephen J. Proctor (lead investigator), Dr Michal Sieniawski (co-investigator) and Ms Jo White (data manager) who were several of the original investigators from the SNLG research group.

Out of the 2025 DLBCL patients, 1391 patients received the standard Cyclophosphamide, Hydroxydaunorubicin (also known as doxorubicin hydrochloride), Oncovin® (vincristine sulfate) and Prednisone (CHOP) chemotherapy regime. To remove the effect of treatment heterogeneity on patient survival, only this subset of patients were included in the analysis.

Fourteen covariates and one measured outcome were selected for the final dataset. Based on the preliminary analysis of the dataset, only 636 (45.7%) of DLBCL patients were found to have complete information on all 14 covariates. The numbers of DLBCL patients experiencing the event of interest (death) and censored observations are 738 and 653 respectively. Only three covariates (age, gender and Ann-Arbor stage) were completely observed. The operational definitions and descriptive summaries of each covariate are given as follows:

- **Age** - Age of patient (in years) at diagnosis.
- **Gender** - Gender for each subject, recorded either 1 or 2 for male and female, respectively.
- **Stage** - Refers to Ann Arbor Stage (Carbone et al., 1971; Shankland et al., 2012). There are 4 categories for this covariate, each corresponds to an Ann-Arbor Stage (See Table 5.2). The number of patients diagnosed with DLBCL stage I, II, III and IV is 255 (18.3%), 386 (27.7%), 354 (25.4%) and 396 (28.5%) respectively.
- **LDH** - This was originally recorded as a continuous covariate representing the patient's serum lactate dehydrogenase level. This is recorded in U/L (units per

Table 5.2: Ann Arbor staging system for Non Hodgkin Lymphoma

Stages	Descriptions
I	The lymphoma is only present in a single lymph node or lymph node region
II	The lymphoma is found in two or more than two nodal regions on the same side of diaphragm
III	The lymphoma is present in the lymph node regions on the both side of diaphragm
IV	Disseminated lymphoma that can be found beyond lymph node regions affecting extranodal sites (e.g. pleura, bone marrow, liver etc)

litre). The LDH ratio was obtained by dividing the raw value of serum lactate dehydrogenase with the upper limit of the reference range for LDH used by each centre of data collection. For analytical purposes, we categorised this covariate into three levels based on the recommendation by Zhao et al. (2014): 1 if LDH ratio $\leq$ 1; 2 if 1 $<$ LDH ratio $\leq$ 3; and 3 if LDH ratio $>$ 3. The values within the normal range of LDH were interval-censored.

- **Haemoglobin** - This represents the amount of haemoglobin (an oxygen carrying compound in the red blood cells). It is a continuous covariate and measured in g/L.

- **WBC** - This represents white blood cell (WBC) count in human blood. The unit of measurement is $\times 10^9$/ L and it was recorded as a continuous covariate. The normal range of WBC in a healthy human is between 4 to 10 $\times$ 10^9/ L.

- **ECOG** - This represents the Eastern Cooperative Oncology Group performance status (Oken et al., 1982). The ECOG scale has 5 ordinal categories. It is utilised for evaluating the fitness of cancer patients which, in turn, can be used for deciding the most suitable therapies for the patients. The descriptions for each ECOG grade are given in Table 5.3.

- **Albumin** - This is a continuous covariate representing the concentration of albumin (a type of protein produced by the human liver) in the serum. The normal reference range for serum albumin in a healthy human is between 35 g/L and 50 g/L (Longmore et al., 2014). Since albumin is rarely above the normal range, albumin is classified into two categories: -1 (normal values) and 1 (low values or hypoalbuminaemia).

Table 5.3: ECOG performance status grading system (Oken et al., 1982)

Grades	Descriptions
0	Fully active,able to carry on all pre-disease performance without functional limitations. Equivalent to Karnofsky performance score of 90 and 100
I	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work. Equivalent to Karnofsky performance score of 70 and 80
II	Ambulatory and capable of all selfcare but unable to carry out any work activities; up and about more than 50% of waking hours. Equivalent to Karnofsky perfomance score of 50 and 60.
III	Capable of only limited selfcare; confined to bed or chair more than 50% of waking hours. Equivalent to Karnofsy performance score of 30 and 40.
IV	Completely disabled; cannot carry on any selfcare; totally confined to bed or chair. Equivalent to Karnofsky performance score of 10 and 20.
V	Dead. Karnofsky performance score of 0.

- **AP** - This covariate represents the concentration of serum alkaline phosphatase (AP). It is an enzyme that is present in the human body, predominantly in the liver, bile duct, kidney and bone. The normal range is between 30 and 150 IU/L (Longmore et al., 2014). Supranormal level of AP indicates increased disease activities in the liver and bone of patients with NHL. For analytical purposes, AP is categorised into normal (-1) or high (1).

- **Urea** - This represents the concentration of urea in a patient's serum. The normal range is between 2.5 and 6.7 mmol/L (Longmore et al., 2014). High serum urea concentration indicates failing kidneys. In this dataset, serum urea concentration is categorised into normal (-1) or high (1).

- **Extranodal**- This is a categorical covariate that indicates the presence of extranodal diseases. This means that lymphoma cells have already disseminated beyond the lymph nodes. This is coded as -1 (absent) or 1 (present).

- **Bulk** - This represents the presence of bulky disease which is based on the size of lymph node mass of 10 cm or more (in any axis) (Pfreundschuh et al., 2008). In this dataset, it is coded either 1 (present) or -1 (absent).

Table 5.4: The descriptive summaries of numerical covariates and percentage of missing data for the SNLG data set ($n = 1391$)

Variable	**Min**	**Max**	**Mean (SD)**	$\mathbf{n_{miss}}$ (%)
Age	18	92	62 (14.21)	0 (0)
Haemoglobin	74	178	126.4 (18.86)	52 (3.7)
WBC	1.1	27.2	8.0 (3.5)	46 (3.3)

- **Bone marrow** - This is a categorical covariate recorded either as 1 (present) or -1 (absent). It indicates DLBCL involving bone marrow which is confirmed by the presence of lymphoma cells in the bone marrow under histopathological examination of bone marrow aspirate or biopsy.

- **Bsy** - This indicates the presence of B symptoms which are: **fever, night sweats and weight loss**, suggesting a poorer course of disease. This is coded as 1 (present) or -1 (absent).

For outcome measures, the only variable used is OS. It is defined as follows.

- **OS** - Overall survival of the patients since diagnosis until death in months (also known as time to death). The event status is recorded as either 1 (died) or 0 (censored). The time is calculated from the date of first diagnosis until the date of death ascertained from the death certificates (event) or to the last date of follow up if no death was noted (censored).

The descriptive summaries and percentages of missing data for the numerical, binary and ordinal covariates are given in Tables 5.4 - 5.5. The pattern of missing covariates and the number of DLBCL patients associated with them are presented in Table 5.6.

Table 5.5: The descriptive summaries of binary and ordinal covariates and percentage of missing data for the SNLG data set ($n = 1391$)

Variable	**Value**	**n (%)**	$\mathbf{n_{miss}}$ (%)
Gender	Male	704 (50.6)	0 (0)
	Female	687 (49.4)	
Stages	I	255 (18.3)	0 (0)
	II	386 (27.7)	

Continued on next page

Table 5.5 – *Continued from previous page*

Variable	**Value**	**n (%)**	**n_{miss} (%)**
	III	354 (25.4)	
	IV	396 (28.5)	
LDH	≤1	365 (26.2)	522 (37.5)
	>1 - 3	339 (24.4)	
	>3	153 (11.9)	
ECOG	0	0 (0)	146 (10.5)
	I	452 (32.5)	
	II	527 (37.9)	
	III	184 (13.2)	
	IV	71 (5.1)	
	V	11 (0.8)	
Albumin	Normal / High	986 (70.8)	97 (7.0)
	Low	308 (22.1)	
AP	Normal	983 (70.7)	78 (5.6)
	High	330 (23.7)	
Urea	Normal	999 (71.8)	51 (3.7)
	High	341 (24.5)	
Extranodal	Present	465 (33.4)	1 (0.1)
	Absent	926 (66.5)	
Bulk	Present	678 (48.7)	109 (7.8)
	Absent	604 (43.4)	
Bone marrow	Present	191 (13.7)	196 (14.1)
	Absent	1004 (72.2)	
Bsy	Present	794 (57.1)	13 (0.9)
	Absent	584 (42.0)	

Table 5.6: The number of missing covariates with the corresponding frequency of DLBCL patients in the SNLG data set

Number of missing covariates	**Frequency**	**Percentage (%)**
0	636	45.72
1	449	32.28
2	188	13.52
3	57	4.10
4	22	1.58
5	18	1.29
6	12	0.86
7	8	0.58
9	1	0.07
Total	1391	100.00

5.4 Lung cancer

5.4.1 Lung cancer- a global and Malaysian perspectives

The worldwide incidence of lung cancer has rapidly risen since the turn of the 20^{th} century. For instance, 228,190 new lung cancer cases and 159,480 lung cancer-associated mortality were estimated for the USA population in 2013. These account for 28% and 26% of cancer-associated mortality for American males and females, respectively (Siegel et al., 2013). In Malaysia, lung cancer was considered rare in the 1950s. The earliest study conducted to estimate lung cancer prevalence in Malaysia was by Marsden (1958) that demonstrated lung cancer was the 8^{th} commonest malignancy in male and not even in the top ten among females. In contrast, according to the Malaysian National Cancer Registry 2006, it had become the 3^{rd} commonest cancer among the Malaysian population after breast and colorectal cancers. The peak incidence of lung cancer is in the 7th decade of life with younger age at diagnosis for non smokers (mean age 54.7) than smokers (mean age 61.7).

There are several putative epidemiological risk factors that have been identified for lung cancer. Smoking remains the predominant risk factor, followed by exposure to environ-

mental pollutants such as nickel, arsenic, asbestos and others. A few genetic susceptibility loci have been identified to contribute towards increased risk of lung cancer among first degree relatives of a proband with lung cancer. Individuals with inflammatory comorbidites such as Chronic Obstructive Pulmonary Disease (COPD), asthma and emphysema are found to have an elevated risk of developing lung cancer whilst hay fever was associated with lower risk of lung cancer.

The lung cancer types can be divided into two large classes; small cell lung cancer (SCLC) which is the more aggressive type, and non-small cell lung cancer (NSCLC). The NSCLC type is further divided into three common subtytpes which are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Of these, squamous cell carcinoma is deemed the more aggressive NSCLC subtype which results in poorer overall survival and progression free survival.

5.4.2 Malaysian advanced non-small cell lung cancer dataset - Hospital Universiti Sains Malaysia (HUSM)

Hospital Universiti Sains Malaysia (HUSM) is a major teaching hospital located in Kota Bharu, the state capital of Kelatan (a Malaysian state that is located near the Thailand and Malaysian border) (Figure 5.3). It is a tertiary medical centre receiving referrals mostly from the Malaysian east coast region. The oncology service is provided by the Department of Nuclear Medicine, Radiotherapy and Oncology which is a spin off from the Department of Medicine since 1995. On the average, it receives 15-25 newly diagnosed primary lung cancer cases per annum.

Based on an initial survey of the hospital records performed in August 2014, there are approximately 600 lung cancer cases which were reported from 1996 until 2014. So far, the data collection process is still ongoing and the data set now has 397 advanced lung cancer cases. All of them received a platinum-based regimen as the first-line palliative chemotherapy regime. Nevertheless, the data gathered so far should be independently verified to ascertain the accuracy of the data.

Based on initial analyses, there are 241 (60.7%) patients that had experienced the event of interest (death) and 156 (49.3%) censored observations. There are 157 (39.55%) patients with complete covariate information. The covariates and measured outcomes of interest are basically similar to the SNLG dataset except that there are several additional

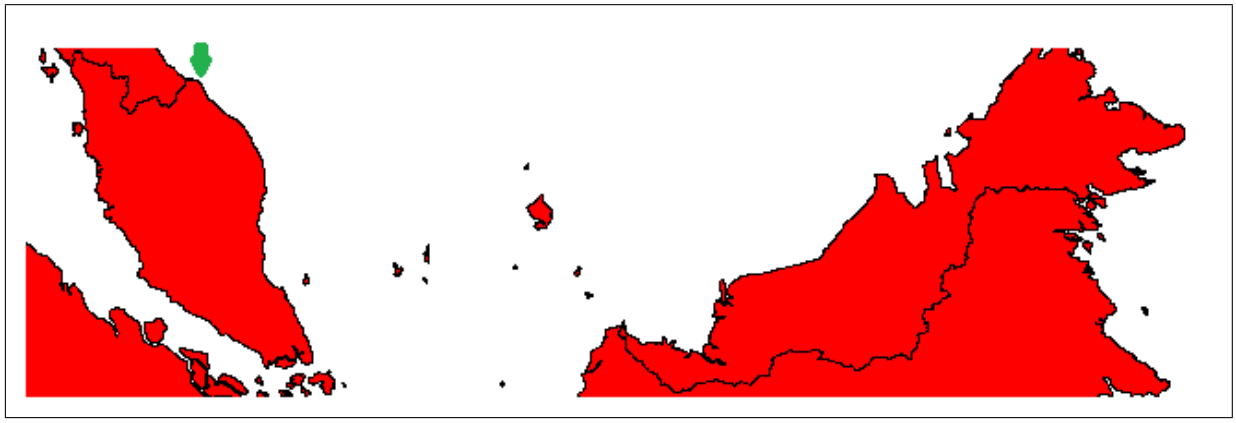

Figure 5.3: Malaysian map showing the location of HUSM (green arrow)

covariates and **Ann Arbor stage, extranodal disease, bulky mass, bone marrow involvement, B symptoms and urea** were excluded since these only pertain to the NHL prognostic index. Furthermore, stage was categorised into either IIIB (locally advanced unresectable lung cancer, coded as -1) or IV (metastasised to other organs, coded as +1) since all lung cancer patients were at terminal stages. Besides, only advanced lung cancer patients with ECOG score of 0-2 were included in the data set since higher ECOG scores excluded the patients from receiving palliative chemotherapy and this would definitely impact patient survival. The operational definitions for the eight additional covariates and two other covariates (stage and histology) that have classification system different from those used for the SNLG data set are given as follows.

- **Race** - The ethnicity of patients, recorded either 1 or 2 as Malays and non-Malays respectively.
- **Smoking** - This is a categorical covariate representing the smoking status of the advanced lung cancer patients at the time of diagnosis. It is recorded as 1, 2, 3 for never smokers, ex-smokers and active smokers respectively. This classification was based on the recommendations by Simonato et al. (2001).
- **Stage** - It is recorded as 1 and 2 which respectively correspond to Tumour, Node, Metastasis (TNM) stage IIIB and IV (Rami-Portas et al., 2017).
- **Weight loss** - It is defined as more than 5% weight loss in comparison to the patient's baseline weight 6 months before the diagnosis of lung cancer (Scott et al., 2002). It is recorded as 1 and 2 which correspond to the absence and presence of pre-treatment weight loss.

- **Brain metastasis** - The presence of lung cancer cells / mass in the brain indicates a widespread disease, indicating poor prognosis (metastasis is defined as the spread of cancer beyond the original / primary site of cancer). This is a categorical covariate coded as : 1 (metastasis absent), 2 (metastasis present).

- **Sodium** - A continuous covariate representing the serum sodium concentration in lung cancer subjects. The usual range is between 135 mmol/L and 145 mmol/L (Longmore et al., 2014). A low serum sodium concentration indicates the presence of Syndrome of Inappropriate ADH (SIADH) secretion, signifying poor prognosis.

- **Neutrophil count** - A continuous covariate that is measured in $\times 10^9$ unit. The normal range is between 2.5 and 7.5 $\times$ 10^9 cells/L (Longmore et al., 2014). A high neutrophil count indicates the presence of intense inflammation inside the tumour environment, indicating poor prognosis.

- **Platelet count** - A continuous covariate. Platelets are a type of blood cell that is responsible for the formation of blood clot. The normal range is between 150 and 450 $\times 10^9$ cells/L (Longmore et al., 2014). It has been documented that the presence of cancer cells induces a hypercoagulable state resulting in an increase in the platelet count. Both neutrophil and platelet counts are modelled as continuous covariates.

- **Histology** - This is a categorical covariate representing the types of lung cancer. There are four categories within this covariate which are coded as follows: **1 - Adenocarcinoma , 2 - Squamous cell carcinoma, 3 - Large cell carcinoma, 4 - Small cell carcinoma**.

- **Epidermal Growth Factor Receptor (EGFR) mutational status** - The presence of the EGFR mutation in lung cancer cells results in a higher response rate to molecular targeted therapy such as gefitinib or erlotinib. Consequently, the survival in those with these genetic features is superior to those without this mutation. It was estimated by Pao et al. (2004) that approximately 35% of lung cancer subjects of East Asian origin harbour this genetic mutation. Hence, this categorical covariate should ideally be included in the development of the prognostic index for advanced lung cancer patients. The data are recorded as 1 for positive EGFR mutation and -1 for absent EGFR mutation.

On further analysis, EGFR mutational status was excluded due to unreliable reporting by the HUSM pathology department. As a result, EGFR was no longer considered in the analysis.

The descriptive summaries of numerical, binary and ordinal covariates of the Malaysian advanced lung cancer data set are given in Tables 5.7 - 5.8. The missing covariate patterns and the corresponding frequency of advanced lung cancer patients are presented in Table 5.9.

Table 5.7: The descriptive summaries of numerical covariates and percentage of missing data for the Malaysian advanced lung cancer data set ($n = 397$)

Variable	**Min**	**Max**	**Mean (SD)**	$\mathbf{n_{miss}}$ **(%)**
Age	41	88	58.9 (10.92)	83 (20.9)
Haemoglobin	72	180	122.9 (21.77)	93 (23.4)
WBC	1.3	11.4	6.2 (2.01)	95 (23.9)
Sodium	125	142	136.0 (3.6)	76 (19.4)
Neutrophil count	0.60	7.60	4.2 (1.40)	95 (23.9)
Platelet count	33.0	882.0	369.8 (139.53)	77 (19.4)

Table 5.8: The descriptive summaries of binary and ordinal covariates and percentage of missing data for the Malaysian advanced lung cancer data set ($n = 397$)

Variable	**Value**	**n (%)**	$\mathbf{n_{miss}}$ **(%)**
Gender	Male	207 (52.1)	0 (0)
	Female	190 (47.9)	
Race	Malay	276 (69.5)	96 (24.2)
	Non-Malay	25 (6.3)	
Smoking	Never smokers	36 (9.1)	83 (20.9)
	Ex-smokers	121 (30.5)	
	Active smokers	157 (39.5)	
Stage	IIIB	127 (32.0)	85 (21.4)

Continued on next page

Table 5.8 – *Continued from previous page*

Variable	**Value**	**n (%)**	**n_{miss} (%)**
	IV	185 (46.6)	
ECOG	0	44 (11.3)	71 (17.9)
	I	196 (49.4)	
	II	86 (21.7)	
LDH	≤1	88 (22.2)	95 (23.9)
	>1 - 3	214 (53.9)	
Albumin	Normal	117 (29.5)	107 (27.0)
	Low	173 (43.6)	
AP	Normal	159 (40.1)	91 (22.9)
	High	147 (37.0)	
Weight loss	Present	208 (52.4)	83 (20.9)
	Absent	106 (26.7)	
Brain metastasis	Present	104 (26.2)	95 (23.9)
	Absent	198 (49.9)	
Histology	Adenocarcinoma	175 (44.1)	87 (21.9)
	Squamous	89 (22.4)	
	Large cell	22 (5.5)	
	Small cell	24 (6.0)	

Table 5.9: The missing covariate patterns for the Malaysian advanced lung cancer data set

Number of missing covariates	**Frequency**	**Percentage (%)**
0	157	39.55
1	2	0.50
2	5	1.26
3	7	1.76
4	22	5.54
5	38	9.57
6	43	10.83
7	37	9.32
8	34	8.56
9	26	6.55
10	15	3.78
11	10	2.52
12	1	0.25
Total	397	100.00

5.5 Summary

In this section, the detailed characteristics of each applied dataset have been described. We shall elaborate on the prior information used in Chapter 6.

Chapter 6

Prior Information and Prior Distribution Construction

6.1 Introduction

In this chapter, the construction of prior distribution will be described for modelling purposes. In 6.2, a short introduction on the conceptual underpinnings of prior elicitation is given. In 6.3 we shall describe and summarise the relevant information obtained from the previous literature for prior constructions for the SNLG data and present the meta analysis results. In 6.4, the information relevant to prior construction for the Malaysian HUSM advanced lung cancer survival model will be presented along with meta analysis results. In 6.5 and 6.6, prior elicitation and construction for Weibull and piecewise constant hazard survival models will be presented, including both hierarchical and autoregressive priors in the piecewise case. In section 6.7, we shall conclude our findings for this chapter.

6.2 A short introduction to prior elicitation

6.2.1 Preliminaries

The purpose of prior elicitation is to obtain and quantify information about uncertain quantities in a probabilistic construction based upon the knowledge acquired from the experts in the respective areas. In other words, the main objective of prior elicitation is to appropriately capture the experts' knowledge and represent it in proper probability distributions. To make things clearer, we shall define experts as individuals whose know-

how and knowledge are worth having for constructing coherent priors (O'Hagan et al., 2006).

In statistics, there are two circumstances where eliciting prior information is important. They are design of experiments (DOE) and Bayesian statistics (O'Hagan et al., 2006). In the former, prior elicitation is important to identify variables whose uncertainty is of considerable magnitude. This is to ensure that the experts' prior knowledge will assist in designing efficient experiments. In the latter, however, prior information is used to augment or supplement the information contained in the data. It is the case, nevertheless, that in Bayesian statistics the prior information may not be very influential since the information contained in the data far outweighs the information represented by the prior. Hence, methodical or systematic prior elicitation will usually be performed if the experts' prior is noticeably very informative or when the amount of information conveyed by the data is limited.

6.2.2 Types of uncertainty

According to O'Hagan et al. (2006), uncertainties can be classified into two groups: **aleatory** and **epistemic** uncertainties. The term aleatory is derived from the Latin *alea*, die, or *aleator*, dice player. Hence, we can straightforwardly deduce that aleatory uncertainty is attributed to uncertainties that arise from randomness in a process. On the other hand, epistemic uncertainty originates from experts' incomplete mechanistic knowledge of a process (the word epistemic is derived from a Greek word, *episteme*, which means knowledge). To better distinguish the differences between these two, we shall consider the following example.

Let us consider an example where we model survival times for some homogeneous group of patients using an exponential (λ) distribution. Beliefs about the survival time, T, for a future patient will involve aleatory uncertainty because, even if we know the value of λ, T is still a random variable, and epistemic uncertainty because we do not know the value of λ. Hence, for example, $\mathrm{Var(T)} = \mathrm{E}_\lambda[\mathrm{Var\ (T|\ }\lambda)]+\mathrm{Var}_\lambda[\mathrm{E}(T \mid \lambda)]$. In this case, the problem which elicitation methods need to overcome is how to separate the epistemic uncertainty, masured by $\mathrm{Var}(\lambda)$, from the aleatory uncertainty, measured by $\mathrm{Var(T|}\ \lambda)$ (and $\mathrm{Var}_\lambda[\mathrm{E(T} \mid \lambda)]$), which both contribute to Var(T). For Bayesian statisticians, both epistemic and aleatory components of uncertainty can be represented by probability distributions.

6.3 Prior information for the Non-Hodgkin Lymphoma data set (SNLG)

6.3.1 Information from previous literature

To search for relevant literature that provides information for prior construction for SNLG model parameters, we used the Patient (or Population), Intervention (or Prognostic Factor), Comparison and Outcome (PICO) search strategy (Schardt et al., 2007). Search engines and databases such as Google Scholar, PubMed and Scopus were utilised to obtain the relevant literature using the following keywords: Non Hodgkin Lymphoma, survival, prognostic factor, prognosticators, survival determinants. Any article that was not available online was manually searched and obtained from the university library (the Walton Medical Library). Overall, 35 studies were reviewed to obtain the necessary information for constructing the prior distributions of the survival parameters and 12 of them had to be discarded due to the following reasons: 1) inadequate information on precision; 2) questionable research methodology and authenticity of data collection process; 3) unsatisfactory sample heterogeneity in unmeasured but important prognostic factors. Only 23 studies whose results are deemed acceptable were used for prior construction for the model parameters. Their hazard ratios and 95% confidence intervals (CIs) are given as follows:

- **Troppan et al. (2014)**: A retrospective analysis of 290 Austrian patients diagnosed with DLBCL between June 2004 and April 2013 at the Division of Hematology, Medical University of Graz, Austria. Cox proportional hazard model was utilised to obtain the hazard ratios and their respective 95% confidence intervals for the relevant survival parameters:
  - Age $>$ 60 years old: 3.52 (95% CI: 1.48, 3.58) [Multiple Cox model]
  - Ann Arbor stage III and IV vs Ann Arbor Stage I and II: 2.38 (95% CI: 1.21, 4.68) [Multiple Cox model]
  - LDH $>$200 IU: 1.62 (95% CI: 0.83, 3.16) [Simple Cox model]
- **Zhao et al. (2014)**: Analyses of 1650 DLBCL patients diagnosed between June 2000 and December 2010 diagnosed at 7 National Comprehensive Cancer Network

(NCCN) research centres, USA. The patients were followed up until the end of December 2011. The hazard ratios of the following survival parameters were obtained using multiple Cox PH model:

  - LDH ratio (LDH-R): $\leq$1: 1.0, >1-3: 2.1 (95% 1.9, 2.7), > 3: 3.3 (95% CI: 2.3, 4.8)
  - ECOG performance status: 1-2 = 0, $\geq$2: 1.9 (95% CI: 1.5, 2.4)
  - Ann Arbor stage: Stage III and IV = 1.5 (95% 1.1, 2.0)
  - Extranodal disease: 1.5 (95% CI: 1.2 , 1.9)

- **Slymen et al. (1990)**: A total of 105 DLBCL patients diagnosed at the Arizona Cancer Center, USA between 1978 and 1987 were included in the analyses which is based on Cox PH model

  - Ann Arbor Stage III and IV: 2.63 (95% CI: 1.11, 6.25)
  - B Symptoms: 1.98 (95% CI: 0.99, 3.95)

- **Li et al. (2012)**: A total of 437 Chinese patients who were diagnosed with DLBCL between January 2001 and May 2010 at 6 Shanghai hospitals, China were included in the analysis.Analyses were based on Cox PH model. Separate regression coefficients were obtained for rituximab (R-CHOP) and non-rituximab-treated (CHOP) patients:

  R-CHOP group:

  - Age >60 years: 1.011 (95 % CI: 0.987, 1.037)
  - Extranodal sites >1: 1.572 (95% CI: 1.059, 2.334)

  CHOP group:

  - Age >60 years: 1.017 (95 % CI: 0.999, 1.034)
  - Extranodal sites >1: 1.172 (95% CI: 0.915, 1.502)

- **Flowers et al. (2013)**: In total, 701 DLBCL patients (533 Caucasians, 144 African Americans) who were recipients of medical care at 2 research centers (Emory University and University of Alabama at Birmingham, USA) between 1981 and 2010 were included in this retrospective analysis. Cox PH method were utilised to obtain the hazard ratios and 95% CIs for each relevant prognostic factor for DLBCL

  - Age> 60 years: 1.53 (95% CI: 1.17, 2.00)
  - Gender (male) : 1.23 (95% CI: 0.97, 1.57)
  - ECOG Performance status (≥ 2): 2.62 (95% CI: 1.97, 3.49)
  - Elevated LDH level: 1.59 (95 % CI: 1.15, 2.21)
  - No of extranodal sites involvement (≥ 2): 0.92 (95% CI: 0.66, 1.28)
  - B symptoms: 1.20 (95% CI: 0.92, 1.58)

- **Chen et al. (2012)**: DLBCL exclusive samples ($n = 100$). Period of recruitment: Jan 2000-Dec 2009. Single institution (Taiwan). All patients were treated with rituximab-containing regime. Three variables are significant on Multiple Cox PH model and an additional four are found to be important prognosticators for overall survival on Simple Cox PH model:

  Multiple Cox PH:

  - ECOG performance ≥ 2 : 4.936 (95% CI: 1.818, 13.403)
  - LDH (defined as $\geq$ 206 IU\L): 3.400 (95%CI: 1.155, 10.007)

  Simple Cox PH:

  - Haemoglobin (defined as $\geq$120g/dL: 0.518 (95% CI: 0.239, 1.120)
  - Albumin (<37 g / dL): 2.451 (95% CI: 0.471, 12.821)
  - B symptoms: 2.952 (95% CI: 1.296, 6.726)
  - Bone marrow involvement: 4.100 (95% CI: 1.526, 11.015)

- **Song et al. (2010)**: A total of 96 Korean patients who were diagnosed with primary extranodal DLBCL in 3 hospitals between April 2003 and October 2008. They received 6 to 8 cycle of R-CHOP as their therapeutic regime. The only reported hazard ratio reported is

  - Tumor bulk ($\geq$ 7.5 cm, in non GC only): 1.420 (95% CI: 0.542, 2.297)

- **Carella et al. (2013)**: This study involves DLBCL patients who received rituximab-containing therapeutic regime and recruited from 43 Brazillian and Italian Institutions (n = 1793). Cox PH model was used with bootstrapping to accommodate for confidence level error ($n_{bootstrap} = 1000$).

  - Male: 1.60 (95% CI: 1.30, 1.96)

- **Takahashi et al. (2012)**: A retrospective analysis of data on DLBCL patients who were recruited from 47 Japanese institutions and treated with R-CHOP from 2003 until 2006 ($n = 1221$). Multivariate Cox PH model was used to obtain thehazard ratios and their respective %95 confidence intervals for the following prognostic factors

  - Age > 60: 2.9 (95% CI: 1.5, 5.6)
  - Stage III and IV: 1.70 (95% CI: 1.0, 2.9)
  - Elevated LDH: 1.9 (95% CI: 1.1, 3.2)
  - ECOG performance status ≥ 2: 1.7 (95% CI: 0.9, 3.1)

- **Chung et al. (2007)**: A total of 489 DLBCL subjects who were diagnosed at Cross Cancer Institute, Edmonton, Canada between 1986 and 1997 were included in this study. The mean age at diagnosis is 68 years (range: 20-100). Fifty five subjects ($n = 55$) have concordant bone marrow involvement out of a total of 489 subjects. Male subjects predominates the sample (>97%).

  - Concordant bone marrow involvement: 1.87 (95% CI: 1.25, 2.81)

- **Carson et al. (2012)**: US army veterans diagnosed with DLBCL between 1^{st} October 1998 and 31^{st} December 2008 ($n = 2534$). Analyses were performed using Cox PH model in these predominantly rituximab-treated subjects (80.8%). B symptoms is a time varying covariate

  - Age (**continuous scale)**: 1.04 (95% CI: 1.03, 1.04)
  - Stage: II: 0.97 (95 % CI: 0.62, 1.05); III: 1.18 (95% CI: 0.78, 1.76); IV: 2.42 (95% CI: 1.74, 3.36) (0-2 months post DLBCL diagnosis)
  - Stage: II: 1.20 (95 % CI: 0.97, 1.48); III: 1.51 (95% CI: 1.23, 1.84); IV: 1.97 (95% CI: 1.65, 2.35) (> 2 months post DLBCL diagnosis)
  - B symptoms: 0.621 (95% CI: 0.385, 0.859) (0 - 2 months post DLBCL diagnosis)
  - B symptoms: 0.239 (95% CI: 0.104, 0.372) (> 2 months post DLBCL diagnosis)

- **Peyrade et al. (2011)**: A multicentre, single arm, phase II study involving 150 DLBCL patients from France and Belgium. The median age at diagnosis is 83 (range 80-95), with slightly more than a third are man (34%). All were treated with a combination of rituximab with low dose CHOP (R-miniCHOP).
  - Bulky disease (tumour mass $\geq$ 10 cm): 1.4 (95% CI: 0.6, 2.9)
  - Serum albumin ($\leq$35 g / L): 3.2 (95% CI: 1.4, 7.1)
  - Number of extranodal sites $>$ 1: 1.2 (95% CI: 0.6, 2.4)
- **Dalia et al. (2014)**: This study involves a cohort of 124 American patients with DLBCL, recruited between 2007 and 2010 from a single institution (Moffitt Cancer Center, Tampa, Florida). The median age of the subjects is 58 years (range 20-84), with 62% of them are males.
  - Serum albumin ($\leq$ 37 g / dL): 3.846 (95% CI: 6.250, 1.587)
  - Elevated LDH (hazard ratio obtained from Simple Cox PH Model): 2.11 (95% CI: 0.98, 4.52)
  - Extrandodal Disease (hazard ratio obtained from Simple Cox PH Model): 1.28 (95% CI: 0.69.2.35)
- **Frederiksen et al. (2012)**: A retrospective analysis of Danish Non Hodgkin Lymphoma patients whose data were obtained from The Danish National Lymphoma Registry (LYFO) database (similar to our SNLG database). The sample size of this study is 6234.
  - WHO performance status (another name for ECOG performance status): Class I: 1.80 (95% CI: 1.65, 1.95); Class II: 3.31 (95% CI: 2.91,3.76); Class III: 4.23 (95% CI: 3.53, 5.07); Class IV: 7.37 (95% 5.79, 9.39)
  - Ann Arbor Stage: Stage II: 1.10 (95% CI: 0.81, 1.50); Stage III: 1.18 (95% CI: 0.93, 1.49); Stage IV: 1.35 (95% CI: 1.18, 1.53)
  - Elevated LDH: 1.71 (95% CI: 1.55, 1.90)
  - Extranodal lesion: 1.09 (95% CI: 0.98, 1.21)
- **Ngo et al. (2008)**: The only study from South East Asia that provides the hazard ratio estimates for the relevant prognostic factors in DLBCL. In total 279 Singaporean subjects with DLBCL were included in the analysis and most of them were

treated with either R-CHOP or CHOP regimes. Therefore, two different estimates per factor were obtained after subjects were stratified by the types of chemotherapy regimes received.

CHOP group ($n = 183$):

- Age > 60 years: 2.882 (95% CI: 1.793, 4.631)
- LDH (> 380 μ / L): 2.333 (95%CI: 1.181, 4.609)
- Albumin (<37 g / L): 2.294 (95%CI: 1.284, 4.099)
- Ann Arbor stage III and IV: 2.599 (95%CI: 1.526, 4.428)

R-CHOP group ($n = 96$):

- Male: 8.386 (95% CI: 1.848, 38.058)
- Ann Arbor stage III and IV: 6.364 (95% CI: 1.847, 21.929)

- **Ziepert et al. (2010)**: Data from 1062 DLBCL patients treated with chemotherapeutic regime containing rituximab were obtained from three databases of RCTs (MinT, MegaCHOEP and RICOVER-60) and retrospectively analysed.
  - Age(> 60 years): 2.4 (95% CI: 1.7, 3.4)
  - Elevated LDH (based on upper limit of normality used by different trial protocols): 2.2 (95% CI: 1.6, 3.0)
  - ECOG Performance status > 1: 1.8 (95% CI: 1.3, 2.6)
  - Ann Arbor stage III / IV: 1.5 (95% CI: 1.1, 2.2)
  - Extranodal disease >1: 1.3 (95% CI: 0.9, 1.8)

- **Aaldriks et al. (2015)**: Data from a sample of 44 geriatric NHL patients (aged 70 years and above) who were recruited from 1 university and 3 general hospitals in Netherlands. The period of recruitment is between May 2004 and February 2010 were included in the analyses. Tehy were all recipients of R-CHOP based treatment regime. The hazard ratio estimates are
  - Elevated LDH ($\geq$ 250 U / L): 0.92 (95% CI: 0.27, 3.11)
  - Albumin (< 35 g / L): 1.83 (95% CI: 0.81, 4.16)

  - Creatinine (A surrogate measure of renal function that is more specific than urea. Hazard ratio estimate for this parameter is used as a substitute for urea): 1.84 (95% CI: 0.70, 4.86)

- **Yan et al. (1995)**: DLBCL ($n = 60$). Not included since no estimates on hazard ratio for prognostic covariates for DLBCL patients were reported.

- **Oki et al. (2008)**: DLBCL ($n = 221$). One hundred and nineteen ($n = 119$) of NHL patients were treated with CHOP regime whilst the rest ($n = 102$) received R-CHOP. It is a single institution study conducted at Aichi Cancer Centre, Nagoya, Japan).

  - B symptoms: 2.37 (95% CI: 1.16, 4.83)

- **Møller et al. (2003)**: Two hundred and twenty one ($n = 221$) low risk Danish DLBCL patients from Western Denmark covering Jutland and Funen (obtained from the population-based LYFO registry) with IPI score of 0 and who were recipients of a mixture of treatments (anthracycline-based chemotherapy, locoregional radiotherapy or surgery with curative intent) were included in the analyses. The median age at diagnosis is 50 years. The data included in the analysis were restricted to patients who were diagnosed with DLBCL from 1983 until 31$^{\text{st}}$ December 1998. The subjects were then followed-up for three years (the last date of observation: 7$^{\text{th}}$ January 2002 - 134 alive subjects at the end of the observation period).

  - Ann-Arbor stage II vs stage I: 2.7 (95% CI: 1.5, 4.8)
  - Age $>$ 50 years: 1.8 (95% CI: 1.0, 3.3)

- **Cox et al. (2008)**: One hundred and one ($n = 101$) Italian CD20+ aggressive large B cell lymphoma (DLBCL: 89 patients; Primary mediastinal large B cell lymphoma: 12 patients; Developed from or associated with low grade lymphoma: 15 patients) patients receving immunochemotherapy between January 2003 and March 2007 were included in the analysis.

  - ECOG Peformance Status $\geq$2: 5.004 (95% CI: 1.635, 11.754)

- **Maartense et al. (2000)**: A cohort of 1028 Danish patients with NHL of different histologies were included in the final analyses. Old REAL classification system for

NHL was used to identify and group patients with different NHL histologies. The subjects were included prospectively in the registry (NHL registry of the Comprehensive Cancer Center West (CCCW), Netherlands) from June 1981 until December 1989. Hazard ratio estimates were computed using Esteve's relative survival probability model (Esteve et al., 1990).

  - Elevated LDH: 2.39 (95% CI: 1.93, 2.96)
  - Male: 1.235 (95% CI: 1.010, 1.515)

- **Advani et al. (2010)**: Two hundred and sixty seven ($n = 267$) DLBCL subjects aged 60 years and older (median age: 69 years (range: 60-92)) and randomised to R-CHOP arm in the US E4494 intergroup trials (ECOG4494, CALGB9793) were included in the analyses. Cox PH model was used and model fit was assessed using AIC. Concordance index (area under ROC) was used as a measure of discrimination.

  - Elevated LDH: 2.1 (95% CI: 1.4, 3.1)
  - Extranodal sites >1: 2.3 (95% CI: 1.6, 3.2)

- **Hirakawa et al. (2010)**: A retrospective analysis of 152 Japanese DLBCL patients treated with R-CHOP like primary and maintenance therapies. The problem with this study is the results reported as odds ratio (not hazard ratio) despite Cox PH model was used for analysis of overall survival (OS) of this cohort. Only one prognostic factor that is relevant to our study is included in here since no data on other prognostic factors deemed important by the findings of this study such as relative dose intensity of less than 70%, febrile neutropenia and prophylatic use of G-CSF were recorded in our dataset.

  - Albumin (< 35 g / L): 2.397 (95% CI: 0.793, 7.246)

- **Hayward et al. (1991)**: This study is a retrospective analysis of 972 subjects ($n = 972$) with NHL (the early SNLG cohort). The main problem with this study is the heterogeneous subsets of patients were included in the analysis. Besides, the hazard ratios for prognostic factors influencing the overall survival of NHL patients were given but no 95% CI were reported.

  - B symptoms: 0.42
  - Leukocytes: 0.42

  - Ann Arbor stage III and IV: 0.29
  - Age - **continuous**: 0.02

- **Christina et al. (2013)**: This study was conducted among transformed lymphoma patients (from other types of lymphoma such as follicular lymphoma, marginal zone and MALT lymphoma to DLBCL) ($n = 81$) who were identified from the Moffitt Cancer Center Total Cancer Care Database, Tampa, Florida between January 2001 and December 2011. The median age at diagnosis is 60 years, with male to female ratio of 0.72 (58 females). The median time to DLBCL transformation is 3.4 years.
  - B symptoms: 3.1 (95% CI: 1.5, 6.4)
  - Elevated LDH: 2.6 (95% CI:1.0, 6.6)

6.3.2 Meta analysis results for Non-Hodgkin Lymphoma prognostic factors

Ideally, standard Bayesian meta analysis technique should be used to synthesise the results. However, our purpose of performing this meta analysis was solely for obtaining guidance for the choice of prior distributions for the regression coefficient for each Non Hodgkin Lymphoma prognostic factor. Therefore a random effect meta analysis model and an empirical Bayesian estimator were used to account for the heterogeneity of study results. For more information, refer to Viechtbauer (2005, 2010). Meta analysis was performed using the `metafor` package version 2.0.0 (Viechtbauer, 2017) which was implemented in R. The forest plots for each NHL prognostic factors are given in Figure 6.1. The mean and variances for each parameter based on the meta analysis results are presented in Table 6.1.

6.4 Prior information for the Malaysian-HUSM Lung Cancer data set

6.4.1 Information from previous literature

In total, 19 prior studies have been identified as having adequate information on the estimates of the regression coefficients for each prognostic factor for lung cancer. The

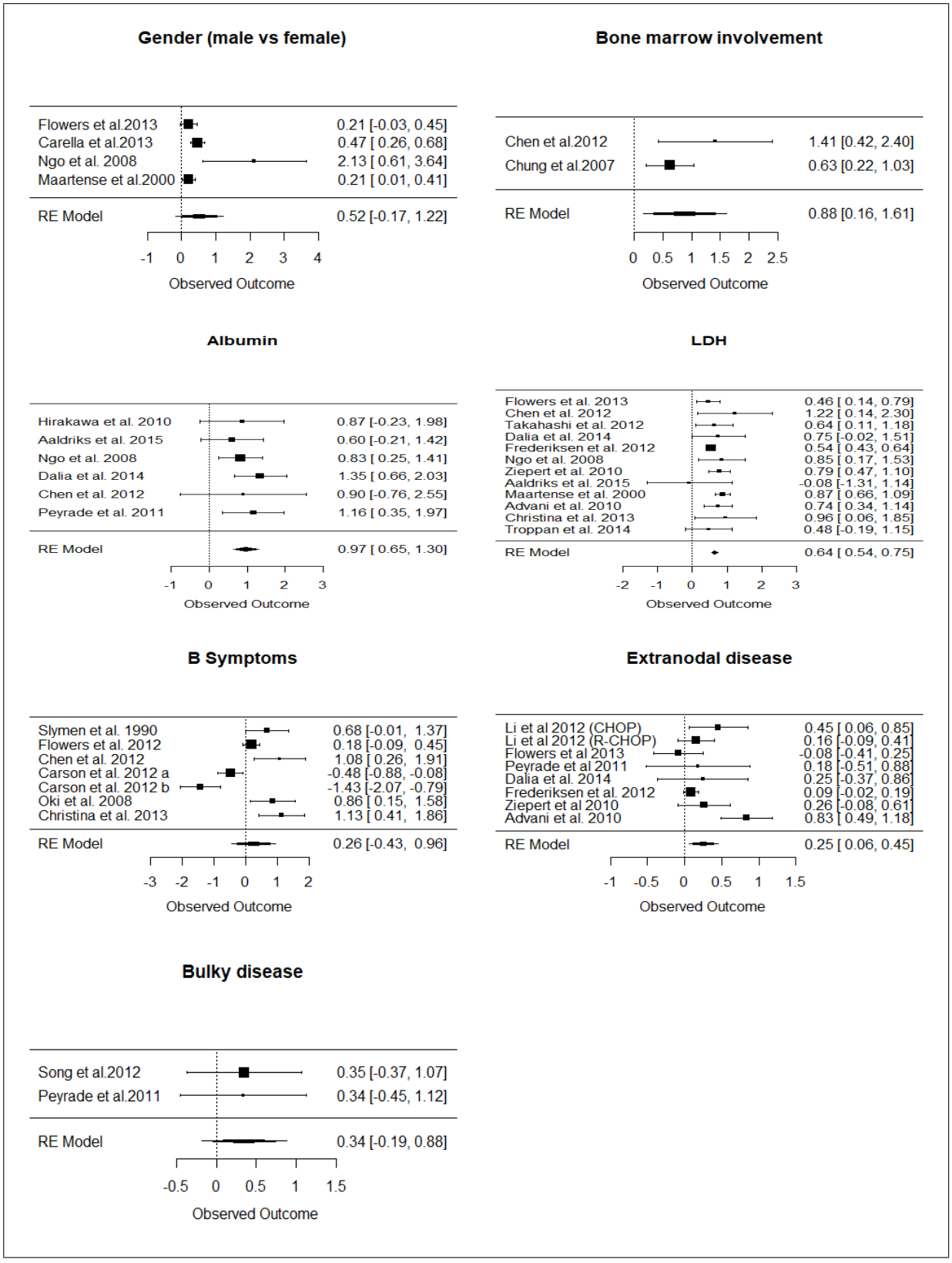

Figure 6.1: Forest plots of NHL prognostic factors

Table 6.1: The synthesised mean and variances for each regression coefficients for survival predictors of Non-Hodgkin Lymphoma obtained from meta analyses

Parameters	**Mean**	**Variances**
β_{Gender}	0.52	0.1258
β_{Marrow}	0.88	0.1359
$\beta_{Albumin}$	0.97	0.0280
β_{LDH}	0.64	0.0026
$\beta_{B-symptoms}$	0.26	0.1266
$\beta_{Extranodal}$	0.25	0.0101
β_{Bulky}	0.34	0.0738

medical literature was searched using the PICO method (Schardt et al., 2007) for relevant literature using the following keywords: advanced lung cancer, prognostic factors, prognosticators, survival determinants. The search engines and databases used were Google Scholar, PubMed, and Scopus databases. Older articles that were not available online were manually searched at the university library (Walton Medical Library, Newcastle University).

We initially found 30 studies that were relevant for extracting information on prognostic factors. However 11 studies had to be excluded for the following reasons: 1) inadequate reporting of measure of precision; 2) poor research methodology; 3) article retraction due to unethical misconduct such as plagiarism and data falsification. The studies that are finally included are as follows:

- **Hespanhol et al. (1995)**: This study involves 411 patients with advanced non-small cell lung cancer (NSCLC) in Portugal, recruited between 1984 and 1990. Twenty one clinical, anatomical, haematological and biochemical parameters were assessed and the following were found to be statistically important prognostic factors for lung cancer survival based on multivariate Cox model.
  - Albumin (normal vs low): 0.588 (95% CI: 0.463,0.747)
  - Weight loss: 1.624 (95% CI: 1.283, 2.056)
  - Stage (IV vs IIIB): 1.576 (95% CI:1.445, 2.310)
  - LDH (high vs low): 1.268 (95% CI:1.010, 1.592)

  - Gender (male vs female): 1.410 (95% CI: 1.048, 1.896)

- **Simmons et al. (2015)**: This study involves 319 patients with advanced NSCLC at 2 university hospitals in Greece (University Hospitals of Heraklion and Larissa). The patients were recruited between September 2006 and February 2010. The hazard ratios for the following prognostic factors were found to be statistically important for lung cancer survival based on Cox model:

  - Gender (male vs female): 1.67 (95% CI: 1.14, 2.34)
  - Weight loss: 1.49 (95% CI: 1.18, 1.88)

- **Leung et al. (2012)**: This study involves 261 patients with surgically inoperable NSCLC at Wishaw General Hospital, Lanarkshire, UK. The patients were recruited and prospectively followed-up from May 2001 to November 2004. The hazard ratios for the following prognostic factors were found to be important for lung cancer survival based on Cox model:

  - Gender (male vs female): 1.09 (95% CI: 0.85, 1.41)
  - Weight loss (yes vs no): 0.86 (95% CI: 0.58, 1.28)
  - WBC:1.23 (95% CI: 0.93, 1.61)
  - Stage (IV vs III): 1.56 (95% CI: 1.21, 2.01)

- **Maeda et al. (2000)**: Data on two hundred and sixty one ($n = 261$) Japanese patients with advanced NSCLC , collected by Okayama Lung Cancer Study Group, Japan between 1978 and 1992, were retrospectively analysed in this study. The hazard ratios for the following prognostic factors were considered important for advanced lung cancer survival based on Cox model:

  - Albumin (normal vs low): 1.69 (95% CI: 1.19, 2.41
  - Stage (IV vs IIIB): 1.71 (95% CI: 1.27, 2.30)

- **Hsu et al. (2012)**: One hundred forty four ($n = 144$) Taiwanese patients with advanced NSCLC were recruited from National Taiwan University Hospital between January 2000 and February 2009. The median survival time of this study cohort was 14.7 months. The hazard ratios for the following prognostic factors were considered important for advanced lung cancer survival based on Cox model:

  - Gender (male vs female): 1.70 (95% CI: 1.08, 2.68)
  - Smoking status (ever smoker vs non smoker): 0.96 (95% CI: 0.65, 1.41)
  - Weight loss: 2.72 (95% CI: 1.39, 5.30)
  - Haemoglobin (low vs normal): 2.08 (95% CI: 1.15, 3.77)
  - WBC: 1.62 (95% CI: 0.94–2.78)
  - Platelet (high vs normal): 1.70 (95% CI: 1.00, 2.90)
  - Albumin (low vs normal): 3.26 (95% CI: 1.69, 6.30)
  - Stage (IV vs IIIB): 2.62 (95% CI: 1.50, 4.57)

- **Ye et al. (2014)**: Data were collected from 300 Chinese patients with advanced NSCLC attending 40 clinical sites in Mainland China as part of the B9E-AA-B004 clinical trial. The median overall survival time was 15.6 months (95% CI: 12.5, 17.4). The hazard ratios for the following prognostic factors were considered important for advanced lung cancer survival based on Cox model:

  - Gender (male vs female): 1.211 (95% CI: 0.847, 1.733)
  - Smoking status (Never smoker vs ever smoker): 1.08 (95% CI: 0.779, 1.50)

- **Svaton et al. (2014)**: Five hundred and forty four ($n = 544$) Czech patients with advanced stage NSCLC who were treated with Erlotinib at the Department of Pneumology and Phthisiology, Pilsen University Hospital, Czech Republic between 2006 and 2013 were included in this study. The following factors were found to be important determinants of advanced lung cancer survival in this study cohort and their hazard ratios are:

  - Gender (male vs female): 1.25 (95% CI: 0.98, 1.58)
  - Smoking status (Ever smoker vs never smoker): 0.96 (95% CI: 0.70, 1.32)
  - Stage (IV vs IIIB): 1.49 (95% CI: 1.12, 1.97)
  - Sodium (low vs normal): 1.87 (95% CI: 1.47, 2.39)

- **O'Mahony et al. (2016)**: Sixty two ($n = 62$) patients with advanced stage NSCLC who received treatments at Rush University Medical Centre (RUMC), Chicago, USA were recruited into this study. Patient's age, which was recorded as a continuous variable, was found an important determinant for overall survival and the hazard ratio is:

  - Age (continuous): 0.94 (95% CI: 0.99, 1.09)

- **Mandrekar et al. (2006)**: A data set containing the relavant determinants of lung cancer survival were obtained from 1056 patients with advanced stage NSCLC pooled from 9 North Central Cancer Treatment Group (NCCTG) trials in the USA. The patients were recruited between 1985 and 2001. The hazard ratios for the important determinants of overall survival in this study cohort are as follows:

  - Gender (male vs female): 0.99 (95% CI: 0.84, 1.17)
  - Stage (IV vs IIIB): 2.61 (95% CI: 2.04, 3.34)
  - Weight loss (yes vs no): 1.82 (95% CI: 1.29, 2.60)
  - WBC (high vs normal): 1.43 (95% CI: 1.22, 1.67)
  - Haemoglobin (low vs high): 1.51 (95% CI: 1.28–1.78)

- **Park et al. (2009)**: In total, 316 patients with recurrent NSCLC receiving salvage gefitinib treatment at Samsung Medical Centre (SMS), Seoul, Korea between January 2002 and December 2005 were included in this study. Their clinical and laboratory test results were obtained from the SMS database. The median overall survival of this study cohort is 6.4 months and the researchers determined the following prognostic factors were relevant for predicting lung cancer overall survival:

  - Gender (male vs female): 1.51 (95% CI: 1.18, 1.94)
  - Smoking status (current smoker vs never smoker): 1.34 (95% CI: 1.02, 1.75)
  - Stage (IV vs IIIB): 1.07 (95% CI: 0.80, 1.43)
  - Brain metastasis (yes vs no): 1.13 (95% CI: 0.85, 1.50)
  - ALP (high vs low): 1.50 (95% CI: 1.13, 2.00)
  - Albumin (low vs normal): 1.45 (95% CI: 1.10, 1.93)
  - WBC (high vs normal): 1.38 (95% CI: 1.03, 1.86)

- **Du et al. (2013)**: Overall, 258 advanced NSCLC patients who were diagnosed at Kaifeng second people's hospital, Henan, China between March 2001 and September 2011 were included in this study cohort. The median overall time is 20 months and only high platelet count is a significant prognosticator for overall survival

  - Platelet (high vs low): 4.15 (95% CI: 3.09, 5.59)

- **Kim et al. (2014)**: Data on 854 patients with inoperable advanced NSCLC diagnosed at the Seoul University College of Medicine between January 2007 and September 2009 were retrospectively analysed in this study. The median overall time is 340.5 days (i.e approximately a year) in those with high platelet counts and 569.5 days in those with normal platelet levels. The important prognostic factors identified in this study and their hazard ratios are:
  - Gender (male vs female): 1.55 (95% CI:1.30, 1.84)
  - Stage (IV vs IIIB - **estimated**): 1.98 (95% CI: 1.45, 2.69)
  - Platelet (high vs low): 1.51 (95% CI: 1.14, 2.00)
- **Ulas et al. (2014)**: This study involves data which were retrospectively collected from 462 advanced NSCLC patients who were treated at the Department of Medical Oncology, Oncology Teaching and Research Hospital, Ankara, Turkey between 2000 and 2010. In this study cohort, the median overall survival time is 11 months with a median follow-up time of 44 months. The important determinants of overall survival for this study cohort are:
  - Gender (male vs female): 1.18 (95% CI: 0.86, 1.63)
  - Smoking status (Ever smoker vs non smoker): 1.16 (95% CI: 0.87, 1.55)
  - Brain metastasis (yes vs no): 1.56 (95% CI: 1.26, 1.98)
  - Haemoglobin (low vs normal): 1.13 (95% CI: 0.92, 1.39)
  - WBC (high vs normal): 1.55 (95% CI: 1.27, 1.89)
  - Platelets (high vs normal): 1.11 (95% CI: 0.88, 1.39)
  - ALP (high vs normal): 1.51 (95% CI: 1.16, 1.97)
  - LDH (high vs normal): 1.31 (95% CI: 1.00, 1.70)
  - Albumin (low vs normal): 1.28 (95% CI: 0.98, 1.67)
- **Zhang and Ran (2015)**: In total, 773 NSCLC patients at advanced stages who were diagnosed and treated at Kimmel Cancer Centre, Thomas Jefferson Hospital, Greater Philadelphia area, USA between 1998 and 2011 were included as this study cohort. The median overall survival time is 7.0 months in the training cohort. The important prognostic factors determined from this study and their hazard ratios are:

  - Age (continuous): 1.005 (95% CI: 0.997, 1.013)
  - Stage (IV vs IIIB): 1.53 (95% CI: 1.24, 1.89)

- **Lim et al. (2018)**: In total, 217 terminal stage NSCLC patients with malignant pleural effusion who were diagnosed and treated at 6 hospitals in Korea between January 2012 and July 2016 were included in this study. The median overall survival time is 16.2 months and nearly two thirds of the patients (62.6%) received conventional chemotherapy regimens as the first line of treatment whilst the rest were recipients of targeted therapies directed against either Epithelial Growth Factor Receptor (EGFR) or Anaplastic Lymphoma Kinase (ALK) translocational mutations. The hazard ratios for the major determinants of overall survival in this study cohort are:

  - Gender (male vs female): 1.620 (95% CI: 1.057, 2.481)
  - Smoking status (ever smoker vs never smoker): 1.166 (95% CI: 0.971, 1.399)
  - Platelet (high vs normal): 1.523 (95% CI: 1.059, 2.191)
  - Haemoglobin (low vs normal): 1.245 (95% CI: 0.810, 1.914)
  - Albumin (low vs normal): 1.724 (95% CI: 1.039, 2.861)

- **Matsunuma et al. (2014)**: In total, 69 terminal stage NSCLC patients who received palliative care at the Palliative Care Unit of Komatsu Municipal Hospital, Ichikawa, Japan in 2012 were included in this retrospective study. The median overall survival time is 30 days, a very short survival time due to the terninal stage of illness. The important prognostic factors determined from this study and their hazard ratios are:

  - Sodium (low vs normal): 2.17 (95% CI: 1.01, 4.68)
  - Albumin (low vs normal): 2.37 (95% CI: 1.05, 5.36)

- **Fiala et al. (2016)**: On the whole, clinical and laboratory data on 457 advanced stage NSCLC patients treated with erlotinib (a type of targeted therapy directed against EGFR mutation) who were diagnosed and treated at Faculty Hospital, Pilsen, Czech Republic between between 2005 and 2014 were retrospectively analysed. The median overall survival time is 10.0 months. The hazard ratios of important determinants of lung cancer survival in this study cohort are:

  - Gender(male vs female): 1.08 (95% CI: 0.87, 1.33)
  - Smoking status (ever smoker vs never smoker): 1.16 (95% CI: 0.86, 1.56)
  - Stage (IV vs IIIB): 1.29 (95% CI: 1.01, 1.64)
  - Albumin (low vs normal): 1.66 (95% 1.16, 2.38)

- **Teramukai et al. (2009)**: In total, 338 advanced stage NSCLC patients who hadn't received any chemotherapy and received usual care between March 2001 and April 2005 from 45 institutions which were under the Japan Multinational Trial Organisation LC00-03 were were included in this study. They then received two separate regimens of chemotherapy. The hazard ratios of relevant overall survival determinants are:

  - Gender (male vs female): 1.35 (95% CI: 0.98, 1.85)
  - Smoking status (current smoker vs never smoker): 1.56 (95% CI: 1.18, 2.06)
  - Weight loss (yes vs no): 1.30 (95% CI: 0.96, 1.76)
  - Stage (IV vs IIIB): 1.24 (95% CI: 0.88, 1.75)
  - LDH (high vs normal): 1.57 (95% CI: 1.20, 2.05)

- **Schad et al. (2018)**: This is a non-controlled cohort study involving 158 stage IV NSCLC patients receiving mistletoe as an adjunct to conventional chemotherapy regimes. The study participants were recruited from multiple centres in Germany between February 2010 and June 2016 based on the data available from the German Network Oncology registry. The median overall survival times were respectively 17.0 and 8.0 months in those receiving conventional chemotherapy and mistletoe and conventional chemotherapy alone. The hazard ratios of the determinants of lung cancer survival are as follows:

  - Age (continuous): 1.00 (95% CI: 0.98, 1.03)
  - Gender (male vs female): 1.65 (95% CI: 1.04, 2.62)
  - Smoking status (current smoker vs never smoker): 1.41 (95% CI: 0.62, 3.22)

- **Li et al. (2015)**: This is a retrospective analysis of a data set containing clinical and pathological information on 64 stage IV NSCLC patients treated in Beijing hospitals from August 2010 to July 2015. No detailed information was given on the types of treatments received by the patients. The hazard ratios of the determinants of lung cancer survival for this study cohort are as follows:

- Gender (male vs female): 0.71 (95% CI: 0.345, 1.424)
- Brain metastasis (yes vs no): 1.754 (95% CI: 0.844, 3.647)
- ALP: 1.957 (95% CI: 1.082, 3.542)

6.4.2 Meta analysis results for Malaysian advanced lung cancer prognostic factors

Ideally, standard Bayesian meta analysis techniques should be used to synthesise the results. However, our purpose in performing this meta analysis was solely for obtaining guidance for the choice of prior distribution for the regression coefficient for each advanced lung cancer prognostic factor. Therefore a random effect meta analysis model was used and an empirical Bayesian estimator was utilised to account for the heterogeneity of study results. For more information, refer to Viechtbauer (2005, 2010). Meta analysis was performed using the metafor package version 2.0.0 (Viechtbauer, 2017) which was implemented in R. The forest plots for the NHL prognostic factors are given in Figures 6.2 and 6.3. The mean and variance for each parameter based on the meta analysis results are presented in Table 6.2.

Table 6.2: The mean and variances for each advanced lung cancer parameters obtained from meta analyses

Parameters	**Mean**	**Variances**
β_{Age}	0.64	0.0140
β_{Gender}	0.26	0.0027
$\beta_{Albumin}$	0.40	0.0307
β_{LDH}	0.27	0.0003
$\beta_{Platelet}$	0.58	0.0517
β_{Stage}	0.47	0.0066
$\beta_{Weight\ loss}$	0.28	0.0135
β_{WBC}	0.32	5.274×10^{-5}
$\beta_{Smoking}$	0.16	0.0033
$\beta_{Haemoglobin}$	0.31	0.0120
β_{ALP}	0.43	0.0089
$\beta_{brain\ metastasis}$	0.38	0.0165

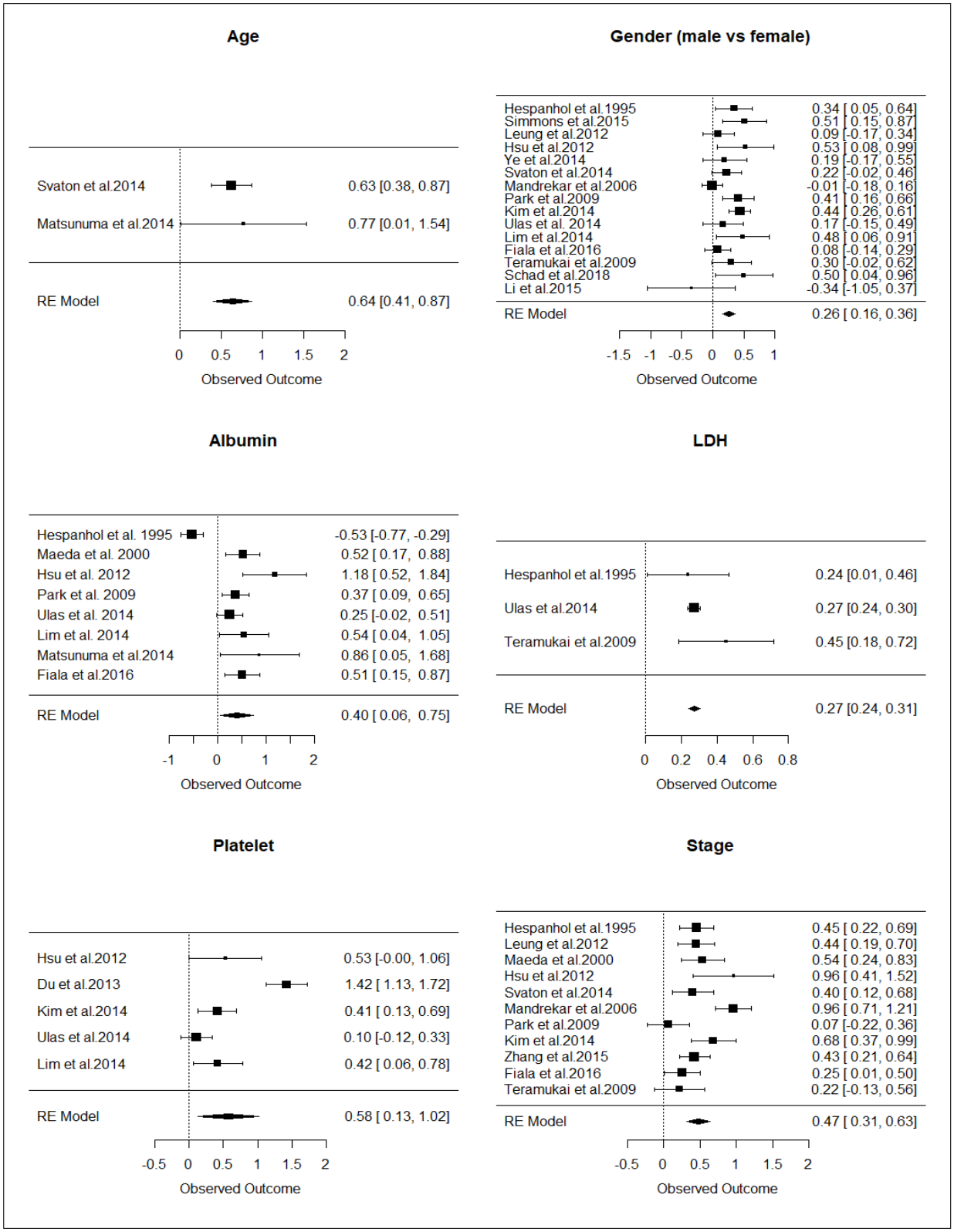

Figure 6.2: Forest plots of meta analysis results for advanced lung cancer prognostic factors

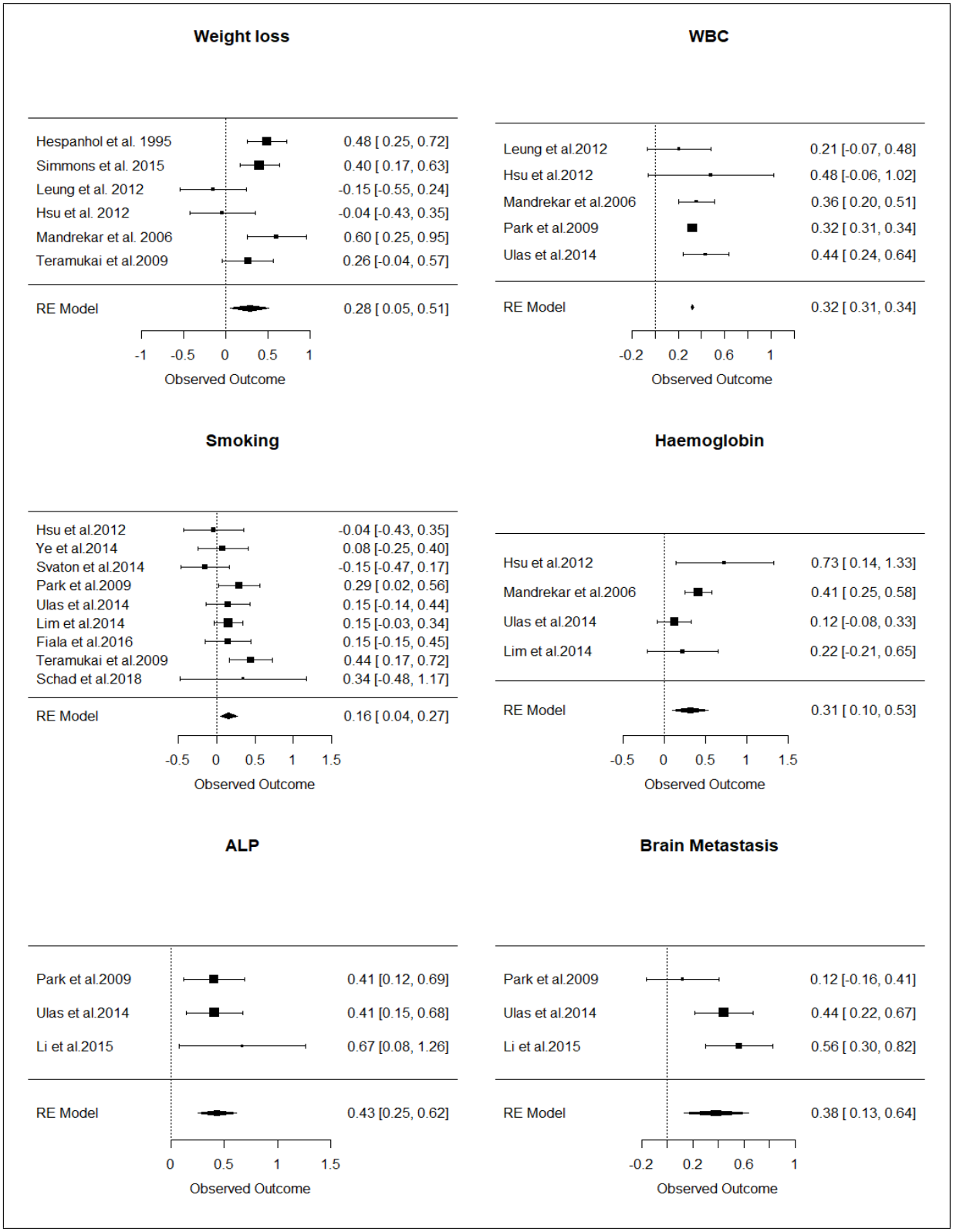

Figure 6.3: Forest plots of meta analysis results for advanced lung cancer prognostic factors

6.5 Prior construction for Weibull models

To obtain the prior distribution for each parameter in a Weibull survival model, we can use the information such as that presented in Sections 6.3 and 6.4. Three types of priors are required: prior for Weibull shape parameters α, prior for baseline log hazard (β_0) and prior for the coefficients of linear predictors in the Weibull survival model. A suitable elicitation method can be based on the method proposed by Consul (2016).

In the remainder of Section 6.5, we will construct the prior distributions for the SNLG non-Hodgkin's lymphoma data as an example. This prior specification will be based on information from published studies. In Chapter 7, we will use this prior to assess the performance of our INLA method in these circumstances.

6.5.1 Prior construction for the shape parameter for a Weibull survival model

In constructing the prior for a Weibull shape parameter, α, we will not be able utilise any information from the previous literature since all of them utilised the semi-parametric Cox Proportional Hazard approach. Nevertheless, we can construct the prior distribution for α by considering the median survival time, t_m, upper quartile, $t_{q(3)}$ and lower quartile, $t_{q(1)}$ of survival times of a large imagined future sample. The probability of death before the median survival time is given by

$$\Pr(T < t_m) = 0.5.$$

The individuals' survival function at the median time of survival is

$$S(t_m) = \exp(-\lambda t_m^{\alpha}) = 0.5,$$

since S(t_m) = 1 - F(t_m). Then, with simple algebraic manipulation, we can conveniently express t_m as

$$t_m = \Big(\frac{\log 2}{\lambda}\Big)^{1/\alpha}.$$

Similarly, we can use the same mathematical procedures to express the survival function for both lower and upper quartiles. This will lead to

$$t_{q(1)} = -\Big(\frac{\log\ 0.25}{\lambda}\Big)^{1/\alpha}$$

and

$$t_{q(3)} = -\Big(\frac{\log\ 0.75}{\lambda}\Big)^{1/\alpha}.$$

The ratio of upper quartile to lower quartile and hence the α shape parameter can be computed as follows:

$$t_{q(1)}/t_{q(3)} = \Big(\frac{\log\ 0.25}{\log\ 0.75}\Big)^{1/\alpha}$$

$$= \Big(4.818842\Big)^{1/\alpha}.$$

Then, we can take logarithms on both sides of the equation and rearrange to obtain

$$\alpha \ = \ \frac{\log\ (0.4818842)}{\log(t_{q(1)}) - \log(t_{q(3)})}. \tag{6.1}$$

To complete our prior specification for α, we have to elicit further information on $t_{q(1)}$ and $t_{q(3)}$ from the experts (presumably from a group of oncologists with long experience in managing non Hodgkin lymphoma patients) or introspectively based on information from past studies. Since α is always positive, we might give this quantity a gamma distribution ($\alpha \sim \text{Ga}(g_\alpha, h_\alpha)$). To construct a prior distribution for α, we have first to elicit a range of values for $t_{q(1)}$ and $t_{q(3)}$ which can be performed by using the hypothetical future sample (HFS) as expounded by O'Hagan et al. (2006).

We can start by firstly asking our expert to identify an average SNLG patient that is usually seen at a clinical setting. This patient should be the one that has the minimum uncertainty about his survival times, based on the expert's opinion. Then, we can enquire further by asking the expert to assess the median survival time, t_m, by asking the expert the time when there is an equal probability that the patient will die before and after it.

To elicit $t_{q(1)}$, we can ask the expert to further consider the case where an average SNLG patient who died before t_m. Then, we ask the expert to assess the time when such a patient would die prior to and after this time with an equal probability, provided the patient died before t_m. This will be our point estimate for $t_{q(1)}$. We may also elicit the point estimate for $t_{q(3)}$ by asking the expert to identify the time when an average patient is equally likely to die before and after this time, provided that the patient is still alive after t_m. This should be our assessment for $t_{q(3)}$. We can then obtain an assessment for $\varphi = t_{q(3)}/t_{q(1)}$. Nevertheless, we still have to quantify the uncertainty about φ.

To achieve this, we need to elicit φ_1, φ_2 and φ_3 which corresponds to the first, second and third quartiles of $\boldsymbol{\varphi}$, the multiplier of $t_{q(1)}$. This can be converted to $Q_1(\alpha)$, $Q_2(\alpha)$ and $Q_3(\alpha)$ using (6.1). The parameters g_α and h_α are then chosen to suit $Q_1(\alpha)$ and $Q_3(\alpha)$. Due to the fact that h_α is the scale parameter for the gamma distribution, only g_α needs to be specified and this can be achieved using numerical iteration. Then, we can find h_α using the following relationship: $h_\alpha = q_{g_\alpha}(0.75)/Q_{3(\alpha)}$ and $q_{g_\alpha}(0.75)$ represents the upper quartile of $\text{Ga}(g, 1)$ distribution. For further details, see Consul (2016).

6.5.2 Prior construction for the baseline log hazard

The baseline hazard is a hazard experienced by a baseline patient. It is convenient to choose as the baseline a patient with a typical, or roughly average, covariate pattern. For instance, consider the SNLG data. In this case, we choose the baseline patient to be an individual whose age is 60, who is female, and has normal values or categories for each predictor. Let us consider that survival time approximately follows an exponential distribution. That is $\gamma \simeq 1$. Based on the current literature and experience managing such a cohort of non-Hodgkin Lymphoma patients, we can rationally deduce the survival time of a baseline non-Hodgkin lymphoma patient (i.e low risk patients) to be more than 5 years if they receive the current standard chemotherapy regime for non-Hodgkin lymphoma patients (rituximab-based chemotherapy). Hence, we postulate that the plausible range of median survival time for baseline non-Hodgkin lymphoma patients to be between 5 years and 10 years (i.e 60 months to 120 months). We can represent this by the inequality

$$60 < t_m < 120.$$

$$\text{That is } \exp(-120\lambda) < \exp(-\lambda t_m) < \exp(-60\lambda).$$

$$\text{So, } -120\lambda < \text{-log(2)} < -60\lambda,$$

$$\frac{\log(2)}{120} < \lambda < \frac{\log(2)}{60},$$

$$\log\Big(\frac{\log(2)}{120}\Big) < \log(\lambda) < \log\Big(\frac{\log(2)}{60}\Big),$$

$$\text{and} \quad -5.154005 < \log(\lambda) < -4.460857.$$

We might then give $\log(\lambda)$ a Gaussian distribution. We can obtain the mean value and standard deviation for $\log(\lambda)$ by solving the two linear equations

$$\begin{aligned} \mathrm{E}[\log\,(\lambda)] - 2\sigma &= -5.154005, \\ \mathrm{E}[\log\,(\lambda)] + 2\sigma &= -4.460857. \end{aligned}$$

The solution for these equations will be the prior mean and standard deviation for log (λ). Hence

$$\sigma \;=\; 0.173287 \qquad \text{and} \qquad \mathrm{E}[\log(\lambda)] = -4.807431.$$

Since $\log(\lambda) = \beta_0$, this culminates in

$$\beta_0 \sim \mathcal{N}(-4.807, 0.03003)$$

or, in precision form

$$\beta_0 \sim \mathcal{N}(-4.807, 33.30^{-1}).$$

6.5.3 Prior construction for the coefficients of linear predictors

To obtain the prior means and standard deviations for coefficients of linear predictors, we shall use the method proposed by Ren and Oakley (2014) and elucidated further by Wilson and Farrow (2017). The derivations of prior means and standard deviations for continuous and 2-level categorical covariates shall be demonstrated as examples on how to elicit prior distribution parameters based on the information from the literature. We shall also assume that the prior distributions of coefficients in linear predictors are Gaussian. For continuous covariates, the age of the patient will be chosen as an example. For baseline patient i, the hazard $h_i(\mathrm{t}) = \exp\,\big(\beta_0\big)$. Then, let us consider another patient k who is 10 years older than the baseline patient, with the same covariate pattern as the baseline patient for other covariates. Hence, the hazard for patient k can be expressed as $h_k(\mathrm{t}) = \exp\,\big(\beta_0 + 10\beta_{age}\big)$. Hence, the hazard multiplier for patient k can be derived as

$$\lambda_k = \frac{h_k(t)}{h_i(t)} = \frac{\exp\Big(\beta_0 + 10\beta_{age}\Big)}{\exp\Big(\beta_0\Big)} = \exp\Big(10\beta_{age}\Big).$$

In the non-Hodgkin Lynphoma case, we identified two previous studies that gave estimates of β_{age}: Carson et al. (2012) and Hayward et al. (1991). The estimates for β_{age}, according to these two studies, are 0.02 and 0.039. However, only Carson et al. (2012) gave a 95% CI of the estimate, which is between 0.0247 and 0.0430, and therefore the results could not be combined using meta analysis. The mean of β_{age}, based upon these estimates, is hence 0.0295. We can also obtain the standard deviation of this estimate based on the upper and lower limits of the 95% CI

$$(0.0430 - 0.0247)/4 = 0.004575.$$

Therefore, the prior distribution for β_{age} is

$$\beta_{age} \sim \mathcal{N}(2.950 \times 10^{-2}, 2.093 \times 10^{-4}).$$

For a 2-level categorical covariate, the prior distribution for $\beta_{B-symptoms}$ shall be shown. Again, we compare the baseline patient i with patient j who has the same covariate pattern as the baseline patient i, except for B-symptoms. The hazard multiplier is therefore

$$\lambda_j = \frac{h_j(t)}{h_i(t)} = \frac{\exp\Big(\beta_0 + \beta_{B-symptoms}\Big)}{\exp\Big(\beta_0\Big)} = \exp\Big(\beta_{B-symptoms}\Big).$$

From the previous literature, we identified 6 studies that gave estimates of $\beta_{B-symptoms}$: Christina et al. (2013), Slymen et al. (1990), Flowers et al. (2013), Oki et al. (2008), Carson et al. (2012) (2 estimates) and Song et al. (2010). The estimates for $\beta_{B-symptoms}$, according the synthesis of results from these 7 studies is given in Table 6.1. We choose that the mean of $\beta_{B-symptoms}$, based upon this meta analysis result, is thus 0.26. Since the variance of $\beta_{B-symptoms}$ is 0.1266 (Table 6.1), the standard deviation of $\beta_{B-symptoms}$ is hence 0.3558. However, we may make the prior variance for $\beta_{B-symptoms}$ bigger to account for the heterogeneous nature of our study population and therefore we choose a prior variance double the one we obtained from meta analysis. Hence the prior variance is 0.2532 (or a prior SD of 0.5032). Therefore, the prior distribution for $\beta_{B-symptoms}$ is

$$\beta_{B-symptoms} \sim \mathcal{N}(0.26, 0.2532).$$

The prior construction for an m unordered categorical factor is as follows. Suppose that we want to construct a prior for m unordered categorical covariates. For a model where

an intercept term is included, the degrees of freedom would be $m-1$. To ensure that the parameters are identifiable, we can constrain them using a sum to zero constraint or assign a zero value to one of the parameters which is usually the baseline or reference category. If we choose to use a corner point constraint, this may indicate that our prior knowledge is greater for this baseline category. As an example, let us say that we endeavour to build a prior for a four-level categorical covariate. In this case, based on our prior knowledge we may regard the first level of that categorical covariate as our reference level. The structure of that categorical covariate with level 1 as the reference level is shown in Table 6.3.

Table 6.3: Corner constraint structure with level 1 as the reference category

Level	Contrast		
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

The consequence of using this prior structure is that a different prior variance is assigned to the reference level in contrast with other levels. In practice, however, our knowledge about the reference category may not be more substantial than our knowledge about the other levels. Fortunately, other methods can be used to deal with this issue. One of them is that we could use orthogonal contrasts to ensure that the parameters could have the sum-to-zero property. Besides, an orthogonal contrast may also ensure that the parameters are exchangeable. As a result, all the parameters might possess identical means, variances and covariances between two parameters.

To demonstrate this, suppose we have a four-level factor and we aim to construct an orthogonal contrasts with sum-to-zero constraint for the parameters. We can achieve this by using this scheme of orthogonal contrast which is presented in Table 6.4.
However, it is more cumbersome to create a contrast scheme if the number of levels is not a power of two. To address this problem, we could use the scheme presented in Table 6.5 for an m-level category.

As an example, an appropriate orthogonal constrast scheme for a 5-level factor is depicted in Table 6.6.

Table 6.4: Scheme of orthogonal contrasts with sum-to-zero constrain

Level	Contrast		
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

Table 6.5: An orthogonal contrast scheme for a category whose levels are not multiple of two

Level	Contrast				
1	-1	-1	-1	....	-1
2	1	-1	-1	....	-1
3	0	2	-1	....	-1
4	0	0	3	....	-1
⋮	⋮	⋮	....	⋮	
m	0	0	0	....	m - 1

Table 6.6: Example of a scheme where the number of levels is 5

Level	Contrast			
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	0	2	-1	-1
4	0	0	3	-1
5	0	0	0	4

To exemplify further, we might wish to build a prior structure for a categorical covariate where all levels of that covariate should be exchangeable based on the orthogonal contrast with sum-to-zero constraint as shown in Table 6.5. To achieve this, we may convert $m-1$ uncorrelated random variables into m random variables that have the zero-sum constraint. These random variables may serve as parameters for an m-level categorical covariate. For instance, let us say that we have $m-1$ independent zero - mean

random variables $\delta_1, \delta_2, \ldots, \delta_{m-1}$ with $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_{m-1})^T$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T$. Let $\text{Var}(\delta_j) = w_j$, $\boldsymbol{\beta} = M\boldsymbol{\delta}$ where M represents a $m \times (m-1)$ matrix (Table 6.5). In this case, every column is a contrast such that the total is ensured to be zero.

If we make $(m-1)^2 w_{m-1} = \upsilon$, hence, $\text{Var}(\beta_m) = \upsilon$. Therefore,

$$w_{m-1} = \frac{\upsilon}{(m-1)^2}. \tag{6.2}$$

For $1 \leq i \leq m$, we need

$$\text{Var}(\beta_i) = \upsilon = (i-1)^2 w_{i-1} + \sum_{k=i}^{m-1} w_k$$

and

$$\text{Var}(\beta_{i+1}) = \upsilon = i^2 w_i + \sum_{k=i+1}^{m-1} w_k$$

. Thus

$$(i-1)^2 w_{i-1} - i^2 w_i + \sum_{k=i}^{m-1} w_k - \sum_{k=i+1}^{m-1} w_k = 0$$

$$(i-1)^2 w_{i-1} - i^2 w_i + w_i = 0$$

and

$$w_{i-1} = \frac{(i^2-1)w_i}{(i-1)^2} = \frac{(i+1)}{(i-1)} w_i. \tag{6.3}$$

We are also required to demonstrate the equality of covariances by showing what the values are. So

$$\text{covar}(\beta_m, \beta_{m-1}) = -(m-1)w_{m-1} = -(m-1)\frac{\upsilon}{(m-1)^2} = \frac{-\upsilon}{m-1}.$$

If $i = 2, \ldots, m-1$

$$\text{covar}(\beta_i, \beta_{i-1}) = -(i-1)w_{i-1} + \sum_{k=i}^{m-1} w_k$$

and

$$\text{covar}(\beta_{i+1}, \beta_i) = -iw_i + \sum_{k=i+1}^{m-1} w_k.$$

Hence

$$\begin{aligned}\text{covar}(\beta_i, \beta_{i-1}) - \text{covar}(\beta_{i+1}, \beta_i) &= iw_i - (i-1)w_{i-1} + w_i \\ &= (i+1)w_i - (i-1)w_{i-1} \\ &= (i+1)w_i - (i-1)\frac{(i+1)}{(i-1)}w_i \\ &= 0\end{aligned}$$

and

$$\text{covar}(\beta_1, \beta_2) = \text{covar}(\beta_2, \beta_3) = \text{covar}(\beta_{m-1}, \beta_m) = -\frac{\upsilon}{m-1}.$$

Clearly, for $i = 3, \ldots, m$ and $j = 1, \ldots, i-2$

$$\text{covar}(\beta_i, \beta_j) = \text{covar}(\beta_i, \beta_{i-1}).$$

Therefore, for all $i \neq j$,

$$\text{covar}(\beta_i, \beta_j) = -\frac{\upsilon}{m-1}. \tag{6.4}$$

Besides, we may wish to weaken the exchangeability assumption and this can be achieved by permitting the prior mean for each parameter to be dissimilar and this can be achieved by making the differences between the parameters and prior means exchangeable. A set of exchangeable parameters might be constructed if $\beta_j - \text{E}(\beta_j)$ for $j = 1, \ldots, m$. Besides, the variances might also be different whilst preserving $m-1$ degrees of freedom. Suppose that $\beta_1^*, \ldots, \beta_m^*$ forms a set of exchangeable quantities that also has the sum-to-zero constraint. We may then let $\beta_j = m_j + s_j\beta_j^*$ for some selection of m_j, s_j which permits the prior mean to be non-identical. For a further exposition on the general structure of prior construction for categorical covariates, refer to Farrow (2003).

6.5.4 Prior for the frailty variance

The specification of an appropriate prior distribution for the frailty variance poses a difficult but fascinating question. This lies in the fact that it is cumbersome to visualise the actual effects of frailties on the survival time, especially in the univariate case. One such problem is that it makes the unconditional lifetime distribution for a given covariate profile not be a Weibull distribution even though the conditional distribution of the lifetimes given the frailty term is Weibull. However this fact itself offers the possibility of a way to consider prior beliefs about the frailty variance.

Suppose that we consider individual frailties Z_i to have a gamma distribution. So $Z_i \sim \text{Ga}(g, g)$. Now the marginal survival probability at time t is

$$S(t) = \text{E}_Z\left[\exp(-Z\lambda_i t^\alpha)\right] = \int_0^\infty [\Gamma(g)]^{-1} g^g z^{g-1} e^{gz} e^{-z\lambda_i t^\alpha}\, dz = \left(\frac{g}{g+\lambda_i t^\alpha}\right)^g.$$

If we write $v_z = g^{-1}$ for the frailty variance, then

$$S(t) = (1 + v_z\lambda_i t^\alpha)^{-1/v_z}.$$

Solving this for t, we obtain

$$t = \left[\frac{S(t)^{-v_z} - 1}{v_z\lambda_i}\right]^{1/\alpha}.$$

Let us write $Q_1,\ Q_2$ and Q_3 for the values of t corresponding to the three quartiles of the lifetime distribution where $S(Q_i) = S_i = 1 - i/4$ for $i = 1, 2, 3$. Then

$$\frac{Q_i}{Q_j} = \left[\frac{S_i^{-v_z} - 1}{S_j^{-v_z} - 1}\right]^{1/\alpha}.$$

So, by eliciting values for two such ratios, we can obtain two simultaneous equations which can be solved numerically to obtain values for α and g = v_z^{-1}. By eliciting distributions, or at least quantiles, for the ratios, we can obtain a joint distribution for α and g. Suitable ratios could be Q_2/Q_3 and Q_1/Q_2.

In practical terms, we might give both g and α gamma prior distributions, compute the corresponding distributions of the ratios to see whether they seem reasonable and then adjust the hyperparameters of the prior distributions accordingly. If it is preferred to use a lognormal distribution for the frailties then the method above might still be used as a reasonable approximation.

A possible alternative approach might be based on the fact that the presence of frailties causes non-proportionality of hazard effects. Consul (2016) discussed elicitation of prior distributions for covariate effects. This method could be adapted so that, as well as considering patients from time $t = 0$, we compare the conditional survival distributions of patients who are still alive at some later time. The presence of frailties will cause the hazard ratios to be different at the later time.

6.6 Prior construction for a piecewise constant hazard model

The prior distribution for this model is the Gaussian process which specifies the joint distribution of $\beta_1,\ldots,\beta_p$, where $\boldsymbol{\beta}_j = \beta_{0j},\ldots,\beta_{pj}$. Two types of priors will be discussed here: hierarchical and autoregressive priors. These priors are based on the work by Fung (2017). We shall describe the hierarchical priors first for the piecewise constant hazard model.

6.6.1 Hierarchical priors

Here, we shall consider building hierarchical priors for the baseline log hazard, β_0, in the piecewise constant hazard model for the SNLG dataset.

Suppose that $\beta_{0,j}$ is the baseline log hazard at interval j and β_0 is the mean baseline log hazard. We can build the hierarchical prior for the baseline log hazard by conditioning $\beta_{0,j}$ on β_0. Then we can specify the hyperprior distribution for β_0 with hyperparameters m and ν_2. This can be represented by

$$\beta_{0,j} \mid \beta_0 \sim \mathcal{N}(\beta_0, \nu_1),$$

$$\beta_0 \sim \mathcal{N}(m, \nu_2).$$

In this case, we can clearly observe that the marginal variance of $\beta_{0,j}$ is divided into two separate components: ν_1 is the variance of the conditional distribution, $\beta_{0,j} \mid \beta_0$, and ν_2 is the variance of β_0. The marginal variance for $\beta_{0,j}$ is hence $\nu = \nu_1 + \nu_2$. In section 6.5.2, we assessed the variance of β_0 to be 0.03002838. In this case, we therefore let the marginal variance of β_{0j} for every single time interval be 0.03002838. The inter-interval covariance is given by

$$\text{Cov } (\beta_{0,j}, \beta_{0,k}) = \nu_2, \qquad\qquad j \neq k$$

and the correlation between time intervals is given by:

$$\text{Cor } (\beta_{0,j}, \beta_{0,k}) = \frac{\nu_2}{\nu_1 + \nu_2}, \qquad\qquad j \neq k$$

Since we judge that the correlation between the log hazards in different time intervals is

high (0.95), we shall obtain

$$\nu_2 = 0.95\ (\nu_1 + \nu_2). \tag{6.5}$$

However, we know that $\nu_1 + \nu_2 = \nu$. In the NHL example, we have $\nu = 0.03002838$. We can thus substitute this value into 6.5 and obtain

$$\nu_2 = 0.028526961$$

and

$$\nu_1 = 0.03002838 - 0.028526961 = 0.001501419.$$

Hence, the hierarchical prior structure has the final form:

$$\beta_{0,j} \mid \beta_0 \sim \mathcal{N}(\beta_0, 0.001501),$$

$$\beta_0 \sim \mathcal{N}(-4.807, 0.02853).$$

Assessment of $\text{Corr}(\beta_{0,j}, \beta_{0,k})$ can be done using a hypothetical future sample method. Suppose that we have assessed prior means and variances for $\beta_{0,j}$ and $\beta_{0,k}$ where k>j. Suppose that we are now told to imagine observing that, out of n_j patients alive at time τ_{j-1}, X_j die in interval I_j. We should then reassess our median for the number X_k of n_k patients alive at the beginning of interval I_k who would die in this interval. This allows us to calculate $\text{E}(\beta_{0,k} \mid X_j)$ and, using Bayes linear kinematics, we can deduce the correlation. For further details, refer to Wilson and Farrow (2010, 2017).

6.6.2 Autoregressive priors

Another approach to specifying a prior distribution for the piecewise constant hazard model is by using an autoregressive process prior. This approach is more rationally plausible since we can now specify the log hazards in time intervals that are closer together to be more correlated than the time intervals that are further apart. We shall firstly explain the fundamental aspects of autoregressive processes before demonstrating how we can use this stochastic process for the construction of autoregressive priors.

Borrowing the notation used by Fung (2017), we shall denote $\{Y_t\}$ as a first order autoregressive (i.e. AR(1)) process. This can be written as

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + \epsilon_t$$

with constants μ and ϕ, and white noise process, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, provided that $\mid \phi \mid < 1$. The term $\phi(Y_{t-1} - \mu) + \epsilon_t$ can be viewed as an infinite moving average (MA), since an AR process can be written in MA form and vice versa. Hence, this AR(1) process can be written as

$$Y_t = \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}.$$

The expectation and variance of Y_t are

$$E[Y_t] = \mu \qquad \text{and} \qquad \text{Var}[Y_t] = \frac{\sigma^2}{1-\phi^2},$$

if $\mid \phi \mid < 1$ (based on the limit of the sum of a geometric progression). The lag-k autocovariance and autocorrelation functions are given by

$$\gamma(k) = \frac{\sigma^2 \phi^k}{1-\phi^2} \qquad \text{and} \qquad \rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k.$$

We can then use these results to construct autoregressive priors for the baseline log hazard for the SNLG and Malaysian advanced lung cancer data sets. For the SNLG data sets, the baseline hazards are denoted by $\beta_{0,1}, \beta_{0,2}, \ldots, \beta_{0,10}$. For $j = 2, 3, \ldots, 10$, the AR(1) equation for the baseline log hazard is given by

$$\beta_{0,j} = \beta_0 + \phi(\beta_{0,j-1} - \beta_0) + \epsilon_j, \qquad \epsilon_j \sim \mathcal{N}\big(0, p^{-1}\big).$$

Hence, the expectation and variance of the conditional distribution of $\beta_{0,j} \mid \beta_0, \beta_{0,j-1}$ are given by

$$\text{E}(\beta_{0,j} \mid \beta_0, \beta_{0,j-1}) = \beta_0 + \phi(\beta_{0,j-1} - \beta_0)$$

and

$$\text{Var}(\beta_{0,j} \mid \beta_0, \beta_{0,j-1}) = \frac{1}{p}.$$

Hence, the conditional distribution of $\beta_{0,j} \mid \beta_0, \beta_{0,j-1}$ is

$$\beta_{0,j} \mid \beta_0, \beta_{0,j-1} \sim \mathcal{N}\big(\beta_0 + \phi(\beta_{0,j-1} - \beta_0), p^{-1}\big). \tag{6.6}$$

We can then specify a hierarchical prior for each parameter in (6.6). Firstly, we give

Gaussian priors for β_0 and $\beta_{0,1} \mid \beta_0$

$$\beta_0 \sim \mathcal{N}(m, \nu) \qquad \text{and} \qquad \beta_{0,1} \mid \beta_0 \sim \mathcal{N}\Big(\beta_0, \frac{1}{p_1}\Big).$$

If we equate the variance obtained for $\beta_{0,1} \mid \beta_0$ with the prior variance obtained previously, $(\frac{\sigma^2}{1-\phi^2})$, since we assume the process is stationary, we obtain

$$p_1 = p(1 - \phi^2).$$

Hence, the marginal variance is

$$\gamma(0) = \nu + \frac{1}{p(1-\phi^2)}.$$

In the SNLG case, we obtain $\gamma(0) = 0.03002838$. The covariance between $\beta_{0,j}$ and $\beta_{0,j+k}$ and the lag-k autocorrelation are given by

$$\gamma(k) = v + \frac{\phi^{|k|}}{p(1-\phi^2)}$$

and

$$\begin{aligned} \rho(k) &= \frac{\gamma(\mathrm{k})}{\gamma(0)} \\ &= \frac{\nu + \left(\dfrac{\phi^{|k|}}{p(1-\phi^2)}\right)}{\nu + \left(\dfrac{1}{p(1-\phi^2)}\right)} \\ &= \frac{\nu p(1-\phi^2) + \phi^{|k|}}{\nu p(1-\phi^2) + 1}. \end{aligned} \tag{6.7}$$

If we write $\varrho = \nu p(1-\phi^2)$, the final form of (6.7) is therefore

$$\rho(k) = \frac{\varrho + \phi^{|k|}}{\varrho + 1}.$$

Then we have to specify the value for ϕ. We can make the neighbouring time-intervals more strongly correlated than those that are further apart, provided that $\phi > 0$. In the SNLG case, we specify ϕ to be 0.95 and $\nu = \frac{\gamma(0)}{2} = \frac{0.03002838}{2} = 0.01501419$. Then we can calculate p_1 and p as follows. We have

$$\gamma(0) = \nu + \frac{1}{p_1} = 0.01501419 + \frac{1}{p_1}$$

since $\gamma(0)$ is the marginal variance of 0.03002838, so we obtain

$$0.01501419 + \frac{1}{p_1} = 0.03002838.$$

Hence,

$$p_1 = 66.6036596$$

and

$$p = \frac{66.6036596}{1 - 0.95^2} = 683.1144575.$$

The lag-k autocorrelation function can be utilised to prove that the adjacent time intervals are more highly correlated than the ones that are further apart. Based on the values of p, ν and ϕ that we selected, $\varrho = 1$. The lag 1 and 9 autocorrelations are therefore

$$\rho(1) = \frac{1 + 0.95^1}{1 + 1} = 0.9750$$

and

$$\rho(9) = \frac{1 + 0.95^9}{1 + 1} = 0.8151.$$

From the result, it is thus proven that the time intervals that are next to each other are more highly correlated than the time periods that are more distant apart. This occurs since our selected value for $\phi(0.95)$ is very near to unity.

Hence, the final prior structure for the SNLG model has the form:

$$\beta_0 \sim \mathcal{N}\left(-4.807, \frac{0.03003}{2} \right),$$

$$\beta_{0,1} \mid \beta_0 \sim \mathcal{N}\left(\beta_0, \frac{0.03003}{2} \right),$$

and

$$\beta_{0,j} \mid \beta_0, \beta_{0,j-1} \sim \mathcal{N}\left(\beta_0 + \phi(\beta_{0,j-1} - \beta_0), 0.001464 \right), \quad j = 2, 3, \ldots, 10$$

where $\frac{1}{p} = \frac{1}{683.1144575} = 0.001464$. For further details, refer to Fung (2017) and Chatfield

Table 6.7: The prior mean and variances for each SNLG parameters

Parameters	**Mean**	**Variances**
β_{Age}	0.0295	0.00002093
β_{Gender}	0.52	0.2516
β_{Marrow}	0.88	0.2718
$\beta_{Albumin}$	0.97	0.056
$\beta_{B-symptoms}$	0.26	0.2532
$\beta_{Extrannodal}$	0.25	0.0202
β_{Bulky}	0.34	0.1476

(2003).

The prior distributions for all coefficients of linear predictors for SNLG data set are given in Table 6.7.

6.7 Prior construction for Malaysian advanced lung cancer data set

To construct prior distributions for the Malaysian-HUSM advanced lung cancer data set, we follow the same methodology. For the Weibull shape parameter, there is a paucity of information with respect to t_{q1} and t_{q3} that is available from prior studies. To construct the baseline log hazard for the piecewise constant hazard model for this data set, we assume that the median survival time, t_m for typical lung cancer patients is between 10 months and 36 months. Therefore

$$\beta_0 \sim (-3.301, 0.1025).$$

For coefficients of linear predictors and 2-level categorical variables, the prior mean and variances are given in Table 6.2. However, we make the prior variances double the ones that we obtained from meta-analysis to accommodate heterogeneity of patients' characteristics due to differences in population characteristics. Therefore, the parameters for the prior distributions are given in Table 6.8.

Table 6.8: The prior mean and variances for each advanced lung cancer parameters

Parameters	**Mean**	**Variances**
β_{Age}	0.64	0.0280
β_{Gender}	0.26	0.0054
$\beta_{Albumin}$	0.40	0.0614
β_{LDH}	0.27	0.0006
$\beta_{Platelet}$	0.58	0.1034
β_{Stage}	0.47	0.0132
$\beta_{Weight\ loss}$	0.28	0.0270
β_{WBC}	0.32	1.0548×10^{-4}
$\beta_{Smoking}$	0.16	0.0066
$\beta_{Haemoglobin}$	0.31	0.0240
β_{ALP}	0.43	0.0178
$\beta_{brain\ metastasis}$	0.38	0.0330

To construct a hierarchical prior for the piecewise constant hazard model, we know that $\nu = 0.1025495$ and we fix the correlation between the time intervals at 0.95. Therefore, using equation (6.5), we obtain $\nu_0 = 0.09742202$ and $\nu_1 = 0.00512748$. Hence

$$\beta_{0,j} \mid \beta_0 \sim \mathcal{N}(\beta_0, 0.005127),$$

$$\beta_0 \sim \mathcal{N}(-3.310, 0.1025).$$

For the autoregressive prior, using the same methodology as in section 6.6.2, we obtain the following prior structure for the baseline log hazard:

$$\beta_0 \sim \mathcal{N}\left(-3.310, \frac{0.1025}{2}\right),$$

$$\beta_{0,1} \mid \beta_0 \sim \mathcal{N}\left(\beta_0, \frac{0.1025}{2}\right),$$

and

$$\beta_{0,j} \mid \beta_0, \beta_{0,j-1} \sim \mathcal{N}\left(\beta_0 + \phi(\beta_{0,j-1} - \beta_0), 0.004999\right), \quad j = 2, 3, \ldots, 7$$

where $\frac{1}{p} = \frac{1}{200.0285} = 0.004999$. For further details, refer to Fung (2017) and Chatfield (2003).

6.8 Summary

In this chapter, we discuss the prior construction for both our survival models, primarily for the SNLG data set. In chapter 7, we will use these prior distributions in illustrative analyses of the data sets. We will use these analyses to assess the performance of our proposed INLA-within-MCMC method.

Chapter 7

Practical application

7.1 Introduction

In this chapter, we shall demonstrate the applications of the INLA method expounded in chapter 4 to the analysis of several data sets. In 7.2, we shall show the applications of INLA to survival models for the kidney infection (7.2.1) and the SNLG (7.2.2) data sets, both without missing covariate information. In 7.3, we show INLA-MCMC performance with respect to a Weibull survival model (7.3.1) and a piecewise constant hazard model (7.3.2) in the presence of missing covariate information using the SNLG data set. In 7.4 we will apply our INLA-MCMC algorithm to the Malaysian advanced lung cancer data set using both a Weibull model and a piecewise constant hazard model.

7.2 Analysis without missing covariate information

7.2.1 Analysis of kidney infection data

7.2.1.1 Data setup

In this section, the differences in the data setup between those used for `rjags` and `R-INLA` are demonstrated.

From the data setup, shown in Table 7.1, we can clearly see that the data setup for `R-INLA` is similar to the usual data setup for survival analysis and GLM modelling. On

Table 7.1: Partial display of the data setup for rjags and R-INLA when fitting survival model for kidney infection data set

RJAGS				**R-INLA**		
Event	Placement	Survival time	Censoring time	Time	Event	Placement
0	1	0.05	1.05	0.05	1	1
0	1	0.05	1.05	0.05	1	1
0	1	0.05	1.05	0.05	1	1
0	1	0.15	1.15	0.15	1	1
0	1	0.35	1.35	0.35	1	1
0	1	0.45	1.45	0.45	1	1
0	1	0.45	1.45	0.45	1	1
0	1	0.55	1.55	0.55	1	1
0	1	0.85	1.85	0.85	1	1
0	1	0.85	1.85	0.85	1	1
0	1	0.95	1.95	0.95	1	1
0	1	1.05	2.05	1.05	1	1
0	1	1.15	2.15	1.15	1	1
0	1	1.55	2.55	1.55	1	1
0	1	1.65	2.65	1.65	1	1
0	1	1.85	2.85	1.85	1	1
0	1	2.35	3.35	2.35	1	1
0	1	2.65	3.65	2.65	1	1
1	1	NA	0.25	0.25	0	1
1	1	NA	0.25	0.25	0	1
1	1	NA	0.65	0.65	0	1
1	1	NA	0.65	0.65	0	1
1	1	NA	0.75	0.75	0	1
1	1	NA	0.75	0.75	0	1
1	1	NA	0.75	0.75	0	1
1	1	NA	0.75	0.75	0	1
1	1	NA	0.85	0.85	0	1
1	1	NA	0.95	0.95	0	1
1	1	NA	1.05	1.05	0	1
1	1	NA	1.15	1.15	0	1

the other hand, the data setup for `rjags` requires that the event and censoring times are split into two separate vectors. Besides, the censoring indicators used for `rjags` are also different since the event is coded as 0 whilst in `R-INLA`, event is coded as 1. For censored observations, they are coded as 1 in the `rjags` setup and 0 in the `R-INLA` setup. This can be trivially achieved with a couple of lines R code to convert `rjags` data structure to the other and vice versa.

7.2.1.2 Exponential proportional hazard model

The model specification for the exponential proportional hazard model for this data has been shown in Section 2.11. We first fitted the exponential survival model using an MCMC procedure (Section 2.11) as in Martino et al. (2011), implemented using the rjags package (Plummer, 2016) in R environment (R Development Core Team, 2008) on a 3.20GHz Ergo Desktop with Intel(R) Core (TM) i7-4790S CPU and 8.00 GB of random access memory (RAM). Two parallel chains were used. The number of iterations was set at 10^5 per chain with a burn-in of 10^3. In the case of INLA, three versions of the method were used, using the Gaussian, Laplace and simplified Laplace approximations (Section 4.3.1, INLA step 2). We also provided the posterior summaries of regression coefficients obtained by numerical integration using a trapezoidal rule (Burden et al., 2015). The results of the analyses are given in Table 7.2 and Figure 7.1.

From Figure 7.1, the MCMC trace plots indicate good mixing. The densities calculated using MCMC and INLA are clearly very similar in the cases of both the treatment and intercept terms, indicating both MCMC and INLA give similar mean and standard deviations. This is supported by the results presented in Table 7.2 which show the posterior means and standard deviations obtained via MCMC and INLA are close to each other. However, the time taken to obtain results is much shorter for INLA than MCMC. We also provide the results and time taken for simple numerical quadrature and it is clear that INLA is almost as fast as numerical quadrature in obtaining the posterior means and standard deviations of the model parameters.

7.2.1.3 Weibull proportional hazard model

The model specification for the Weibull proportional hazard model used for the kidney infection data is given follows. The lifetime is

$$T_i \sim \text{Weibull}(\alpha, \lambda_i)$$

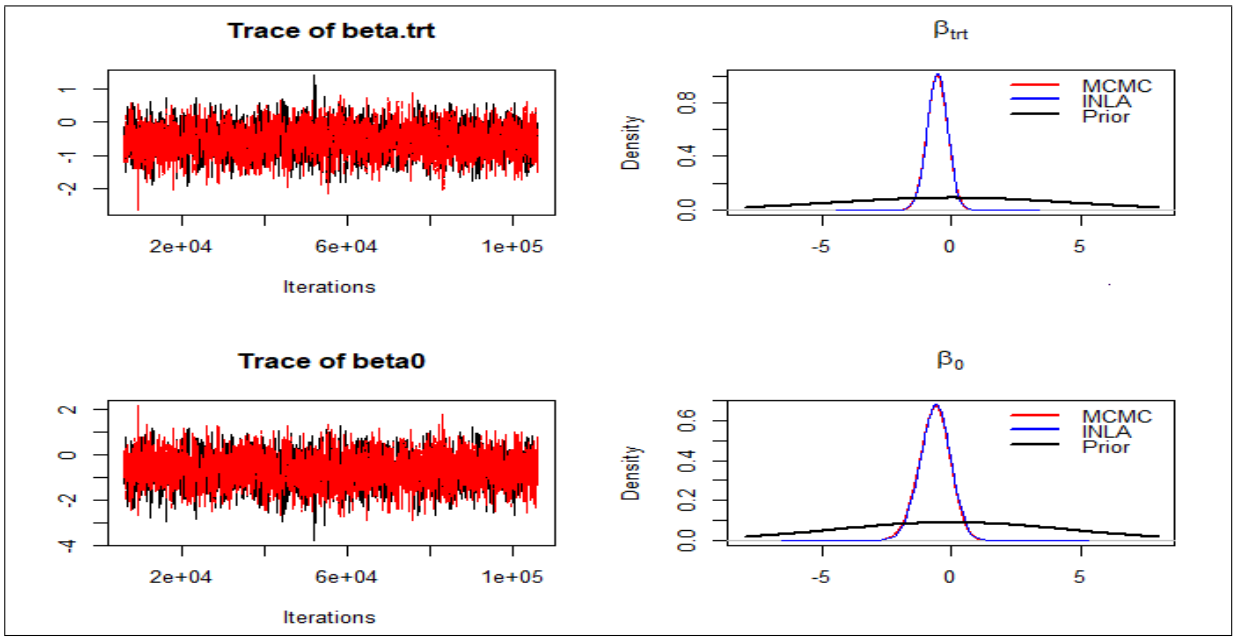

Figure 7.1: MCMC trace plots and comparisons of posterior distributions of regression coefficients obtained via INLA and MCMC

Table 7.2: The comparison of results for β_0 and β_{trt} obtained via different methods

Method	$\boldsymbol{\beta_0}$ mean (SD)	$\boldsymbol{\beta_{trt}}$ mean (SD)	**Time(seconds)**
Prior	0 (4.4721)	0 (4.4721)	-
MCMC (RJAGS)	-0.6604 (0.5992)	-0.5451 (0.3971)	2180.50
INLA(Gaussian)	-0.6357 (0.5902)	-0.5364 (0.3922)	0.64
INLA(Laplace)	-0.6552 (0.5876)	-0.5485 (0.3891)	0.83
INLA(SL)	-0.6193 (0.5902)	-0.5366 (0.3922)	0.71
Numerical quadrature	-0.6647 (0.5936)	-0.5415 (0.3936)	0.50

with density

$$f_i(t \mid \alpha, \lambda_i) = \alpha\lambda_i t^{\alpha-1} \exp(-\lambda_i t^{\alpha})$$

and survival function

$$S_i(t \mid \alpha, \lambda_i) = \exp(-\lambda_i t^{\alpha}),$$

where

$$\boldsymbol{\beta} = (\beta_0, \beta_1)^T \quad \text{and} \quad \lambda_i = \exp(\eta_i) = \exp(\beta_0 + x_i\beta_1).$$

Hence,

$$L(\alpha, \boldsymbol{\beta} \mid D) = \alpha^{\sum \delta_i} \exp\Big\{ \sum_{i=1}^{n} \delta_i \boldsymbol{z}_i^T \boldsymbol{\beta} + \sum_{i=1}^{n} \Big[\delta_i(\alpha-1)\log(t_i) - \exp(\boldsymbol{z}_i^T\boldsymbol{\beta}) t_i^{\alpha}\Big]\Big\}.$$

We give $\boldsymbol{\beta}$ a bivariate normal prior

$$\boldsymbol{\beta} \sim N_2(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$

and

$$\alpha \sim \mathrm{Ga}(\alpha_0, \kappa_0),$$

independently of β_0 and β_1. The joint posterior distribution for α and $\boldsymbol{\beta}$ is thus given by

$$\pi(\alpha, \boldsymbol{\beta} \mid D) \propto \mathrm{L}(\alpha, \boldsymbol{\beta} \mid D)\ \pi(\alpha \mid \alpha_0, \kappa_0)\ \pi(\beta_0 \mid m_0, v_0)\pi(\beta_1 \mid m_1, v_1).$$

We assigned values

$$\boldsymbol{\mu}_0 = (0, 0)^T,$$
$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix},$$
$$\alpha_0 = 1.1$$

and

$$\kappa_0 = 1.1.$$

For the building of the Weibull regression model, the covariates are introduced via λ through $\lambda_i = \exp(\boldsymbol{z}_i^T\boldsymbol{\beta})$. Hence, the corresponding hazard function can be written as

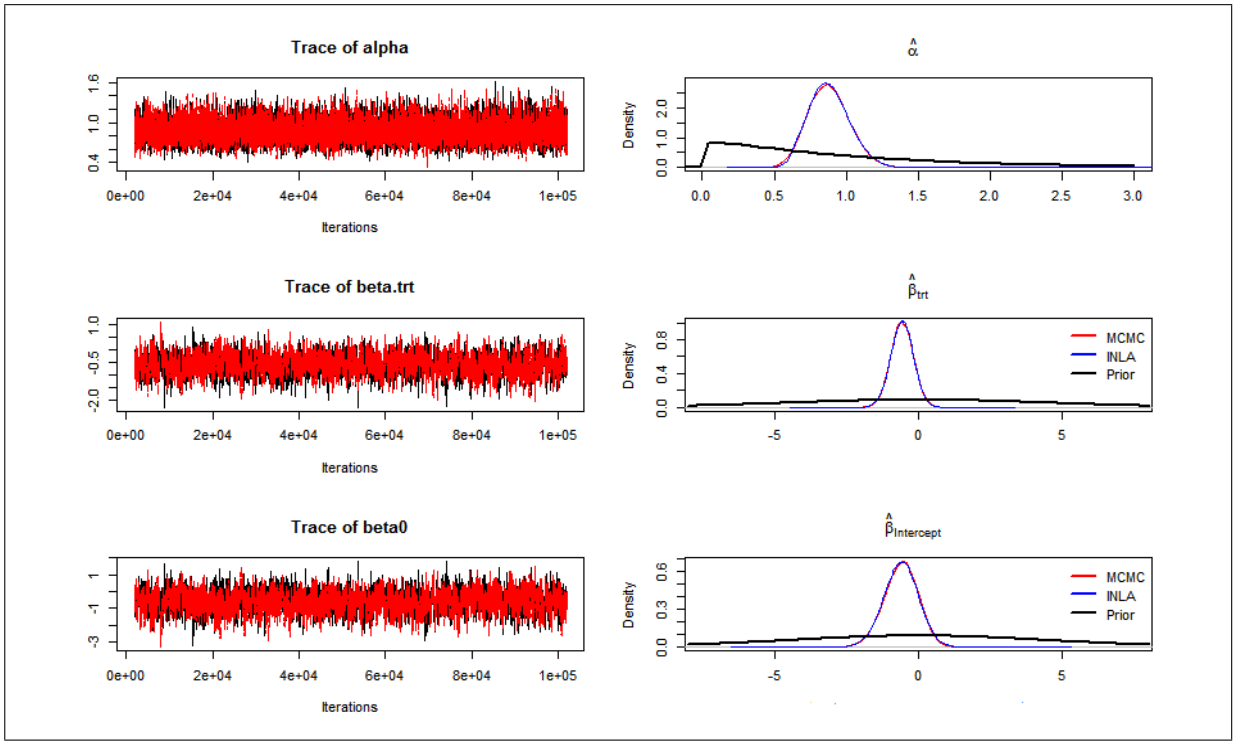

Figure 7.2: Comparisons of posterior distributions of regression coefficients obtained via INLA and MCMC

$h_i(t \mid \alpha, \lambda_i) = \alpha t^{\alpha-1}\lambda_i$. Therefore, the joint posterior density is

$$\pi(\alpha, \boldsymbol{\beta} \mid D) \propto \alpha^{\alpha_0+\sum\delta_i-1} \exp\left\{ \sum_{i=1}^{n} [\delta_i \boldsymbol{z}_i^T \boldsymbol{\beta} + \delta_i(\alpha-1)\log(t_i)) - t_i^{\alpha}\exp(\boldsymbol{z}_i^T\boldsymbol{\beta})] - \kappa_0\alpha \right.$$

$$\left. -\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0) \right\}.$$

In this case we compare MCMC (Section 2.11) and the three INLA methods (Section 4.3.1, INLA Step 2).

The number of iterations of MCMC (in RJAGS code) was set at 10^5 per chain with a burn-in set at 10^3. Two parallel chains were used. The trace plots and posterior means and standard deviations are presented as Figure 7.2 and Table 7.3.

From Figure 7.2, the trace plots indicate good mixing. The densities calculated using MCMC and INLA are clearly very similar, indicating both MCMC and INLA give similar mean and standard deviations. This is supported by the results presented in table 7.3 which show the posterior means and standard deviations obtained via MCMC and INLA

Table 7.3: The comparison of posterior summaries of β_0, β_{trt} and α for obtained via different methods

Method	$\boldsymbol{\beta}_0$ mean (SDs)	$\boldsymbol{\beta}_{trt}$ mean (SDs)	$\boldsymbol{\alpha}$ Mean (SDs)	**Time(sec)**
Prior	0 (4.4721)	0 (4.4721)	1 (0.9535)	-
MCMC (RJAGS)	-0.5698 (0.3933)	-0.5988 (0.5939)	0.8787 (0.1453)	4238.45
INLA(Gaussian)	-0.5872 (0.5931)	-0.5526 (0.3926)	0.8797 (0.1415)	1.96
INLA(Laplace)	-0.6067 (0.5905)	-0.5647 (0.3895)	0.8797 (0.1415)	3.65
INLA(SL)	-0.5708 (0.5931)	-0.5528 (0.3926)	0.8797 (0.1415)	2.72

are close to each other. However, the time taken to obtain results is much shorter for INLA than MCMC.

7.2.2 Analysis of Non-Hodgkin Lymphoma (SNLG) data set

7.2.2.1 Basic notations

In this section, the basic notations used for covariates in the SNLG data will be expounded and this is given in Table 7.4.

After labelling each covariate with the appropriate notation, we can then construct a design matrix, V, with S columns. The total number of columns, S, is determined by the total number of covariates and the total number of levels within the categorical covariates. There are now two choices of constraints to use for the categorical variables: 1) corner point constraint, 2) zero-sum constraint. However, the corner point constraint has one shortcoming. We assign a different prior variance to the baseline category (level 1) compared to the rest of the levels, as if we have better prior information about level 1 than the rest of the levels of that factor. Consequently, the alternative method, the zero-sum constraint, is more rationally appealing since it allows all the parameters to have similar means, variances and covariances.

For this model, we calculate the number of regression parameters (S) is 24, allowing m-1 for a factor with m levels. The overall number of parameters is thus 25 after taking into consideration an additional parameter, α, which is the shape parameter for Weibull distribution.

Table 7.4: The notations used for each covariate that may influence the survival time of SNLG cohort

Covariates	Notation	Types of covariates
Age	x_1	Continuous (quantitative)
Haemoglobin	x_2	Continuous (quantitative)
White blood cell count (WBC)	x_3	Continuous (quantitative)
Sex (Gender)	x_4	2-level categorical covariate
Albumin	x_5	2-level categorical covariate
AP	x_6	2-level categorical covariate
Urea	x_7	2-level categorical covariate
Extranodal disease	x_8	2-level categorical covariate
Bone marrow involvement	x_9	2-level categorical covariate
B symptoms	x_{10}	2-level categorical covariate
Bulky disease	x_{11}	2-level categorical covariate
Stage	x_{12}	4-level ordinal covariate
ECOG	x_{13}	5-level ordinal covariate
LDH	x_{14}	3-level ordinal covariate

7.2.2.2 Modelling strategies

For the SNLG dataset, we assume that the survival times follow a Weibull distribution with a shape parameter α and a scale parameter λ. We also centre the quantitative covariates (haemoglobin, wbc and age) and used orthogonal contrast for categorical covariates.

7.2.2.3 Covariate centering

There are three continuous covariates in this dataset: age, haemoglobin level and white blood cell count. To obtain proper prior specification, these quantitative variables can be standardised. For example, we can centre age as follows, $x_1 = x_{\mathrm{A}}$ - 60 where x_{A} is the subject's age.

Table 7.5: An outline of zero-sum constraint for a 4-level factor

Level	Contrast		
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

7.2.2.4 Orthogonal contrasts with zero-sum constraint

As stated previously, we can use orthogonal contrasts with a zero-sum constraint to avoid the shortcomings of a corner-point constraint scheme. When we use this strategy, we are now able to make the parameters exchangeable, thus making them have the same means, variances and covariances. As an example, the orthogonal contrast with zero-sum constrain for a 4-level factor is given by Table 7.5. This zero-sum-constraint was employed for Ann-Arbor stage, a 4-level categorical covariate in the SNLG dataset.

In this case, the effects of the four levels will sum to zero. Nevertheless, this is a trivial example since this scheme is only effective for a factor where the number of levels is a power of 2. The scheme will break down if this is violated. To generalise such a scheme to factors whose total number of levels is not a power of 2, we present the form of the generalised orthogonal contrast with zero-sum constraint in Table 7.6.

Table 7.6: The general outline for a factor whose levels are not a power of 2

Level	Contrast				
1	-1	-1	-1	....	-1
2	1	-1	-1	....	-1
3	0	2	-1	....	-1
4	0	0	3	....	-1
⋮	⋮	⋮	....	⋮	
m	0	0	0	....	m - 1

As another example, we present the outline of the zero-sum constraint for a factor

Table 7.8: The orthogonal contrast with zero-sum constraint for a 3-level factor (LDH)

Level	Contrast	
1	-1	-1
2	1	-1
3	0	2

with 5 levels in Table 7.7. This zero-sum constraint was then used for ECOG performance status which is a 5-level factor.

Table 7.7: The orthogonal contrast with zero-sum constraint for a 5-level factor (ECOG)

Level	Contrast			
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	0	2	-1	-1
4	0	0	3	-1
5	0	0	0	4

Finally, we provide the zero-sum-constraint scheme for LDH, a 3-level factor in SNLG dataset. This is given in Table 7.8.

For further information on calculating the variances of factors that have been orthogonally constrained, refer to section 6.5.3 and equations (6.2)-(6.4).

7.2.3 Analysis of SNLG Data with full covariate information

Recall from subsection 2.5.1.2, the log likelihood for survival data that follows a Weibull distribution is given by

$$\ell = \log(\boldsymbol{\beta}, \alpha \mid R) = \sum_{i \in E} [\log\{\lambda_i\} + (\alpha - 1)\log\{t_i\}] + n_D \log\{\alpha\} - \sum_{i=1}^{n} \lambda_i t_i^{\alpha}$$

where E is the set of subjects who died and n_D is the number of subjects who died. The regression coefficients are represented by

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_K)^T$$

where the intercept term is represented by β_0 and β_k represents the coefficient of covariate k. From Section 2.5.1.1, we see that the linear predictors are related to λ_i by

$$\lambda_i = \exp(\eta_i), \qquad \eta_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k}.$$

We can then assign multivariate normal prior distribution to $\boldsymbol{\beta}$

$$\boldsymbol{\beta} \sim N_{K+1}\ (\boldsymbol{\mu}, W).$$

In this case, $\boldsymbol{\mu}$ represents the vector of prior means for the regression coefficients and W is the (K+1 $\times$ K+1) variance-covariance matrix. The multivariate normal prior distribution for $\boldsymbol{\beta}$ has pdf

$$f_0(\boldsymbol{\beta}) = (2\pi)^{-K/2} \mid W \mid^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}[(\boldsymbol{\beta} - \boldsymbol{\mu})^T W^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu})]\Big\}.$$

If we take logarithms of the multivariate normal prior distribution and assume that $\beta_0, \ldots, \beta_k$ are independent and the shape parameter α is also independent of $\boldsymbol{\beta}$ as well, the logarithm of the density becomes

$$f_1(\boldsymbol{\beta}) = -\frac{K}{2}\log\{2\pi\} - \frac{1}{2}\sum_{k=0}^{K} \log \mid W_k \mid -\frac{1}{2}\sum_{k=0}^{K} \frac{\{\beta_k - \mu_k\}^2}{W_k}$$

where W^{-1} is a diagonal matrix with diagonal elements, $W_0, W_1, W_2, \ldots, W_K$. In this case, W_0 represents the variance of the intercept term and W_k is the variance of the k^{th} regression coefficient. Since the shape parameter α must be positive, we assign a gamma prior for α with shape parameter a and rate parameter b

$$\alpha \sim \text{Ga}(a, b).$$

Hence, the density of the prior distribution for α has the form

$$\pi(\alpha \mid a, b) = \frac{b^a}{\Gamma\ (a)} \alpha^{a-1} \exp\{-b\alpha\}.$$

Then, the joint prior density of $\boldsymbol{\beta}$ and α is given by $\pi(\boldsymbol{\beta}, \alpha)$. Therefore, using Bayes' theorem, the joint posterior distribution of β and α is obtained by

$$\pi(\boldsymbol{\beta}, \alpha \mid D) \propto \text{Prior} \times \text{Likelihood},$$

$$= \text{Q}\ \pi(\boldsymbol{\beta}, \alpha)\ L(\boldsymbol{\beta}, \alpha \mid D).$$

Using simple algebra and taking logs on both side, the log posterior density of $\boldsymbol{\beta}$ and α has the following form

$$\begin{aligned}
\log\ \Big(\pi[\boldsymbol{\beta}, \alpha \mid D]\Big) &= \log\Big[\text{Q}\Big] + \log\Big[\pi(\boldsymbol{\beta}, \alpha)\Big] + \log\Big[\text{L}(\boldsymbol{\beta}, \alpha \mid D)\Big] \\
&= \log\big(Q\big) + \ \big(\alpha - 1\big)\log\big(\alpha\big) - b\alpha - \frac{K}{2}\log\big(2\pi\big) - \\
&\quad \frac{1}{2}\sum_{k=0}^{K}\log\big(W_k\big) - \frac{1}{2}\sum_{k=0}^{K}\frac{\big(\beta_k - \mu_k\big)^2}{W_k} + \sum_{i\in \text{E}}\big[\log(\lambda_i) + \\
&\quad \big(\alpha - 1\big)\log(t_i)\big] + n_D\log(\alpha) - \sum_{i=1}^{n}\lambda_i t_i^{\alpha}.
\end{aligned}$$

Since the joint posterior density is analytically intractable, we have to use numerical methods. We can use an MCMC algorithm to obtain the sampled values from the posterior distribution. We apply a zero-sum constraint for categorical covariates in this SNLG Weibull survival model. Hence, the linear predictor has the following form

$$\eta_i = \beta_0 + \beta_{i,1}x_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \ldots + \beta_{11}x_{i,11} + \sum_{k=1}^{3}\beta_{12,k}\delta_{i,12,k} +$$

$$\sum_{k=1}^{4}\beta_{13,k}\delta_{i,13,k} + \sum_{k=1}^{5}\beta_{14,k}\delta_{i,14,k}.$$

with δ represents the variable for an orthogonal contrast. A Metropolis within Gibbs sampling algorithm was used which was implemented using RJAGS. Two parallel MCMC chains were employed with different starting values to ensure that the convergence had

been satisfactorily achieved. After a burn-in period of 2000 iterations, 150,000 iterations for each chain were obtained as posterior samples. The trace and autocorrelation plots were examined for convergence and adequate mixing. INLA was also used to obtain the posterior means and standard deviations of the parameters, using the simplified Laplace option. The posterior summaries are given in Table 7.9.

As we can see from the results in Table 7.9, the posterior means and standard deviations (and hence precisions) obtained using MCMC and INLA are closely similar to each other. In this case, INLA is by far more efficient than MCMC (computation time: 3.58 seconds vs 17 minutes 26 seconds, Intel i7® single core 8 GB RAM). Apart from that, we can also clearly observe that the posterior standard deviations of all parameters acquired through MCMC or INLA are much smaller compared to prior standard deviations. This signifies increases in precision due to the combination of our prior knowledge with information contained in the data through likelihood. As a result, this causes decreases in our uncertainty over all parameters.

The plots of posterior distributions obtained for each parameter (24 regression parameters and the shape parameter, α, for the Weibull distribution) with both MCMC and INLA are given in Figures 7.3 - 7.5. As we can see, the posterior distributions obtained via INLA (blue curves) nearly perfectly match the posterior distributions obtained using MCMC (red curves).

7.3 Analysis of SNLG data with missing covariate information

7.3.1 Weibull lifetime distribution

We also fitted the Weibull survival model on the whole data set which includes those cases with missing covariate information. The results are presented in Table 7.10. We can clearly see that the posterior means and standard deviations obtained using MCMC are very similar to those obtained via the INLA-MCMC algorithm. Figure 7.6 depicts the prior (black) and posterior densities of selected regression coefficients obtained via MCMC (red) and INLA-MCMC (blue) and it can be clearly seen that both match each other. The times taken for MCMC and INLA-MCMC to obtain the posterior means and standard deviations are 4200 seconds and 209 seconds, respectively.

Table 7.9: The posterior means and standard deviations (SDs) for the unknown parameters ($n = 636$)

Parameter	Posterior mean (MCMC)	Posterior SD (MCMC)	Posterior mean (INLA)	Posterior SD (INLA)
β_0(intercept)	-0.4843	0.1384	-0.4842	0.1401
β_1(Age)	0.0275	0.0046	0.0275	0.0046
β_2(HB)	-0.0069	0.0032	-0.0069	0.0032
β_3(WBC)	0.0287	0.0182	0.0288	0.0182
β_4(Sex)	0.0576	0.0569	0.0579	0.0570
β_5(Albumin)	-0.0820	0.0699	-0.0820	0.0697
β_6(Ap)	0.0568	0.0686	0.0561	0.0683
β_7(Urea)	-0.0154	0.0642	-0.0157	0.0639
β_8(Extranod)	0.0289	0.0641	0.0289	0.0641
β_9(Bulk)	0.1702	0.0570	0.1701	0.0570
β_{10}(Marrow)	0.2308	0.0865	0.2308	0.0868
β_{11}(Bsy)	-0.0439	0.0616	-0.0438	0.0616
$\delta_{12,1}$(Stage)	0.0377	0.0481	0.0379	0.0481
$\delta_{12,2}$(Stage)	0.0964	0.0428	0.0966	0.0426
$\delta_{12,3}$(Stage)	0.0636	0.0405	0.0636	0.0406
$\delta_{13,1}$(ECOG)	0.0197	0.0379	0.0198	0.0378
$\delta_{13,2}$(ECOG)	0.0524	0.0390	0.0525	0.0389
$\delta_{13,3}$(ECOG)	0.0894	0.0458	0.0893	0.0457
$\delta_{13,4}$(ECOG)	-0.1236	0.0921	-0.1234	0.0910
$\delta_{14,1}$(LDH)	0.0567	0.0377	0.0569	0.0377
$\delta_{14,2}$(LDH)	0.1366	0.0408	0.1365	0.0407
α (Shape)	0.8454	0.0389	0.8452	0.0382

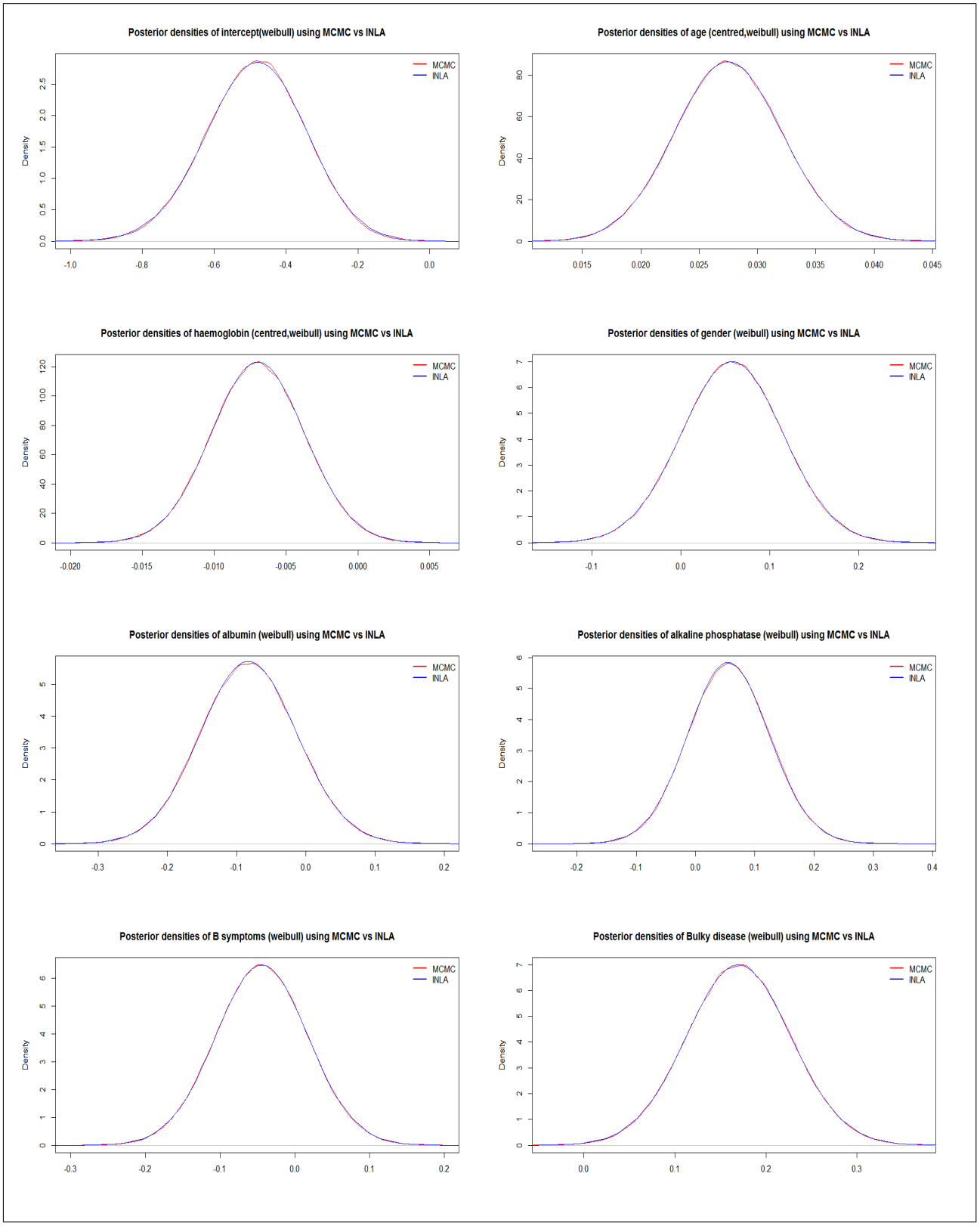

Figure 7.3: Comparisons of posterior means (SDs) of β_0 until β_{bulk} using MCMC and INLA

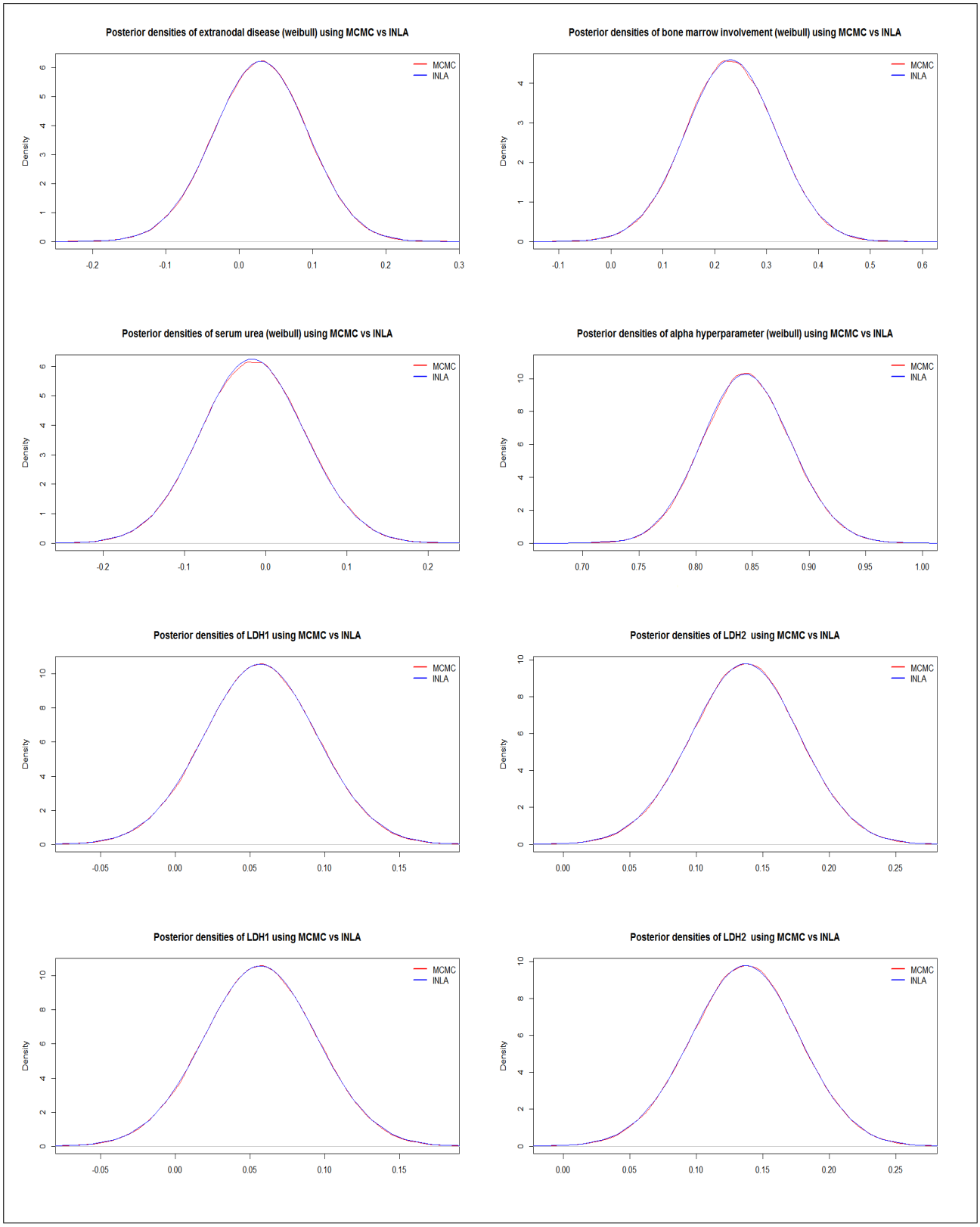

Figure 7.4: Comparisons of posterior means (SDs) β_{bsy} until $\delta_{2,\text{LDH}}$ using MCMC and INLA

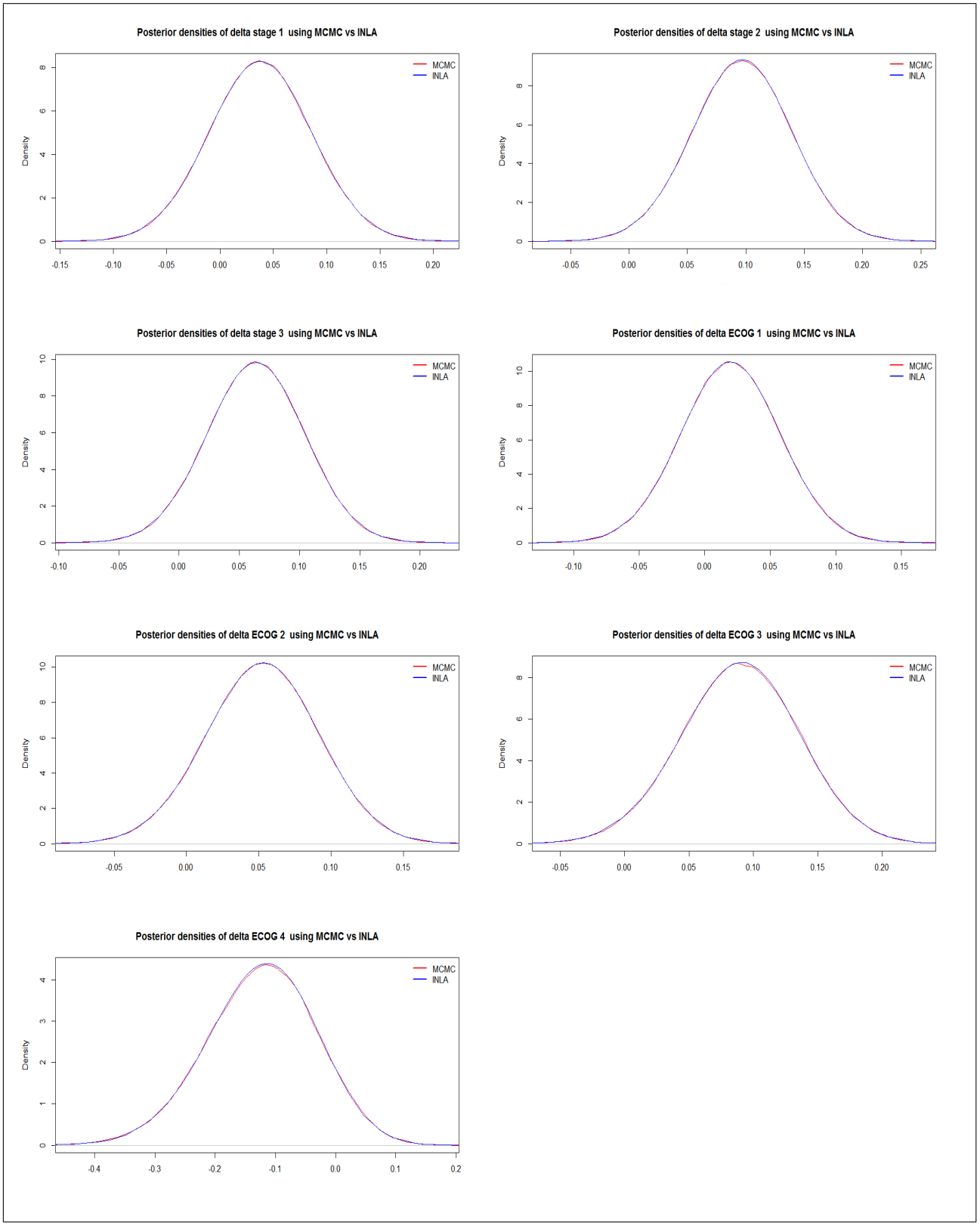

Figure 7.5: Comparisons of posterior means (SDs) of regression coefficients $\delta_{1,\text{stage}}$ until $\delta_{4,\text{ECOG}}$ using MCMC and INLA

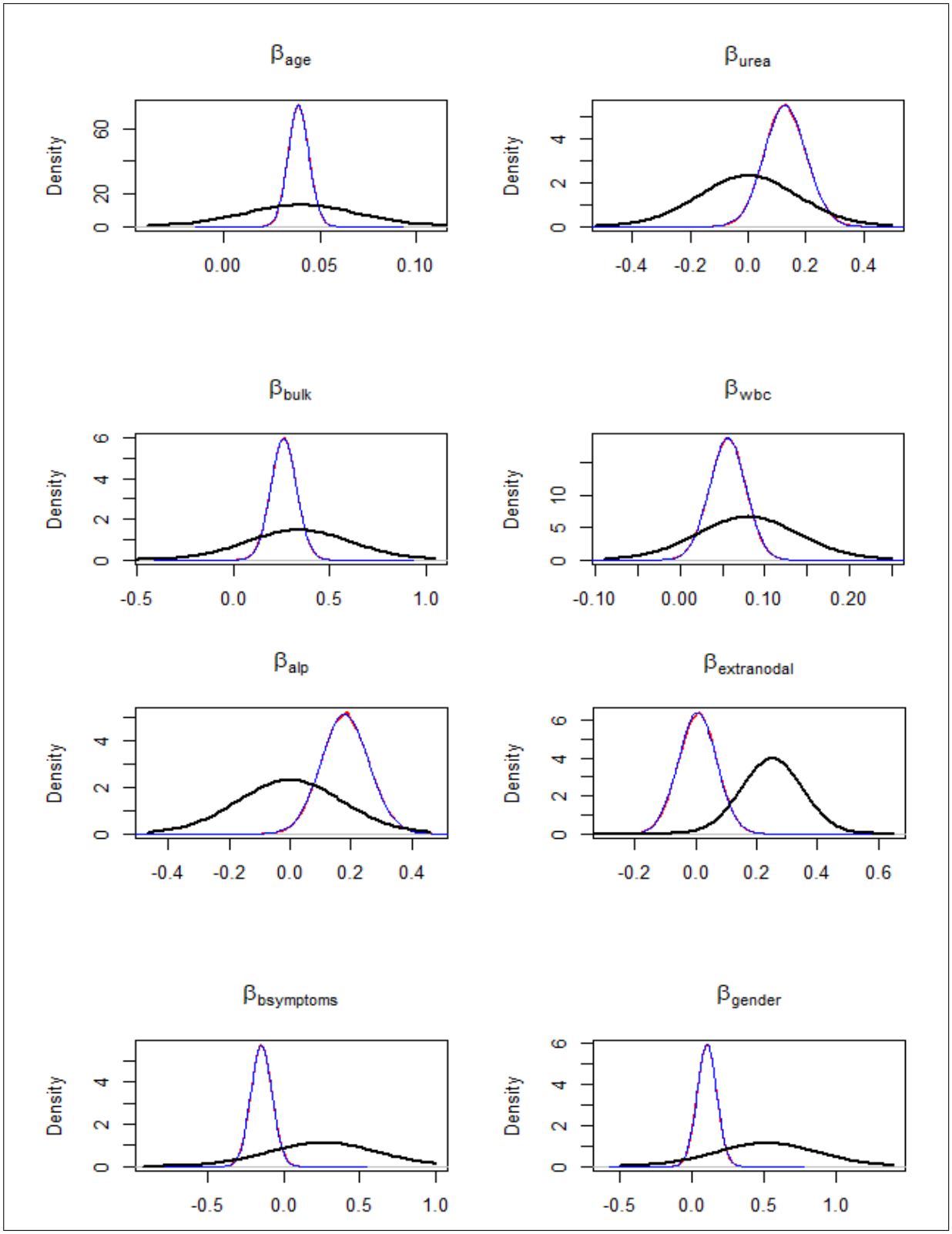

Figure 7.6: Comparisons of posterior distributions of regression coefficients obtained via INLA-MCMC (red) and MCMC (blue)

Table 7.10: The posterior means and standard deviations (SDs) for the unknown parameters ($n = 1391$)

Parameter	Post.mean (MCMC)	Post.SD (MCMC)	Post.mean (INLA-MCMC)	Post. SD (INLA-MCMC)
β_0(intercept)	-2.6998	0.1348	-2.7032	0.1350
β_1(Age)	0.0390	0.0054	0.0390	0.0054
β_2(HB)	-0.0140	0.0036	-0.0140	0.0036
β_3(WBC)	0.0557	0.0212	0.0559	0.0212
β_4(Sex)	0.1076	0.0671	0.1086	0.0671
β_5(Albumin)	-0.3724	0.0858	-0.3738	0.0853
β_6(Ap)	0.1825	0.0779	0.1817	0.0779
β_7(Urea)	-0.1295	0.0729	-0.1298	0.0727
β_8(Extranod)	0.0059	0.06251	0.0048	0.0624
β_9(Bulk)	0.2614	0.0667	0.2621	0.0669
β_{10}(Marrow)	-0.1859	0.1043	0.1881	0.1042
β_{11}(Bsy)	-0.1464	0.0697	-0.1469	0.0694
$\delta_{12,1}$(Stage)	0.0718	0.0495	0.0721	0.0496
$\delta_{12,2}$(Stage)	0.0948	0.0467	0.0949	0.0465
$\delta_{12,3}$(Stage)	0.2077	0.0414	0.2085	0.0411
$\delta_{13,1}$(ECOG)	0.0210	0.0386	0.0210	0.0386
$\delta_{13,2}$(ECOG)	0.0216	0.0418	0.0218	0.0417
$\delta_{13,3}$(ECOG)	-0.0194	0.0530	-0.0200	0.0532
$\delta_{13,4}$(ECOG)	-1.1050	0.1194	-1.1040	0.1210
$\delta_{14,1}$(LDH)	0.0634	0.0388	0.0633	0.0387
$\delta_{14,2}$(LDH)	0.1076	0.0446	0.1075	0.0447
α (Shape)	0.8397	0.0406	0.8392	0.0397

7.3.2 Analysis of SNLG Data with missing covariate information using piecewise constant hazard model

In this section, we fit a piecewise constant hazard model to the SNLG dataset. For this task, all subjects including those with missing information were included ($n = 1391$). The likelihood, L, is given by

$$\mathrm{L} = \prod_{i=1}^{n} \prod_{\mathrm{j}=1}^{\mathrm{r}} \left\{ \lambda_{ij} \right\}^{\delta_{ij}} \exp\left\{ -\lambda_{ij} \left\{ t_{ij} - \tau_{j-1} \right\} \right\}$$

where λ_{ij} represents the hazard experienced by patient i at time interval j, δ_{ij} is an indicator whether patient i died in interval j. In each interval, there are three things that may occur to patient i: surviving through that single interval, event (death) occurs in that single interval or censoring occurs. For further information, refer to 2.7.1.

Besides, a frailty term was also included in the model to weaken the dependent effects of covariates. In this case, we assumed the frailty term was log-normally-distributed, with the prior for its variance having a gamma distribution. Full distributional specification of the frailty term (denoted as Z) is given by

$$\log(Z) \sim N\ (0, \sigma_z^2) \qquad \text{and} \qquad \sigma_z^2 \sim \mathrm{Ga}(1.1, 0.53).$$

The next step is to choose a sensible value for the autoregressive parameter that we consider in 6.6.2. To do this, we have to judge how much variance of the parameters in the next time interval is explained by the values in the preceding time interval. In this case, it is assumed that 90% of the parameters' variances in the subsequent time period are explained by the value in the preceding period. Hence, our coefficient of determination, R^2 , 0.90. Consequently, our correlation, r, is 0.95. With respect to prior means and variances, we used the same values that we used in Section 6.3.2 (Table 6.1).

From the results shown in Table 7.11, the age effect decreases temporarily before it increases as the time progresses. Besides, we can clearly see, the mean and standard deviations obtained via the INLA-MCMC algorithm are really close to the mean and standard deviations obtained using MCMC only. This is more evident if we look at Figures 7.7 - 7.8 which show clear matching densities of each parameter obtained using MCMC and INLA-MCMC algorithms. The time taken for MCMC and INLA-MCMC

Table 7.11: Posterior means and standard deviations (in brackets) in each interval of time for some chosen covariates ($n = 1391$)

j	τ_j	β_0 (MCMC)	β_{age} (MCMC)	β_0 (INLA-MCMC)	β_{age}(INLA-MCMC)
1	0.316	-0.8515(0.2177)	0.06324(0.0137)	-0.8313 (0.2169)	0.0624 (0.0137)
2	0.669	-0.1731(0.1603)	0.0459 (0.0072)	-0.1609(0.1360)	0.0458(0.0720)
3	1.070	0.0847(0.1537)	0.0292(0.0072)	0.0932(0.1517)	0.0291(0.0072)
4	1.532	-0.0045(0.1761)	0.0321(0.0720)	0.0075(0.1734)	0.0321(0.0072)
5	2.079	-0.7924(0.2024)	0.0220(0.0078)	-0.7761(0.2007)	0.0220(0.0079)
6	2.749	-1.8137(0.2096)	0.1723(0.0083)	-1.8076(0.2102)	0.1723(0.0083)
7	3.612	-1.9534(0.2501)	0.1477(0.0088)	-1.9443(0.2503)	0.1477(0.0088)
8	4.828	-1.8861(0.2601)	0.1316(0.0089)	-1.8785(0.2636)	0.1316(0.0532)
9	6.908	-1.8405(0.2820)	0.1177(0.0087)	-1.8327(0.2842)	0.1177(0.0312)
10	∞	-1.4652(0.2969)	0.1065(0.0091)	-1.4581(0.2968)	0.1064(0.0091)

to compute posterior means and SDs are 4650.33 and 191.11 seconds, respectively. This shows that INLA-MCMC is the more efficient algorithm.

7.4 Malaysian advanced lung cancer data set

7.4.1 Weibull lifetime model with missing covariate information

Now we turn our attention to the Malaysian advanced lung cancer data and fit a Weibull model. From Table 7.12 and Figure 7.9, we can clearly see that posterior means and standard deviations obtained via MCMC and INLA-MCMC are very close to each other. The density plots clearly match each other and the posterior summaries of all covariates for the MCMC and the INLA-MCMC are very close to each other. The computational time for the MCMC algorithm is 5142 seconds whilst for the INLA-MCMC algorithm, the computational time is 314.57 seconds. Therefore, the INLA-MCMC algorithm is more efficient than MCMC algorithm for computing the posterior summaries of this data set.

7.4.2 Analysis of the Malaysian advanced lung cancer data set with missing covariate information using piecewise constant hazard model

We also fitted a piecewise constant hazard model for the Malaysian advanced lung cancer data set with the presence of missing covariate information. From the results presented in Table 7.13, we can clearly see that the posterior summaries for β_0 and $\beta_{neutrophil}$ for each time interval are very close to each other. Besides, Figure 7.10 demonstrates that the density plots for each parameter obtained using MCMC and INLA-MCMC algorithms again clearly match each other. The computing time is 412 seconds for the INLA-MCMC algorithm and 3972.19 seconds for the MCMC algorithm. Again, this indicates that the INLA-MCMC is a much more efficient algorithm than MCMC.

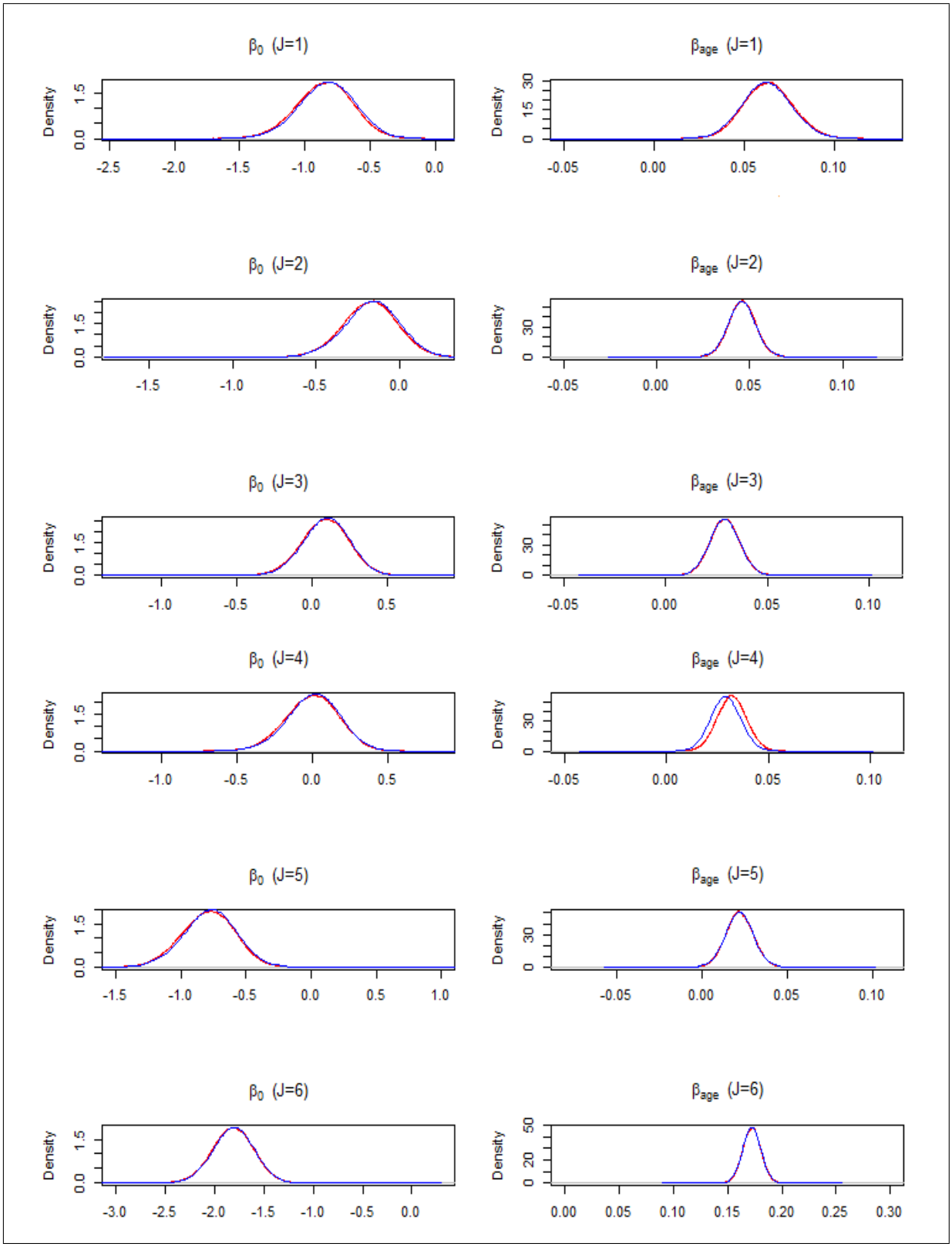

Figure 7.7: Plots of the effects of intercept and age across time intervals (J=1 to J=6) obtained via MCMC(red) and INLA-MCMC(blue)

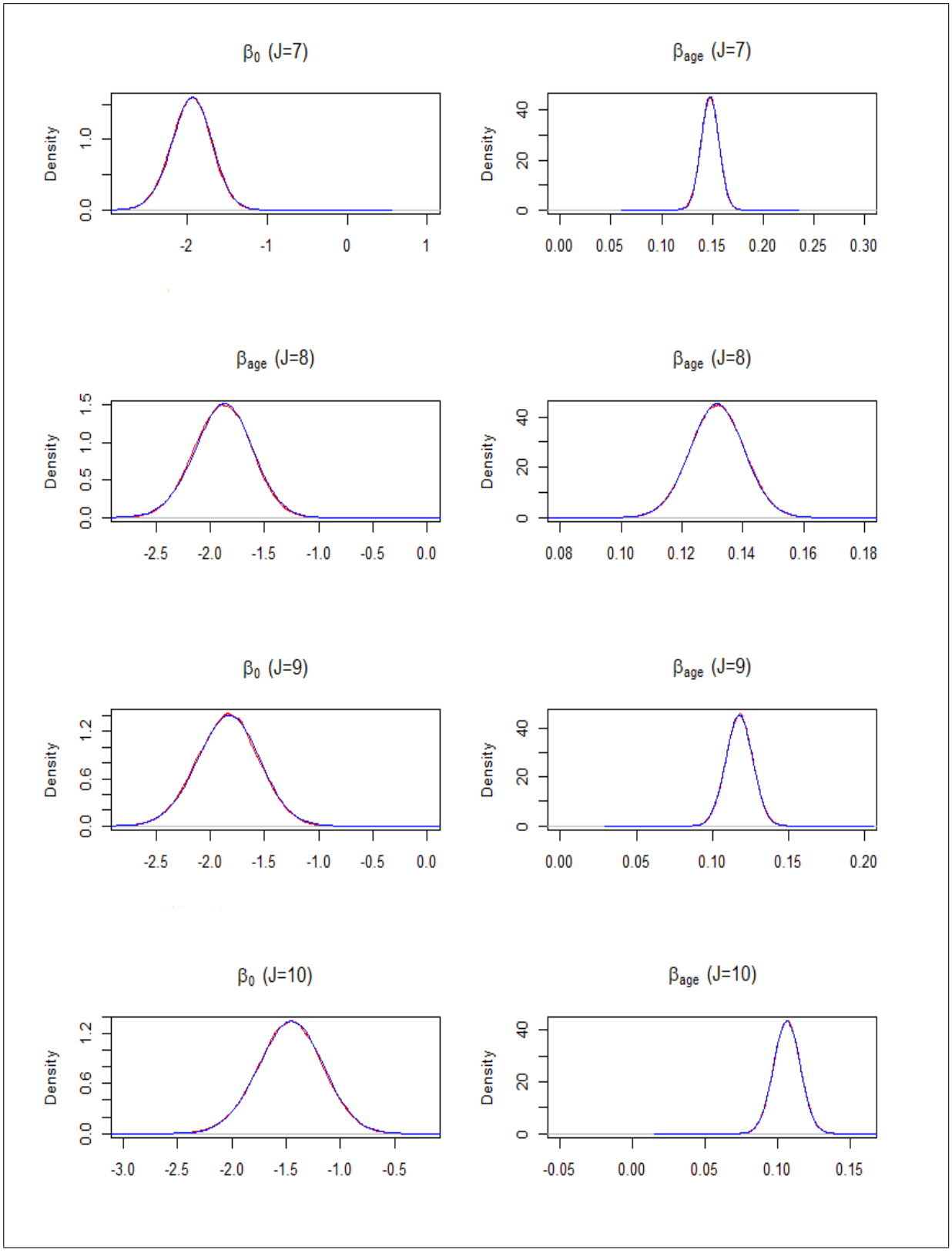

Figure 7.8: Plots of the effects of intercept and age across time intervals (J=7 to J=10) obtained via MCMC(red) and INLA-MCMC(blue)

Table 7.12: The posterior means and standard deviations (SDs) for the unknown parameters ($n = 397$)

Parameter	Post.mean (MCMC)	Post.SD (MCMC)	Post.mean (INLA-MCMC)	Post.SD (INLA-MCMC)
β_0	-1.8699	0.3950	-1.8573	0.3919
β_{Age}	-0.0044	0.0053	-0.0044	0.0053
$\beta_{Albumin}$	1.0813	0.0894	1.0794	0.0844
$\beta_{Brainmets}$	0.0314	0.0635	0.0311	0.0631
β_{Hb}	0.0038	0.0129	0.0040	0.0128
$\beta_{Neutrophil}$	-0.0479	0.0255	-0.0487	0.0255
$\beta_{Platelet}$	0.0005	0.0059	0.0005	0.0005
β_{Race}	-0.0717	0.0904	-0.0719	0.0896
β_{Sodium}	-0.0490	0.0041	0.0049	0.0032
β_{WBC}	-0.0036	0.0294	-0.0311	0.0292
α_{lung}	2.8240	0.1265	2.8220	0.1278

Table 7.13: Posterior means and standard deviations (in brackets) in each interval of time for some chosen covariates ($n = 397$)

j	τ_j	β_0 (MCMC)	β_{neut} (MCMC)	β_0 (INLA-MCMC)	β_{neut}(INLA-MCMC)
1	2.003	-3.259 (0.1718)	0.0453 (0.0093)	-3.2610 (0.1727)	0.0454 (0.0093)
2	4.374	-3.662 (0.1364)	0.0451 (0.0093)	-3.6633 (0.1368)	0.0452 (0.0093)
3	7.275	-3.273 (0.1090)	0.0461 (0.0093)	-3.2869 (0.1088)	0.0461 (0.0093)
4	11.015	-3.181 (0.0927)	0.046 (0.0092)	-3.1994 (0.0933)	0.0466 (0.0093)
5	16.286	-2.732 (0.0814)	0.0486 (0.0092)	-2.7332 (0.0816)	0.0486 (0.0093)
6	11.912	-2.539 (0.0774)	0.0485 (0.0092)	-2.5400 (0.0774)	0.0486 (0.0092)
7	25.297	-2.546 (0.0762)	0.0488 (0.0092)	-2.5463 (0.0763)	0.0488 (0.0092)

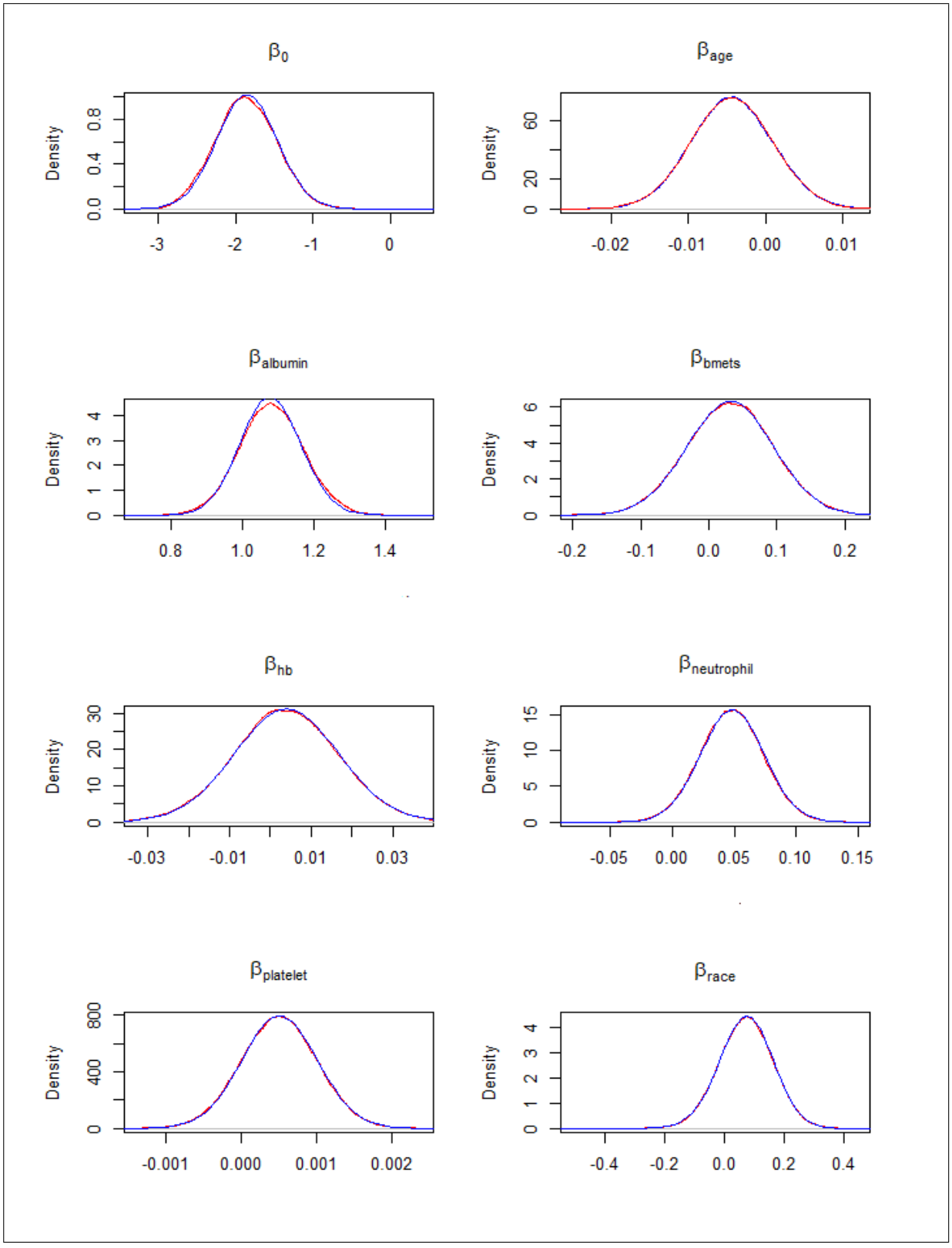

Figure 7.9: Selected plots of posterior distributions for regression coefficients obtained via MCMC (red) and INLA-MCMC(blue)

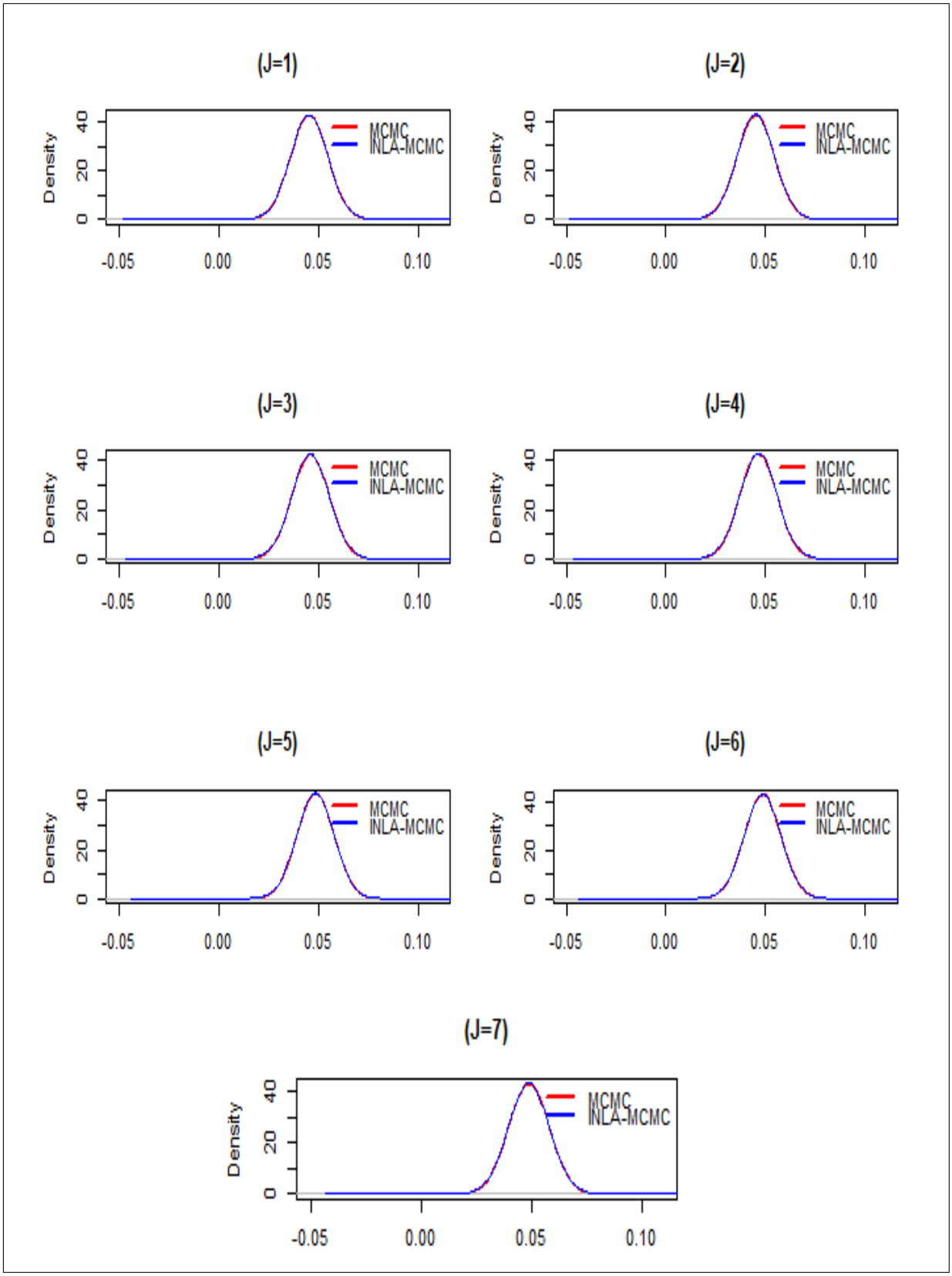

Figure 7.10: Plots of the effects of neutrophil accross time intervals (J=1 to J=7)

7.5 Summary

In this chapter, we have shown that INLA-MCMC performs as well as MCMC in obtaining posterior means and standard deviations of the parameters of our models based on its performance in subsections 7.3 - 7.4. However, INLA-MCMC has a much shorter computing time than MCMC and thus is the more efficient algorithm. We have thus been able to demonstrate the usefulness of INLA in fitting models to survival data, even when there are many cases with missing covariate values. This significantly increases the range of problems where INLA may be used efficiently since large reductions in computing time can be achieved using the new INLA-MCMC algorithm.

Chapter 8

Conclusion and Future work

8.1 Thesis summary

The main contribution of this thesis is an extension of the applicability of INLA by developing the capability to deal with missing covariate information in survival analysis. We successfully extended the INLA-MCMC algorithm (Gomez-Rubio and Rue, 2018) to deal with the situation and applied it to parametric Weibull survival models and the semiparametric piecewise constant hazard models, with and without frailties. In the piecewise constant hazard case, we used both hierarchical and autoregressive prior structures.

In Chapter 2, we introduced the fundamentals of survival analysis and Bayesian approaches. In Chapter 3, we demonstrated the fundamental properties of Gaussian Markov random fields (GMRF). We also proved the sparsity of the precision matrix, $\mathbf{Q}$, and derived the recursive formula for computing the marginal variances and GMRF distribution conditioned on linear constraints. We then demonstrated how the GMRF can be seen as a Bayesian hierarchical model and how to modify the GMRF to accommodate non-normal likelihood functions

In Chapter 4, we introduced the basic concepts of INLA. We gave two examples, a simple regression model with intercept term and a Poisson regression model for the piston failures data set, to show how INLA can be used as a fast approximation for computing posterior distributions. The basic concepts of the INLA within MCMC approach were introduced in Section 4.6. We gave an overview of the INLA within MCMC algorithm and how this algorithm can be modified to circumvent the missing covariate problem in survival analysis.

In Chapter 5, we gave an overview of the three data sets used for modelling purposes. In Chapter 6, we showed how prior information can be obtained from previous studies and combined using meta analysis. In Section 6.5, we showed how prior distributions were constructed for Weibull shape parameters, baseline log hazards, coefficients of linear parameters and a frailty variance. In Section 6.6, we demonstrated how hierarchical and autoregressive priors could be constructed for piecewise constant hazard models.

In Chapter 7, we showed how INLA-MCMC performs for three different kinds of data sets, the kidney infection, SNLG and Malaysian advanced lung cancer data sets. We also demonstrated that INLA-within MCMC is superior to MCMC in handling missing covariate information for models with reasonable numbers of parameters and hyperparameters. In our case, the performance of INLA was more efficient when fitting the piecewise constant hazard models compared to Weibull survival models. Nevertheless, INLA is still faster than MCMC, especially when computing the marginal likelihood which when combined with the conditional posterior distributions of model parameters would yield the marginal posterior distributions of our parameters.

8.2 Discussion and Limitations

8.2.1 Introduction

In this section we identify points where there are limitations in the findings so far and where further work would strengthen the findings of the thesis, throw more light on the questions addressed or provide extra information useful to potential users of the methods.

8.2.2 Simulation study

In the context of Bayesian inference, given the combination of a particular model, prior specification and data set, there is a posterior distribution for the model parameters which is obtained by applying Bayes' rule. In this thesis, we have been concerned with how to calculate summaries of the posterior distribution in the case of survival models, particularly when some values of covariates are missing.

We have compared the performance of INLA and our proposed INLA-within-MCMC method with more conventional MCMC in terms of computational speed and in terms of the ability of the INLA methods to reproduce the results obtained using MCMC. The comparisons have been done using real data sets and, in the examples where some

covariate values were missing, the missingness was the actual missingness in the real data sets. It is possible that the ability of the INLA methods to reproduce the results obtained using MCMC, and the relative computational speed might be affected by changes in the proportion of missing covariate values. This was not investigated in the thesis. To investigate this question, we might analyse data sets with different missingness proportions, increasing in steps from no missingness up to a high proportion of covariate values missing.

It may be that, rather than just the proportion of missing values, the configuration of missingness, that is which values are missing, or other features of the data might have an effect. Therefore we might randomly generate data sets with randomly allocated missingness and use a number of such data sets at each level of missingness. This whole simulation experiment would be likely to be computationally time-consuming so it might be best to use a relatively simple problem. In the case of a simple problem with few unknowns, we could also compute the posterior using numerical quadrature to provide a deterministic comparison rather than assuming that MCMC accurately gives the correct values of posterior summaries. However, this would be impractical with missing covariate values.

Another area for further investigation is sensitivity to the choice of prior distribution. It may be that, with models of the type which we investigate here, the posterior, or aspects of it, may be sensitive to the choice of prior. It may also be that there are particular types of prior distribution, for example very imprecise priors, which cause difficulties for either the INLA methods or MCMC and may lead to discrepancies between the results obtained. This will be discussed further in Section 8.2.3.2.

8.2.3 Bayesian meta analysis and sensitivity analysis

8.2.3.1 Meta analysis and prior distribution construction

In Chapter 6, detailed information from previous studies on Non-Hodgkin Lymphoma and advanced lung cancer has been presented, including the estimated values of coefficients of prognostic factors for Non Hodgkin Lymphoma. However, we only used an empirical Bayesian technique to synthesise the results which is an unsatisfactory method since we did not use a full Bayesian meta analysis. This information could be used in a more meaningful way by considering the following approaches.

Firstly, we could implement a standard Bayesian meta analysis of information obtained from previous literature combined with full prior elicitation from the experts. In this way,

better prior distributions for our parameters of interest could be obtained. For more information, refer to O'Hagan et al. (2006). Besides, we could also improve the construction of our prior distribution by extracting the individual patient data (IPD) from the published Kaplan-Meier curves using the Guyot algorithm (Guyot et al., 2012). These recreated data sets could then be used to generate prior distributions for the Weibull distribution parameters, for instance, by fitting Weibull models to these recreated data sets. However, the poor quality of the image for the available Kaplan-Meier curves published in older journals might hamper the process of IPD extraction since it is dependent upon the ability of the digital software such as DigitizeIt (Bormann, 2012) to establish accurately the coordinates of published survival curves (Guyot et al., 2012).

To account for differences in the studies over time, we could use a meta-regression technique to obtain, for instance, the prior means of our covariates of interest (Thompson and Higgins, 2002). It has been demonstrated in a number of studies that the mean covariate effect may vary across the time when the study was conducted as represented by the year of the publication of the study (Berkey et al., 1998; Bollen et al., 2007; Fiocco et al., 2012). This temporal heterogeneity of effect may be attributed to improvements in treatment modalities, changes in the characteristics of the patient cohort and other changes (Berkey et al., 1998; Bollen et al., 2007; Fiocco et al., 2012). This is particularly true in our case since there had been several published works which demonstrated pronounced temporal heterogeneity of covariate effects for the non-Hodgkin lymphoma and advanced lung cancer cases (Gao et al., 2002; Matakidou et al., 2005). Besides, meta-regression technique can also be used to adjust the other sources of heterogeneity such as study sample size, types of studies (e.g. case control vs cohort studies), different cut-off values used to classify the covariate of interests that may affect the accuracy of our prior summary and the construction of prior distributions in general (Castillo et al., 2014; Zhang and Ran, 2015). Hence, the results obtained from meta-regression analysis can clearly facilitate the construction of more accurate prior distributions for our covariates of interest.

We could also use multivariate meta-analysis methods to take into account the correlations induced by the same (within-study correlation) and different (between-study correlation) studies contributing to potentially a number of different covariate meta-analyses (Bujkiewicz et al., 2013). The within-study correlation could be assessed using the IPD obtained from other observational cohort studies or clinical trials whilst the between-study correlation could be evaluated using the summary statistics from individual studies. However, caution should be exercised since the use of multivariate meta analysis methods

might inflate the degree of uncertainty around the studied outcomes instead (Cooper et al., 2011). Possible causes include the heterogeneity in treatment effects, baseline characteristics of study participants and study designs (Song et al., 2001).

We could also perform random effect meta-analysis using the combination of IPD and the summary statistics from previous studies to generate posterior predictive distributions. Ades et al. (2005) and Dias et al. (2013) reasoned that, in the presence of heterogeneity of effects between studies, the predictive distribution of a future study is a more cogent description of the degree of uncertainty around the effect of interest than the distribution of mean effect. As a result, the predictive distributions obtained from prior studies can subsequently be used to derive the prior distributions for the analyses that had been undertaken in our study.

Alternatively, we could also use the power prior distributions for regression models as proposed by Ibrahim and Chen (2000) to construct informative prior distributions. Briefly, suppose that historical data from previous studies is given by $D_0 = (n_0, y_0, X_0)$, where n_0 represents the sample size of the previous study, y_0 is the $n \times 1$ response vector, X_0 is the $n \times p$ matrix for covariates and θ represents the indexing parameters. The power prior distribution of θ for the present study is then given by

$$\pi(\theta \mid \alpha_0, D_0) \propto L(\theta \mid D_0)^{\alpha_0} \; \pi_0(\theta \mid c_0)$$

where c_0 is the hyperparameter for the initial prior (before knowing the existence of the historical data), α_0 is the scalar parameter representing the weights of the historical data relative to the likelihood of the current study, $L(\theta \mid D_0)$ is the likelihood of θ in the current study given the historical data, D_0. This hierarchical power specification can be made complete by assigning a prior distribution for α_0, resulting in the joint power prior distribution for θ and α_0 that is represented by

$$\pi(\theta, \alpha_0 \mid D_0, \alpha_0) \propto L(\theta \mid D_0)^{\alpha_0} \; \pi_0(\theta \mid c_0) \; \pi(\alpha_0 \mid \gamma_0)$$

where γ_0 is the hyperparameter vector. One of the candidates for $\pi(\alpha_0 \mid \gamma_0)$ is a beta prior. Nevertheless, priors that have either truncated gamma or truncated normal form may also serve as alternatives since all of them possess the same computational and theoretical consistency (Ibrahim and Chen, 2000).

8.2.3.2 Sensitivity to prior distribution

The results obtained using the informative priors should be evaluated with respect to their sensitivity to changes in prior specification (sensitivity analyses). We could repeat the analyses, starting with very imprecise priors and increasing the prior precision in steps to assess the effects. However, the effects on the results might depend on the combination of particular features of the data with the degree of precision in the prior. Therefore, as in Section 8.2.2, multiple simulated data sets should be generated and by these, the sensitivity of results obtained to the different prior specifications could be adequately investigated.

8.2.4 Predicted survival curves

Using the analyses of Chapter 7, we could compute predicted survival curves for patient groups with a specific covariate profile. In fact, since we have a missing data model, we could obtain the predicted survival curves without specifying the full covariate profile. For instance, if we would like to obtain a predicted survival curve for a male Non-Hodgkin Lymphoma (NHL) patient aged 60, we could do this by integrating over the conditional distributions of the unspecified covariates. This can be done easily in MCMC and INLA-MCMC since we can use MCMC sampling for the missing covariate values. In both cases, a "dummy patient" could be included in the analysis and a survival function could then be computed at a range of time points for this patient. We could then sample the survival function values and since the survival function is a probability, and therefore an expectation, we could subsequently find the posterior mean of the survival function at each of the time points. These could then be plotted against time to obtain the predicted survival curve for that particular patient.

8.3 Future work: Multi-state survival models

8.3.1 General overview of multi-state survival models

The inspiration to use a multi-state model for lung cancer patients comes from the seminal work of Armero et al. (2016) on a disability model for lung cancer. Multistate models are a class of stochastic models that involve the probability of occupying a number of discrete states in continuous time. These natural stochastic models are practical for the modelling of discrete systems' evolution. There are two main types of multi-state models: K-progressive model and illness disability model (also known as disability model) (Meira-Machado et al., 2009). For the former, the simplest case will be a model with two states such as dead and alive state with merely one transition probability. Nevertheless, this can be generalised to k-states as shown in Figure 8.1.

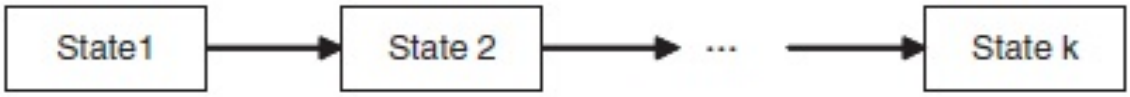

Figure 8.1: K-progressive model (Meira-Machado et al., 2009)

As an example, let us consider a three-state model for a hypothetical scenario of survival of breast cancer patients which is illustrated in Figure 8.2.

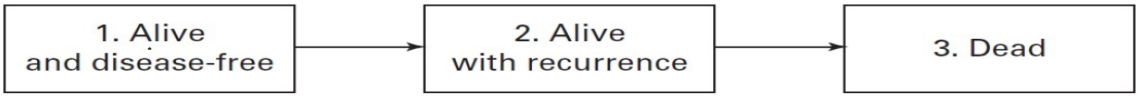

Figure 8.2: Breast-cancer three-state model (adapted from (Meira-Machado et al., 2009))

One possible approach is to decouple the whole multi-state model into a variety of survival models. This is achieved by fitting distinct hazard functions for all possible transitions using proportional hazard regression models whilst at the same time making suitable adjustment to the risk set. The effects of covariates can be assumed to act linearly on the log hazards. However, this is a very simplified version of modelling multi-state survival models and this makes it less useful if the covariate effects do not have constant linear effects on the log hazards.

Another alternative is dependent on whether the Markov assumption is completely fulfilled. This will result in two different models: Cox-Markov model or Cox semi-Markov model. In the former, the Markovian assumption is completely fulfilled, by which the future states that an individual will occupy will only depend on the current state. Within this fully Markovian framework, we may assume that the hazard function, h_{vj} (t;($\mathbf{Z}$)) = h_{vj_0}(t) exp ($\boldsymbol{\beta}_{vj}^T \boldsymbol{Z}$), where h_{vj} (t;($\mathbf{Z}$)) is the hazard function for transitions from state v to state j with covariate vector $\mathbf{Z}$, $h_{vj_0}(t)$ is a non-negative baseline hazard function and $\boldsymbol{\beta}_{vj}$ is a vector of regression parameters. The number of covariates, which may be, continuous or categorical, is p.

Nevertheless, the complete Markov assumption is not truly tenable in the real-life setting. Under a full Markovian framework, we assume that the future health of patients who were recently afflicted with breast cancer is similar to that of breast cancer patients who have been sick for a long time. Besides, for patients who experience recurrent breast cancer (exemplified by state 2 in Figure 8.2), we may be interested in the sojourn time of patients in state 1 (alive and healthy). Hence, the Cox semi-Markov model may be used for the modelling purposes. The difference with this approach from the previous full-Markovian approach is really subtle: the future of the states is not dependent on the absolute time at the current state (i.e. the time lapsed since the entry of a subject into the initial state), but rather on the time duration spent in the current state (Foucher et al., 2007). Hence, the difference lies in the way we model the transition from state 2 to state 3. The hazard functions are given by

$$h_{12}(t; \boldsymbol{Z}) = h_{12_0}(t) \exp(\boldsymbol{\beta}_{12}^T \boldsymbol{Z})$$

and

$$h_{23}(t - T_{12}; \boldsymbol{Z}) = h_{23_0}(t - T_{12}) \ \exp(\boldsymbol{\beta}_{23}^T \boldsymbol{Z})$$

where T_{12} is the time when a patient enters into state 2 and $t - T_{12}$ is the time duration spent in state 2. This is similar to resetting the clock since, when a patient enters a new state, the clock time is reset to zero again.

To address the possible non-linearity effects of covariates on the hazard function, we may employ additive multi-state models, by which the transition of individuals from one

state to another state can be modelled by the hazard function

$$h_{vj_0}(t)\exp\left(\sum_{i=1}^{q} f_{i,vj}(Z_i)\right)$$

where $f_{i,hj}(.)$, i=1,2,3,. . .,q represent smooth functions of the covariates for breast cancer survival. This approach is more useful than the previous one since we relax the assumption of a linear effect of covariates on breast cancer survival. Hence, we assume such covariate effects may affect the breast cancer survival via unspecified smooth functions.

For the second type of model (illness-disability model, also known as illness-death model), the individuals or patients were disease-free initially. Subsequently, they contracted the disease (state 2) and died (state 3) later. For non-lethal diseases, the patients may return to state 1 again after receiving successful treatments. If this is a possibility, the model can be styled as a *bi-directional model*. Apart from that, patients may also experience death instantaneously, thus bypassing state 2. Figure 8.3 illustrates an example of an illness-disability model with three states.

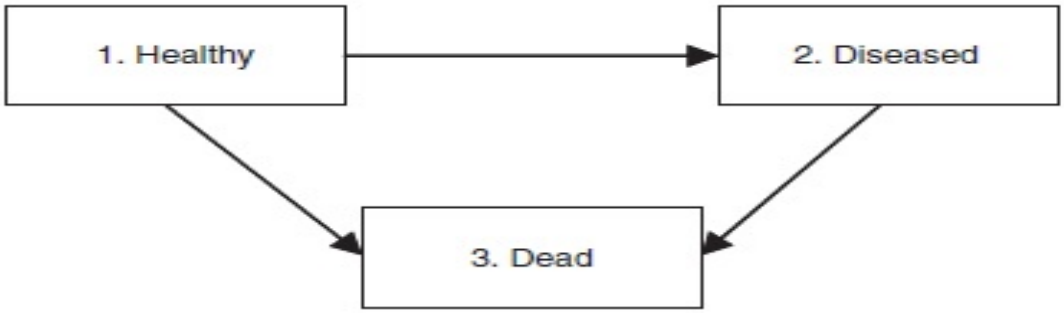

Figure 8.3: Illness-disability model (Meira-Machado et al., 2009)

As a further example, let us consider an illness-disability model for surgically-operable cancer patients who later experience recurrence, distant spread (metastases) and death. For clarity, Figure 8.4 gives a further example of such a model.

8.3.2 Multi-state models for lung cancer survival

Borrowing the methodological framework used by Armero et al. (2016) a stage IV lung cancer patient will occupy the following three states at a specific time t during the course of their diseases:

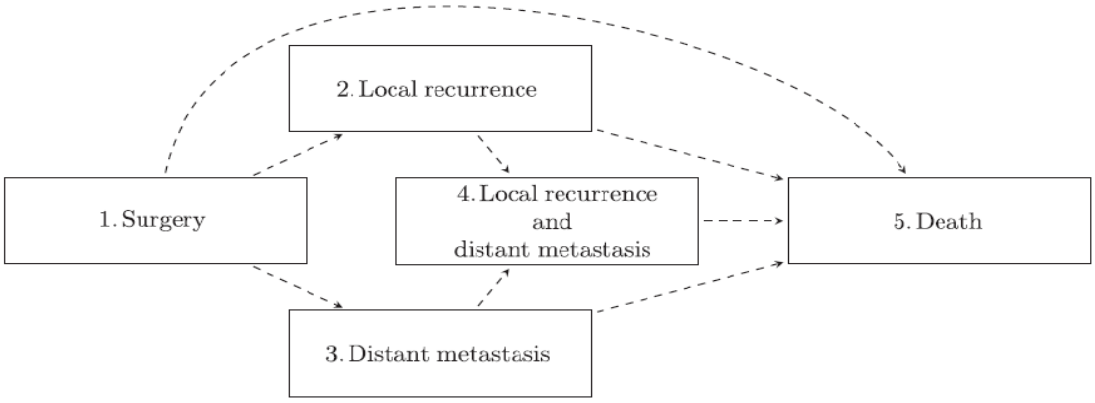

Figure 8.4: Multi-state model for surgically-operable cancer patients (Putter et al., 2007)

- **State I:** Stable disease (defined according to Response Evaluation Criteria in Solid Tumor version 1.1 (RECIST) criteria (Eisenhauer et al., 2009)).
- **State II:** Progressive disease (increase in diameters of the tumors based on RECIST criteria).
- **State III:** Death.

Hence, we can consider state 1 and state 2 to be transitory states and state 3 as an absorbing state. The illness-disability model for this simplified version of a lung cancer multi-state model can be further elucidated by Figure 8.5.

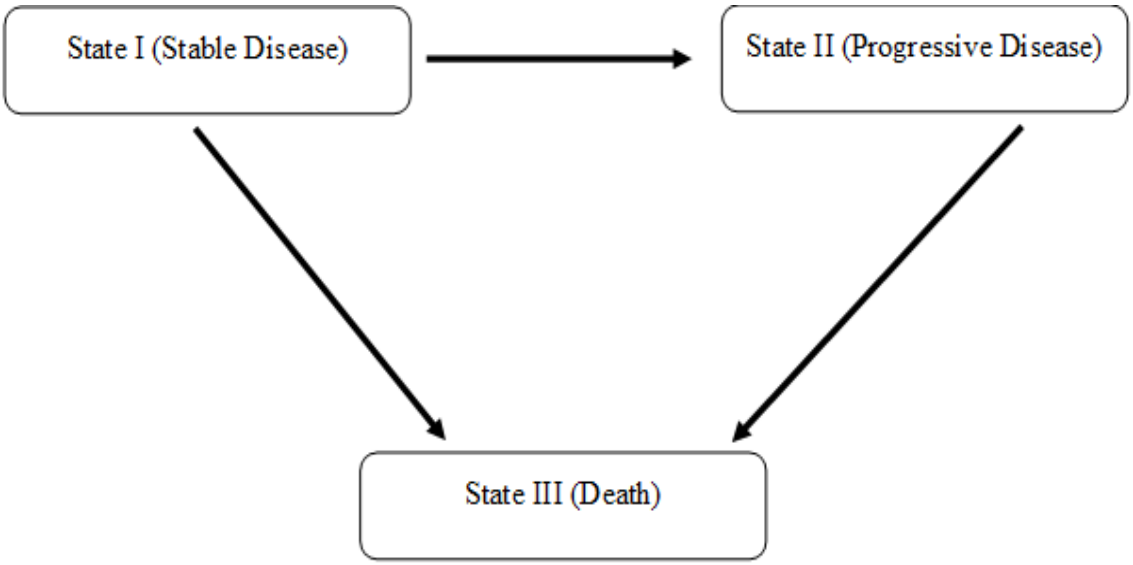

Figure 8.5: Multi-state model used by Armero et al. (2016)

For this scenario, we can assume that transitions between states are determined by a parameter vector, $\boldsymbol{\theta}$ and covariates $\mathbf{x}$. For individuals in state 2, the transition probability from state 2 (progressive stage IV lung cancer) to state 3 (death) is also contingent upon residence time in state 1 (symbolised by T_{12}). These can be represented by the equations

$$p_{1j} = (s, t \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathrm{P}(Z(t) = j \mid Z(s) = 1, \boldsymbol{x}, \boldsymbol{\theta}), s \leq t, j = 2, 3$$

and

$$p_{23} = (s, t \mid \boldsymbol{x}, \boldsymbol{\theta}, t_{12}) = \mathrm{P}(Z(t) = 3 \mid Z(s) = 2, \boldsymbol{x}, \boldsymbol{\theta}, T_{12} = t_{12}), \qquad t_{12} \leq s \leq t.$$

To determine these two probabilities, the hazard function for times between transition, T_{ij}, may be employed since it is congruent with inter-state transition intensities. These can be represented by the two equations

$$h_{1j}(t \mid \boldsymbol{x}_{1j}, \boldsymbol{\theta}_{1j}) = \lim_{\delta t \to 0} \left\{ \frac{\mathrm{P}(T \leq T_{1j} < t + \Delta t \mid T_{1j} \geq t, \boldsymbol{x}_{1j}\boldsymbol{\theta}_{1j})}{\Delta t} \right\}, j = 2, 3$$

and

$$h_{23}(t - t_{12} \mid \boldsymbol{x}_{23}, \boldsymbol{\theta}_{23}, t_{12}) = \lim_{\delta t \to 0} \left\{ \frac{\mathrm{P}(A \mid B, \boldsymbol{x}_{23}, \boldsymbol{\theta}_{23}, T_{12} = t_{12})}{\Delta t} \right\}$$

where A is the event $t - t_{12} \leq T_{23} < t - t_{12} + \Delta t$ and B is the event $T_{23} \geq t - t_{12}$. For a homogenous illness-disability model that posesses a semi-Markovian property (v to j transition intensity at a time point is also dependent upon sojourn time at state v), the relationships between the hazard functions and the transition probabilities can be linked via the equations

$$p_{11}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}) = \exp\Big\{ - \int_s^t [h_{12}(u \mid \boldsymbol{x}_{12}, \boldsymbol{\theta}_{12}) + h_{13}(u \mid \boldsymbol{x}_{13}, \boldsymbol{\theta}_{13})] du \Big\},$$

$$p_{22}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}, t_{12}) = \exp\Big\{ - \int_s^t h_{23}(u - t_{12} \mid \boldsymbol{x}_{23}, t_{12}) du \Big\},$$

$$p_{12}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}, t_{12}) = \int_s^t p_{11}(s, u \mid \boldsymbol{x}, \boldsymbol{\theta}) h_{12}(u \mid \boldsymbol{x}, \boldsymbol{\theta}_{12}) p_{22}(u, t \mid \boldsymbol{x}, \boldsymbol{\theta}, u) du,$$

$$p_{13}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}) = 1 - p_{11}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}) - p_{12}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}),$$

$$p_{23}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}, t_{12}) = 1 - p_{22}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}, t_{12}),$$

and

$$p_{33}(s, t \mid \boldsymbol{x}, \boldsymbol{\theta}) = 1.$$

Nevertheless, a fourth state can be added to the original ideas of Armero et al. (2016). The fourth state is called a responsive disease state (can be either Partial Response (PR) [defined as 30 percent reduction in the sum of diameters of target lesions, using the baseline diameters as reference] **OR** Complete Response (CR) [defined as total disappearance of the target lesion](Eisenhauer et al., 2009; Schwartz et al., 2016). The fourth state can be anticipated to occur quite frequently in Malaysian advanced lung cancer cases since the use of Epidermal Growth Factor Receptor-Tyrosine Kinase Inhibitors (EGFR-TKIs), which is a form of molecularly-targeted therapy for lung cancer that specifically targets the highly expressed cellular growth-promoting signalling cascades, resulting in a total disappearance of target lesions (primary tumor mass in the lungs) in those harbouring EGFR mutations, may cause a dramatic improvement in the survival of the subjects. Therefore, for the Malaysian advanced lung cancer dataset, the following rearrangement of the state space is proposed.

- **State I:** Responsive disease:A minimum reduction of 30% in the total diameters of target lesions compared to their baseline measurements (Eisenhauer et al., 2009)
- **State II:** Stable disease: An inadequate reduction or increase in diameters of the tumors to qualify for responsive disease or progressive disease (Schwartz et al., 2016)
- **State III:** Progressive disease (RECIST 1.1 criteria)
- **State IV:** Death (absorbing state)

Therefore, states I, II and III are considered as transient states and state IV as an absorbing state. The movements between states are governed by the patient's health status. The jump from state I or II to state III occurs when cancer progresses, from state III to IV when cancer patients die because of disease progression, from state I or II directly to state IV if the patient dies because of other reasons (not due to disease progression). There can also be reverse jumps from state III to either state II or I since cancer may stop progressing if new therapeutic regimens are given to the patients who have failed the first line regime (which they might also have previously responded to), resulting in a second complete or partial response period. There will be no jump from state I to state II since they are mutually exclusive. Figure 8.6 summarises the four-state space.

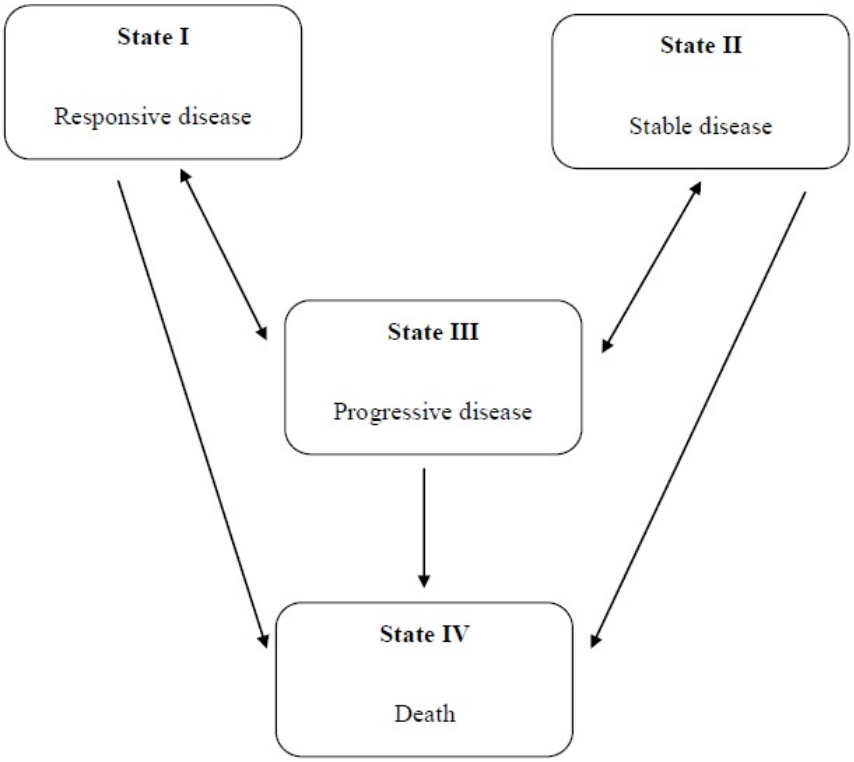

Figure 8.6: State space for the advanced model

8.3.3 Renewal process and non-homogenous Poisson process in survival analysis

Much of the time-to-event modelling for cancer data assumes a non-repairable system. This type of system can only experience a single failure and a parametric time-to-event model using a Weibull distribution will aptly summarise the distribution of the time when the system fails. In contrast, a repairable system presumes that failed items can be restored back into service. For instance, a damaged car can be repaired and refitted back into service, whilst severely damaged cardiac tissues are no longer repairable, requiring a new organ transplantation for such patients. Hence, models for repairable systems have to give allowances to a full sequence of repeated failures and it must also have the capability to reflect the alterations in the system's reliability as time progresses. A counting process approach is thus suitable for modelling a repairable system in survival analysis.

Let N(t) represent the number of failures of a repairable system in one interval [0, t]. Since N(t) is non-negative integer valued, the difference between N(t) and N(s) is the number of failures occurring in the time interval between s and t (provided $s < t$)

(Engelhardt and Bain, 1992). An alternative specification can also be made with respect to sequential failure times: T_1,T_2,...,T_n (Lindqvist et al., 2003). If a system is fully repairable into a new condition, this system can thus be modelled using a parametric assumption with independent and identically distributed survival times (renewal process). However, the modelling of time-to-event data for cancer cases is far more complicated. For instance, the time to chemotherapeutic failure between successive regimes of chemotherapy becomes shorter due to the mutation in cancer cells that confers further selective resistance towards different combinations of chemotherapy. Besides, for curable cancer types, the times between relapses become progressively shorter as evident from the shorter time to relapse for non-Hodgkin lymphoma patients who are cured by the second chemotherapeutic regime, compared to their survival time when they were treated with the first chemotherapeutic regime. Therefore, the repairable system in cancer cases can be assumed to have a profound decrease in its reliability as time progresses (ie. as the system ages).

To illustrate further, we may view the general condition where the renewal process of a series of events is observed for every patient such that the the inter-event times X_j, j = 1,2,... for a patient are iid. Hence, we may imagine a number of individual renewal processes which are censored at possibly different times. Thus, the following scenarios for censoring times may occur (Aalen and Husebeye, 1991):

- The censoring times can be considered to be fixed in advance. The example is at the cessation of a fixed observation period.
- The censoring time itself can be considered as a random variable which is not dependent upon the renewal process. Hence, every patient can be considered to have distinct censoring times or one censoring time that is shared by several patients.
- Censoring time occurs after each patient has completed a predetermined number of renewal periods.
- Censoring time happens when the individuals have completed a certain number of periods or the observation times have exceeded certain limits.
- Censoring time occurs at the first event occuring after a predetermined time-interval.

The other alternative to a renewal process, which can be considered as a generalised homogeneous Poisson process, is to use a non-homogeneous Poisson process (NHPP) to

model cancer survival data. Engelhardt and Bain (1992) showed that this process has a non-constant intensity function as given by

$$\nu(t) = (\beta/\theta)(t/\theta)^{\beta-1}.$$

An intensity function should not be confused with a hazard function since the former is the absolute rate of failure for a repairable system whilst the latter is the relative rate of failure for non-repairable systems (Deshpande et al., 1999). A Weibull process has a mean value function, $M(t)$, given by

$$M(t) = (t/\theta)^{\beta}; \theta > 0, \beta > 0$$

where θ is the scale parameter whilst β is the shape parameter (Engelhardt and Bain, 1992). Besides, Engelhardt and Bain (1992) also demonstrated that $M(t)$ can be alternatively reparameterised as

$$M(t) = \lambda t^{\beta}.$$

If the second parametrisation of $M(t)$ is substituted into the intensity function, $\nu(t)$, this leads to parameterisation of $\nu(t)$ as

$$v(t) = \lambda\beta t^{\beta-1}$$

To model cancer survival data that have the properties of a non-homogenous Poisson process (NHPP), we could use a semiparametric model as proposed by Sinha (1993) based on the conditional intensity function,

$$h(t|z_i, w_i) = I(t)h_0(t)\exp(\beta^T z_i)w_i$$

where w_i is a frailty effect that is patient-specific, z_i is a vector of covariates for the i$^{\text{th}}$ individual, β^T is the vector of unknown parameters, h_0 is the baseline intensity function and I(t) is the indicator of the censoring process. Therefore, w_i can be considered as a multiplicative random effect on the hazard function for a specific patient or individual. The baseline hazard function can be modelled using a piecewise constant hazard function. However, there are a few difficulties in computing the posterior distributions for this model under the INLA framework. Firstly, frailty is usually regarded to be an iid gamma random variable of mean 1 and unknown variance κ

$$W_i \sim Ga(\kappa^{-1}, \kappa^{-1}),$$

for each individual $i = 1,\ldots,N$. As a result, this creates a cumbersome situation if INLA is used for obtaining the posterior distributions. To circumvent this, several alternatives such as using the log normal distribution for frailty could be explored so that the Weibull process can be rewritten into a latent Gaussian model, a prerequisite for Bayesian inference using INLA methodology.

8.4 Conclusions

The main goals of this research were as follows:

1. Compare the performance of different computational approaches (MCMC vs INLA) in Bayesian inference for survival analysis models.
2. Investigate the construction of appropriate forms for the prior distributions of the log hazard, Weibull shape parameters and frailty.
3. Assess the performance of INLA in missing covariates values and develop approaches to improve the flexibility of INLA in this setting.
4. Apply the formulated approaches to practical problems using real medical data sets and evaluate the strengths and limitations of each approach

In Chapter 7, we compared the performance of MCMC and INLA and we showed that INLA was superior to MCMC when obtaining the posterior means and standard deviations of our parameters of interest. In Chapter 6, we showed that the appropriate forms of prior distributions for the log hazard could be computed for both the SNLG and Malaysian advanced lung cancer data sets. We also showed in 6.6, the methods used to construct the prior distributions for the log hazard, Weibull shape parameters and frailty variance. We fulfilled objectives 3 and 4 in Chapter 7 by showing that INLA-MCMC can appropriately handle missing covariate data problems in different kinds of real medical data sets and the strengths and limitations of each approach were described in Chapter 7 and summarised in Chapter 8.

Appendix A

Appendix

A.1 Standard Distributions Used in Survival Analysis

To facilitate better understanding of the model structures developed in this research project, standard distributions used in survival analysis will be elaborated.

A.1.1 Weibull and Exponential Distributions

The Weibull and exponential distributions are the two most commonly distributions used in analysing the lifetimes of subjects. The two-parameter Weibull distribution has the following form:

$$f(x \mid a, b) = \frac{b}{a}\Big(\frac{x}{a}\Big)^{b-1} \exp\Big\{ -\Big(\frac{x}{a}\Big)^{b} \Big\}$$

with b $>$ 0 is the shape parameter and a$>$ is the scale parameter. If k $<$ 1, this means that the failure rate declines as time progresses. If k$>$ 1, the failure rate increases over time. A special case occurs when the failure rate is a constant over time (k =1), where the Weibull distribution is reduced to exponential distribution:

$$f(x \mid a, b) = \frac{1}{a}\, \exp\Big\{ -\frac{1}{a}\, x \Big\}$$

Then we substitute $\frac{1}{a} = \lambda$, resulting in;

$$f(x \mid \lambda) = \lambda\, \exp\Big\{ -\lambda\, x \Big\}$$

with $\lambda > 0$ as the rate parameter.

A.1.2 Lognormal distribution

Let X is a random variable whose logarithm has a Gaussian distribution. Hence, Z = log X is distributed normal. The pdf of this continuous distribution is given by:

$$f(z \mid \mu, \sigma^2) = \frac{1}{\exp(\mathrm{z})\sqrt{2\sigma^2\pi}} \exp\Big\{ - \frac{(\log x - \mu)^2}{2\sigma^2} \Big\}$$

where $\sigma^2 > 0$, $-\infty \leq \mu \leq +\infty$. This distribution is useful for an alternative to gamma distrbution if X $\in$ (0, ∞).

A.1.3 Gamma distribution

Let Y is a random variable that has a positive support and follows a gamma distribution, with a scale parameter,h, and shape parameter g. Hence, the pdf of Y is given as follows:

$$\mathrm{f}(y \mid g, h) = \frac{h^g \; y^{g-1} \exp\big\{ - hy \big\}}{\Gamma \; (g)}$$

where g and h > 0. Γ(g) is a gamma function. This distribution is usually employed for a parameter that has a strictly positive support. One of the examples is the variance of a normal or log-normal distribution.

A.1.4 Log-logistic distribution

Let R is a random variable that has only positive support. Hence the pdf is given by:

$$\mathrm{f}(r \mid \alpha, \beta) = \frac{(\frac{\beta}{\alpha})(\frac{r}{\alpha})^{\beta-1}}{(1 + (\frac{r}{\alpha})^{\beta})^2}$$

where both α and β are strictly positive. This distribution is usually used in survival analysis for modelling events whose rate escalates initially and diminishes eventually. As an example, mortality rate may follow this kind of distribution.

A.1.5 Multivariate normal distribution

This distribution is a Gaussian distribution that has been generalised from one-dimensional univariate Gaussian to multi-dimensional Gaussian distribution. Let say $\mathbf{X}$ is a p-dimensional random vector. The pdf of this distribution is thus given by:

$$\mathrm{f}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{p}{2}} \mid \boldsymbol{\Sigma} \mid^{\frac{1}{2}}} \exp\Big\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \Big\}$$

where $\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu}$ and $\mathrm{Var}[\mathbf{X}] = \boldsymbol{\Sigma}$, which denotes the positive-definite covariance matrix.

A.1.6 Beta distribution

Let V is a continuous random variable whose support is between 0 and 1. Hence, V follows a beta distribution that has a pdf as follows:

$$\mathrm{f}(v \mid \alpha, \beta) = \frac{v^{\alpha-1}(1-v)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$$

where α and β are shape parameters that are strictly positive. B represents the beta function that has the following form:

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\ \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Beta distribution is popular with respect to Bayesian inference because it is a conjugate prior distribution for binomial and geometric distributions.

A.1.7 Multinomial distribution

It is a generalised version of binomial distribution. In n number of multinomial trials that are independent, there are k possible outcomes having fixed success probability. Hence, multinomial distribution is hence the distribution of the outcomes determined by multinomial trials. Let $Q_1, Q_2, \ldots, Q_k$ with $\mathrm{N} = \sum_{i=1}^{k} Q_i = 1$.The probability mass function (pmf) of multinomial distribution is given by:

$$\Pr(Q_1 = q_1,\ Q_2 = q_2\ Q_3 = q_3, \ldots, Q_k = q_k) = \frac{n!}{q_1!, q_2!, q_3!, \ldots, q_k!} p_1^{q_1}\ p_2^{q_2}\ p_3^{q_3} \ldots p_k^{q_k}$$

where $\sum_{i=1}^{k} q_i = n$. Two special cases of this distribution should be noted; a) in a single trial ($n = 1$) with $k = 2$ outcomes, this distribution is reduced to Bernoulli distribution, b) in n independent trial with $k = 2$ outcomes, this multinomial distribution is reduced to binomial distribution.

A.1.8 Dirichlet distribution

This distribution plays a very special role in Bayesian inference, it is a conjugate prior distribution for the multinomial distribution. The pdf is given by:

$$\mathrm{f}(\mathbf{x}) = \frac{\Gamma\Big(\sum_{i=1} a_i\Big)}{\prod_i^d \Gamma\Big(a_i\Big)} \prod_{i=1}^{d} {x_i}^{a_i - 1}$$

where $\mathbf{X} \in \mathcal{S}^d$ and $\mathbf{a} \in \mathbb{R}_+^d$. The expectation and covariance are given as follows:

$$E[x_j] = \frac{a_j}{A}$$

$$\mathrm{Cov}(x_i, x_j) = -\frac{a_i\ a_j}{A^2(A+1)}$$

$$\mathrm{A} = \sum_{i=1}^{d} a_i$$

It can clearly be seen that if d=2, this distribution will be reduced to the beta distribution.

A.2 The Structure of SNLG dataset

The SNLG dataset has the following format (structure):

Table 1A: The structure of partial SNLG data

"t"	"t.cen"	"is.cen"	"age"	"hb"	"wbc"	"gender"	"albumin"
0.0164	1.0164	0	12	-4.706	-6.069	1	2
0.2502	1.2502	0	3	-17.706	-5.970	1	1
0.1689	1.1689	0	-2	-25.706	-5.770	2	1
0.1945	1.1945	0	-8	-42.706	-4.970	2	2

The negative signs for age, haemoglobin level (hb) and white blood cell (wbc) count are attributed to the centering of the continuous quantitative covariates.

A.3 RJAGS and INLA codes for piston-ring failures poisson regression model

```
------------------------------------------------------------
R-JAGS CODES FOR PISTON-RING FAILURES POISSON REGRESSION MODEL
------------------------------------------------------------
modelstring="
model{
for (i in 1:n){
y[i]~dpois(mu[i])
log(mu[i])<-alpha+beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+beta4*
x4[i]+beta5*x5[i]+beta6*x6[i]+beta7*x7[i]+beta8*x8[i]+beta9*
x9[i]+beta10*x10[i]+beta11*x11[i]
}
alpha~dnorm(2.3,0.4545)
beta1~dnorm(0,2)
beta2~dnorm(0,0.6667)
beta3~dnorm(0,0.6667)
beta4~dnorm(0,0.6667)
beta5~dnorm(0,0.6667)
beta6~dnorm(0,4)
beta7~dnorm(0,4)
beta8~dnorm(0,4)
beta9~dnorm(0,1.3333)
beta10~dnorm(0,1.3333)
beta11~dnorm(0,1.3333)
}
"
------------------------------------------------------------
R-INLA CODES FOR PISTON-RING FAILURES POISSON REGRESSION MODEL
------------------------------------------------------------
```

```
formula3=PistonINLA$Failures~PistonINLA$Beta1+PistonINLA$Beta2+
PistonINLA$Beta3+PistonINLA$Beta4+PistonINLA$Beta5+PistonINLA$
Beta6+PistonINLA$Beta7+PistonINLA$Beta8+PistonINLA$Beta9+
PistonINLA$Beta10+PistonINLA$Beta11

model1=inla(formula3,family="poisson",data=PistonINLA
,control.fixed=list(mean=list(Beta1=0,Beta2=0,Beta3=0,
Beta4=0,Beta5=0,Beta6=0,Beta7=0,Beta8=0,Beta9=0,
Beta10=0,Beta11=0),mean.intercept=2.3,prec=list(Beta1=2,
Beta2=0.6667,Beta3=0.6667,Beta4=0.6667,Beta5=0.6667,Beta6=4,
Beta7=4,Beta8=4,Beta9=1.3333,Beta10=1.3333,Beta11=1.3333),
prec.intercept=10.4545),verbose=TRUE,control.compute=list(dic=TRUE,
cpo=TRUE),control.inla=list(strategy="simplified.laplace"))
```

A.4 R function for the numerical quadrature method used for the analysis of kidney infection data set using exponential proportional hazard model

```
kidney2<-function(eta1,eta2,t,status,placement,prior,alpha=1)
 # Evaluates posterior density etc for Weibull model for kidney dialysis data.
 # Exponential model is obtained when alpha is set to 1
 # prior is mean0, mean1, var0, var1, covar for beta
 # placement is 1 for surgical and 2 for percutaneous
 # status is 1 for event, 0 for censored
n1<-length(eta1)
n2<-length(eta2)
step1<-eta1[2]-eta1[1]
step2<-eta2[2]-eta2[1]
eta1<-matrix(eta1,nrow=n1,ncol=n2)
eta2<-matrix(eta2,nrow=n1,ncol=n2,byrow=T)
m1<-prior[1]+prior[2]
```

```
m2<-prior[1]+2*prior[2]
v1<-prior[3]+prior[4]+2*prior[5]
v2<-prior[3]+4*prior[4]+4*prior[5]
c<-prior[3]+2*prior[3]+3*prior[5]
sd1<-sqrt(v1)
sd2<-sqrt(v2)
r<-c/(sd1*sd2)
delta1<-(eta1-m1)/sd1
delta2<-(eta2-m2)/sd2
d<-1-r^2
logprior<- -(delta1^2 + delta2^2 - 2*r*delta1*delta2)/(2*d)
nd<-sum(status)
t1<-t[placement==1]
t2<-t[placement==2]
status1<-status[placement==1]
status2<-status[placement==2]
t1d<-t1[status1==1]
t2d<-t2[status2==1]
lt1d<-log(t1d)
lt2d<-log(t2d)
t1a<-t1^alpha
t2a<-t2^alpha
nd1<-length(t1d)
nd2<-length(t2d)
beta0<-2*eta1-eta2
beta1<-eta2-eta1
lambda1<-exp(eta1)
lambda2<-exp(eta2)
loglik<-nd*log(alpha)+nd1*eta1+nd2*eta2+(alpha-1)*(sum(lt1d)+sum(lt2d))-
lambda1*sum(t1a)-lambda2*sum(t2a)
logpos<-logprior+loglik
logpos<-logpos-max(logpos)
posterior<-exp(logpos)
int<-sum(posterior)*step1*step2
posterior<-posterior/int
```

```
postmean0<-sum(posterior*beta0)*step1*step2
postmean1<-sum(posterior*beta1)*step1*step2
postvar0<-sum(posterior*((beta0-postmean0)^2))*step1*step2
postvar1<-sum(posterior*((beta1-postmean1)^2))*step1*step2
postmean<-c(postmean0,postmean1)
postsd<-sqrt(c(postvar0,postvar1))
ans<-list(density=posterior,postmean=postmean,postsd=postsd)
ans
}
----------------------------
#RUNNING THE MODEL
----------------------------
data<-read.table("Kidney-infec.txt",header=T)
t<-data$time
status<-data$event
placement<-data$placement
source("kidney2R.txt")
prior2<-c(0,0,1000,1000,0)
eta1<-seq(-2,-0.5,0.002)
eta2<-seq(-2.6,-0.8,0.005)
test<-kidney2(eta1,eta2,t,status,placement,prior2)
```

A.5 Bayesian Weibull Survival Model for SNLG data set with completely observed covariate values

```
----------------------------
RJAGS CODE (MCMC)
----------------------------
model
{
for (i in 1:636)
{
```

```
is.cen[i] ~ dinterval (t[i], t.cen[i])
t[i] ~dweib(alpha,lambda[i])
eta[i]<-beta0 +beta.age*age[i]+beta.hb*hb[i]+beta.wbc*wbc[i]
+beta.sex[sex[i]]+beta.albumin[albumin[i]]+beta.ap[ap[i]]+
beta.urea[urea[i]]+beta.extranod[extranod[i]]
+beta.bulk[bulk[i]]+beta.marrow[marrow[i]]+beta.bsy[bsy[i]]
+ beta.stage[stage[i]]+beta.ecog[ecog[i]]+beta.ldh[ldh[i]]
lambda[i]<-exp(eta[i])
}
delta.ecog[1]~dnorm(0,500)
delta.ecog[2]~dnorm(0,333.3)
delta.ecog[3]~dnorm(0,166.6)
delta.ecog[4]~dnorm(0,50)

beta.ecog[1]<- -delta.ecog[1] -delta.ecog[2] -delta.ecog[3]
-delta.ecog[4]
beta.ecog[2]<-  delta.ecog[1] -delta.ecog[2]  -delta.ecog[3]
 -delta.ecog[4]
beta.ecog[3]<-               2*delta.ecog[2]  -delta.ecog[3]
   -delta.ecog[4]
beta.ecog[4]<-              3*delta.ecog[3] -delta.ecog[4]
beta.ecog[5]<-                               4*delta.ecog[4]

delta.stage[1]~dnorm(0,333.3)
delta.stage[2]~dnorm(0,166.6)
delta.stage[3]~dnorm(0,50)

beta.stage[1]<- -delta.stage[1]-delta.stage[2] -delta.stage[3]
beta.stage[2]<-  delta.stage[1]-delta.stage[2] -delta.stage[3]
beta.stage[3]<-              2*  delta.stage[2]-delta.stage[3]
beta.stage[4]<-                                 3*delta.stage[3]

delta.ldh[1]~dnorm(0,500)
delta.ldh[2]~dnorm(0,166.6)
```

```
beta.ldh[1]<- -delta.ldh[1]   -delta.ldh[2]
beta.ldh[2]<-  delta.ldh[1]   -delta.ldh[2]
beta.ldh[3]<-                  2*delta.ldh[2]

delta.sex~dnorm(0.52,3.97)
delta.albumin~dnorm(0.97,17.86)
delta.ap~dnorm(0,33.41)
delta.urea~dnorm(0,33.41)
delta.extranod~dnorm(0.25,49.50)
delta.bulk~dnorm(0.34,6.78)
delta.marrow~dnorm(0.88,3.58)
delta.bsy~dnorm(0.26,3.95)

beta.sex[1]<-delta.sex
beta.sex[2]<- -delta.sex
beta.albumin[1]<-delta.albumin
beta.albumin[2]<- -delta.albumin
beta.ap[1]<-delta.ap
beta.ap[2]<- -delta.ap
beta.urea[1]<-delta.urea
beta.urea[2]<- -delta.urea
beta.extranod[1]<-delta.extranod
beta.extranod[2]<- -delta.extranod
beta.bulk[1]<-delta.bulk
beta.bulk[2]<- -delta.bulk
beta.marrow[1]<-delta.marrow
beta.marrow[2]<- -delta.marrow
beta.bsy[1]<-delta.bsy
beta.bsy[2]<- -delta.bsy
beta.age~dnorm(0.0295,47778.31)
beta.hb~dnorm(0.02,5102.04)
beta.wbc~dnorm(0.08,277.7)
alpha~dgamma(4,4)
beta0~dnorm(-1.5,6.25)
}
```

```
------------------------------------------------------------
INLA WEIBULL WITH ORTHOGONAL CONSTRAINT (FOR ALL COVARIATES)
------------------------------------------------------------
INLAorthogonal<-read.csv("INLAdatacomplete.csv",header=TRUE)

View(INLAorthogonal)

write.csv(INLAorthogonal,"INLAorthogonalwithldh.csv")

SNLGwei1=inla.surv(INLAorthogonal$time,INLAorthogonal$Event)
~age+hb1+wbc1+albumin+ap+bsy+bulk+extranod+gender+
marrow+urea+ldh1+ldh2+stage1+stage2+stage3+ecog1+ecog2
+ecog3+ecog4

SNLGmodelwei1=inla(SNLGwei1,family="weibullsurv",data=
INLAorthogonal,control.inla=list(strategy="laplace",
int.strategy="grid",dz=0.1),control.family=list(list(prior=
"loggamma",param=c(1,0.001))),control.fixed=list(mean=
list(age=0.0295,hb1=0.02,wbc1=0.08,gender=0.52,albumin
=0.97,ap=0,urea=0,extranod=0.25,bsy=0.26,bulk=0.34,marrow=0.88,
ldh1=0,ldh2=0,ecog1=0,ecog2=0,ecog3=0,ecog4=0,
stage1=0,stage2=0,stage3=0),mean.intercept=-1.5,
prec=list(age=47778.31,hb1=5102.04,wbc1=277.7,gender=
3.97,albumin=17.86,urea=33.41,extranod=49.50,bsy=3.95,
bulk=6.78,marrow=3.68,ap=33.41,ldh1=500,ldh2=166.6,
stage1=333.3,stage2=166.6,stage3=50,ecog1=500,ecog2=
333.3,ecog3=166.6,ecog4=50),prec.intercept=6.25),
control.compute=list(dic=TRUE,waic=TRUE,cpo=TRUE,mlik=TRUE),
control.predictor=list(link=1,compute=TRUE),verbose=TRUE)

summary(SNLGmodelwei1)

------------------------------
INLA HPD INTERVALS WEIBULL
------------------------------
```

```
Beta_intercept<-SNLGmodelwei1$marginals.fixed$`(Intercept)`
inla.hpdmarginal(0.95,Beta_intercept)
Beta_age<-SNLGmodelwei1$marginals.fixed$age
inla.hpdmarginal(0.95,Beta_age)
Beta_hb<-SNLGmodelwei1$marginals.fixed$hb1
inla.hpdmarginal(0.95,Beta_hb)
Beta_wbc1<-SNLGmodelwei1$marginals.fixed$wbc1
inla.hpdmarginal(0.95,Beta_wbc1)
Beta_albumin<-SNLGmodelwei1$marginals.fixed$albumin
inla.hpdmarginal(0.95,Beta_albumin)
Beta_ap<-SNLGmodelwei1$marginals.fixed$ap
inla.hpdmarginal(0.95,Beta_ap)
Beta_bsy<-SNLGmodelwei1$marginals.fixed$bsy
inla.hpdmarginal(0.95,Beta_bsy)
Beta_bulk<-SNLGmodelwei1$marginals.fixed$bulk
inla.hpdmarginal(0.95,Beta_bulk)
Beta_extranod<-SNLGmodelwei1$marginals.fixed$extranod
inla.hpdmarginal(0.95,Beta_extranod)
Beta_gender<-SNLGmodelwei1$marginals.fixed$gender
inla.hpdmarginal(0.95,Beta_gender)
Beta_marrow<-SNLGmodelwei1$marginals.fixed$marrow
inla.hpdmarginal(0.95,Beta_marrow)
Beta_urea<-SNLGmodelwei1$marginals.fixed$urea
inla.hpdmarginal(0.95,Beta_urea)
Beta_hyperpar<-SNLGmodelwei1$marginals.hyperpar$
`alpha parameter for weibullsurv`
inla.hpdmarginal(0.95,Beta_hyperpar)
Beta_ldh1<-SNLGmodelwei1$marginals.fixed$ldh1
inla.hpdmarginal(0.95,Beta_ldh1)
Beta_ldh2<-SNLGmodelwei1$marginals.fixed$ldh2
inla.hpdmarginal(0.95,Beta_ldh2)
Beta_ecog1<-SNLGmodelwei1$marginals.fixed$ecog1
inla.hpdmarginal(0.95,Beta_ecog1)
Beta_ecog2<-SNLGmodelwei1$marginals.fixed$ecog2
inla.hpdmarginal(0.95,Beta_ecog2)
```

```
Beta_ecog3<-SNLGmodelwei1$marginals.fixed$ecog3
inla.hpdmarginal(0.95,Beta_ecog3)
Beta_ecog4<-SNLGmodelwei1$marginals.fixed$ecog4
inla.hpdmarginal(0.95,Beta_ecog4)
Beta_stage1<-SNLGmodelwei1$marginals.fixed$stage1
inla.hpdmarginal(0.95,Beta_stage1)
Beta_stage2<-SNLGmodelwei1$marginals.fixed$stage2
inla.hpdmarginal(0.95,Beta_stage2)
Beta_stage3<-SNLGmodelwei1$marginals.fixed$stage3
inla.hpdmarginal(0.95,Beta_stage3)
```

A.6 R functions for splitting the survival time of SNLG subjects into 10 time intervals

```
#------------------------------------------------------------
#TIME INTERVAL CUT-POINTS (9 CUT-POINTS FOR 10 TIME-INTERVALS)
#------------------------------------------------------------

cuts=c(0.316,0.669,1.070,1.532,2.079,2.749,3.612,4.828,6.908)
dat<-read.table("snlgar.txt",header=TRUE)

#-------------------------------------------------------------------
#R FUNCTIONS FOR CONSTRUCTING CUT-POINTS FOR PIECEWISE CONSTANT
 HAZARD MODELS
#-------------------------------------------------------------------
piecewiseirfan<-function(data.in,cuts)
{n.cuts<-length(cuts)
n.patients<-length(data.in[,1])
patients<-1:n.patients
age<-data.in$age
albumin<-data.in$albumin
ap<-data.in$ap
```

```
bsy<-data.in$bsy
bulk<-data.in$bulk
ecog<-data.in$ecog
extranod<-data.in$extranod
hb<-data.in$hb
ldh<-data.in$ldh
marrow<-data.in$marrow
sex<-data.in$sex
stage<-data.in$stage
urea<-data.in$urea
wbc<-data.in$wbc
is.cen<-data.in[,1]
is.cen<-ifelse(is.na(is.cen),1,0)
t.cen<-data.in$t.cen
t<-data.in[,1]
t<-ifelse(is.na(t),t.cen,t)
data<-cbind(t,is.cen,patients)
out<-data[t<=cuts[1],]
alive<-data[t>cuts[1],]
n.alive<-length(alive[,1])
alive[,1]<-rep(cuts[1],n.alive)
alive[,2]<-rep(1,n.alive)
out<-rbind(out,alive)
perj<-rep(1,length(out[,1]))
out<-cbind(out,perj)
for (j in 2:n.cuts)
{data<-data[data[,1]>cuts[j-1],]
outj<-data[data[,1]<=cuts[j],]
outj[,1]<-outj[,1]-cuts[j-1]
alive<-data[data[,1]>cuts[j],]
n.alive<-length(alive[,1])
alive[,1]<-rep((cuts[j]-cuts[j-1]),n.alive)
alive[,2]<-rep(1,n.alive)
outj<-rbind(outj,alive)
perj<-rep(j,length(outj[,1]))
```

```
outj<-cbind(outj,perj)
out<-rbind(out,outj)
}
data<-data[data[,1]>cuts[n.cuts],]
data[,1]<-data[,1]-cuts[n.cuts]
perj<-rep((n.cuts+1),length(data[,1]))
data<-cbind(data,perj)
out<-rbind(out,data)
t<-out[,1]
is.cen<-out[,2]
is.na(t)<-is.cen==1
t.cen<-out[,1]
patients<-out[,3]
perj<-out[,4]
t.cen<-ifelse(is.cen==0,t.cen+1,t.cen)
out<-list(t=t,is.cen=is.cen,t.cen=t.cen,patients=patients,
perj=perj,age=age,albumin=albumin,ap=ap,bsy=bsy,bulk
=bulk,ecog=ecog,extranod=extranod,hb=hb,ldh=ldh,
marrow=marrow,sex=sex,stage=stage,urea=urea,wbc=wbc)
}

#------------------
#APPLYING THE CUTS
#------------------
dat2<-piecewiseirfan(dat,cuts)
```

A.7 Bayesian piecewise constant hazard model using hierarchical priors for baseline hazard for the SNLG data set (including missing covariate information)

```
#Bayesian piecewise constant hazard for the SNLG data set (complete
covariate information) using age, sex, albumin, extranod, bulk, marrow
and bsy covariates. Frailty term is also included.

modelstringfullhierarchical="
model
{
  for (i in 1:7672)
  {
  is.cen[i]~dinterval(t[i], t.cen[i])

  t[i]~dexp(lambda.whole[i])

  lambda.whole[i]<-exp(beta0[perj[i]]+beta.age[perj[i]]*(age[patients[i]])+
  beta.sex[perj[i],sex[patients[i]]]+beta.albumin[perj[i],albumin
  [patients[i]]+beta.extranod[perj[i],extranod[patients[i]]]+beta.bulk
  [perj[i],bulk[patienbts[i]]]+beta.marrow[perj[i],marrow[patients[i]]]+
  beta.bsy[perj[i],bsy[patients[i]]]+frail[patients[i]])
  }
for(c in 1:1396)
   {
frail[c]~dnorm(0,tau.frail)
tau.frail~dgamma(1.1,0.53)

#---------------------------
#Sampling missing data
#---------------------------
```

```
for(i in 1:1391)
   {
bsy[i]~dcat(phi.bsy[])
marrow[i]~dcat(phi.marrow[])
bulk[i]~dcat(phi.bulk[])
extranod[i]~dcat(phi.extranod[])
albumin[i]~dcat(phi.albumin[])
  }

phi.bsy0~dbeta(1,1)
phi.bsy[1]<-phi.bsy0
phi.bsy[2]<-1-phi.bsy0

phi.marrow0~dbeta(1,1)
phi.marrow[1]<-phi.marrow0
phi.marrow[2]<-1-phi.marrow0

phi.bulk0~dbeta(1,1)
phi.bulk[1]<-phi.bulk0
phi.bulk[2]<-1-phi.bulk0

phi.extranod0~dbeta(1,1)
phi.extranod[1]<-phi.extranod0
phi.extranod[2]<-1-phi.extranod0

phi.albumin0~dbeta(1,1)
phi.albumin[1]<-phi.albumin0
phi.albumin[2]<-1-phi.albumin0

#----------------------------------
#Priors for all beta coefficients
#----------------------------------

for (j in 1:10)
     {
```

```
beta0m[j]~dnorm(-4.807,35.05)
beta0[j]~dnorm(beta0m[j],666.04)

beta.age[j]~dnorm(0.64,71.42857143)

delta.sex[j]~dnorm(0.52, 3.97)
delta.sex[j,1]<-delta.sex[j]
delta.sex[j,2]<- -delta.sex[j]

delta.bsy[j]~dnorm(0.26, 3.95)
beta.bsy[j,1]<-delta.bsy[j]
beta.bsy[j,2]<- -delta.bsy[j]

delta.marrow[j]~dnorm(0.88,3.68)
beta.marrow[j,1]<-delta.marrow[j]
beta.marrow[j,2]<- -delta.marrow[j]

delta.bulk[j]~dnorm(0.34,6.78)
beta.bulk[j,1]<-delta.bulk[j]
beta.bulk[j,2]<- -delta.bulk[j]

delta.extranod[j]~dnorm(0.25,49.50)
beta.extranod[j,1]<-delta.extranod[j]
beta.extranod[j,2]<- -delta.extranod[j]

delta.albumin[j]~dnorm(0.97,17.86)
beta.albumin[j,1]<-delta.albumin[j]
beta.albumin[j,2]<- -delta.albumin[j]
  }
}
"
```

A.8 Bayesian piecewise constant hazard model with frailties using autoregressive priors for the SNLG data set with missing covariate information

```
#Bayesian piecewise constant hazard model for the SNLG
data set using all 14 covariates

#-------------------------------------------------------
#Initial values for piecewise constant hazard model
#-------------------------------------------------------

tinits1<-dat$t.cen+5
is.na(tinits1)<-dat2$is.cen==0
tinits2<-tinits1+2
inits1<-list(beta0=c(-6,-6,-6,-6,-6,-6,-6,-6,-6,-6),
                     delta.sex=rep(0.01,10),
                     beta.age=rep(0,10),
                     delta.stage=matrix(0.01,nrow=10,ncol=3),
                     delta.albumin=rep(0.01,10),
                     delta.ap=rep(0.01,10),
                     delta.urea=rep(0.01,10),
                     delta.bulk=rep(0.01,10),
                     delta.bsy=rep(0.01,10),
                     delta.extranod=rep(0.01,10),
                     delta.marrow=rep(0.01,10),
                     delta.ecog=matrix(0.01,nrow=10,ncol=4),
                     beta.hb=rep(0,10),
                     beta.wbc=rep(0,10),
                     delta.ldh=matrix(0.01,nrow=10,ncol=2),
                     phi.bsy0=0.5,phi.ap0=0.5,phi.albumin0=0.5,
                     phi.extranod0=0.5,phi.bulk0=0.5,
                     phi.marrow0=0.5,phi.urea0=0.5,phi.ecog
```

```
                    =c(0.2,0.2,0.2,0.2,0.2),phi.ldh=c(0.3,0.4,
                    0.3),t=tinits1,alpha.weibull=1.1)

inits2<-list(beta0=c(-7,-7,-7,-7,-7,-7,-7,-7,-7,-7),
                    delta.sex=rep(0.01,10),
                    beta.age=rep(0,10),
                    delta.stage=matrix(0.01,nrow=10,ncol=3),
                    delta.albumin=rep(0.01,10),
                    delta.ap=rep(0.01,10),
                    delta.urea=rep(0.01,10),
                    delta.bulk=rep(0.01,10),
                    delta.bsy=rep(0.01,10),
                    delta.extranod=rep(0.01,10),
                    delta.marrow=rep(0.01,10),
                    delta.ecog=matrix(0.01,nrow=10,ncol=4),
                    beta.hb=rep(0,10),
                    beta.wbc=rep(0,10),
                    delta.ldh=matrix(0.01,nrow=10,ncol=2),
                    phi.bsy0=0.5,phi.ap0=0.5,phi.albumin0=0.5,
                    phi.extranod0=0.5,phi.bulk0=0.5,
                    phi.marrow0=0.5,phi.urea0=0.5,
                    phi.ecog=c(0.2,0.2,0.2,0.2,0.2),
                    phi.ldh=c(0.3,0.4,0.3),t=tinits2,
                    alpha.weibull=0.9999)

                    pinits<-list(inits1,inits2)

#-----------------------------------------------------------
#RJAGS codes for piecewise constant hazard model
#-----------------------------------------------------------

modelstring="
model
{
for (i in 1:7672)
```

```
{
is.cen[i]~dinterval(t[i], t.cen[i])

t[i]~dexp(lambda.whole[i])

lambda.whole[i]<-exp(beta0[perj[i]]+beta.sex[perj[i],sex
[patients[i]]]+beta.age[perj[i]]*(age[patients[i]]-62.02157)
+beta.stage[perj[i],stage[patients[i]]]+beta.ecog[perj[i],
ecog[patients[i]]]+beta.albumin[perj[i],albumin[patients[i]]]
+beta.ap[perj[i],ap[patients[i]]]+beta.urea[perj[i],urea
[patients[i]]]+beta.hb[perj[i]]*(hb[patients[i]]-126.363)+
beta.wbc[perj[i]]*(wbc[patients[i]]-8.035688)+beta.extranod
[perj[i],extranod[patients[i]]]+beta.ldh[perj[i],ldh[patients[i]]]
+beta.bulk[perj[i],bulk[patients[i]]]+beta.marrow[perj[i],
marrow[patients[i]]]+beta.bsy[perj[i],bsy[patients[i]]]+
frail[patients[i]])
}

for(c in 1:1391)
{
frail[c]~dnorm(0,tau.frail)
}
tau.frail~dgamma(1.1,0.53)

#-----------------------
#Sampling missing data
#-----------------------

for(i in 1:1391)
{
bsy[i]~dcat(phi.bsy[])
ecog[i]~dcat(phi.ecog[])
ldh[i]~dcat(phi.ldh[])
ap[i]~dcat(phi.ap[])
urea[i]~dcat(phi.urea[])
```

```
hb[i]~dnorm(0,1)
marrow[i]~dcat(phi.marrow[])
albumin[i]~dcat(phi.albumin[])
wbc[i]~dnorm(0,1)
bulk[i]~dcat(phi.bulk[])
extranod[i]~dcat(phi.extranod[])
}

for (j in 1:5)
{
alpha.ecog[j]<-2
}

for(k in 1:3)
{
alpha.ldh[k]<-3
}

phi.ldh~ddirch(alpha.ldh)
phi.ecog~ddirch(alpha.ecog)

phi.bsy0~dbeta(1,1)
phi.bsy[1]<-phi.bsy0
phi.bsy[2]<-1-phi.bsy0

phi.ap0~dbeta(1,1)
phi.ap[1]<-phi.ap0
phi.ap[2]<-1-phi.ap0

phi.urea0~dbeta(1,1)
phi.urea[1]<-phi.urea0
phi.urea[2]<-1-phi.urea0

phi.marrow0~dbeta(1,1)
phi.marrow[1]<-phi.marrow0
```

```
phi.marrow[2]<-1-phi.marrow0

phi.extranod0~dbeta(1,1)
phi.extranod[1]<-phi.extranod0
phi.extranod[2]<-1-phi.extranod0

phi.bulk0~dbeta(1,1)
phi.bulk[1]<-phi.bulk0
phi.bulk[2]<-1-phi.bulk0

phi.albumin0~dbeta(1,1)
phi.albumin[1]<-phi.albumin0
phi.albumin[2]<-1-phi.albumin0

#----------------------------------
#Priors for all beta coefficients
#----------------------------------

delta.albumin[1]~dnorm(0.97,17.86)
delta.ap[1]~dnorm(0,33.41)
delta.sex[1]~dnorm(0.52,3.97)
delta.bsy[1]~dnorm(0.26,3.95)
delta.bulk[1]~dnorm(0.34,6.78)
delta.extranod[1]~dnorm(0.25,49.50)
delta.urea[1]~dnorm(0,33.41)
delta.marrow[1]~dnorm(0.88,3.68)

beta.sex[1,1]<- -delta.sex[1]
beta.sex[1,2]<- delta.sex[1]

beta.albumin[1,1]<- -delta.albumin[1]
beta.albumin[1,2]<- delta.albumin[1]

beta.ap[1,1]<- -delta.ap[1]
beta.ap[1,2]<- delta.ap[1]
```

```
beta.bsy[1,1]<- -delta.bsy[1]
beta.bsy[1,2]<- delta.bsy[1]

beta.bulk[1,1]<- -delta.bulk[1]
beta.bulk[1,2]<- delta.bulk[1]

beta.extranod[1,1]<- -delta.extranod[1]
beta.extranod[1,2]<- delta.extranod[1]

beta.marrow[1,1]<- -delta.marrow[1]
beta.marrow[1,2]<- delta.marrow[1]

beta.urea[1,1]<- -delta.urea[1]
beta.urea[1,2]<- delta.urea[1]

#-----------------------------------------
#Priors for continuous variables and intercept
#-----------------------------------------
beta0r~dnorm(-4.807, 66.60)

beta0[1]~dnorm(beta0r,66.60)

beta.age[1]~dnorm(0.0295,47778.31)

beta.hb[1]~dnorm(0.02,5102.4)

beta.wbc[1]~dnorm(0.08,277.7)

#-------------------------------------------------------------
#Categorical covariates for more than 2 levels (ecog,stage,ldh)
#-------------------------------------------------------------

delta.ecog[1,1]~dnorm(0,50)
delta.ecog[1,2]~dnorm(0,166.6)
```

```
delta.ecog[1,3]~dnorm(0,333.3)
delta.ecog[1,4]~dnorm(0,500)

beta.ecog[1,1]<--delta.ecog[1,1]-delta.ecog[1,2]-delta.ecog[1,3]
-delta.ecog[1,4]

beta.ecog[1,2]<-delta.ecog[1,1]-delta.ecog[1,2]-delta.ecog[1,3]
-delta.ecog[1,4]

beta.ecog[1,3]<-2*delta.ecog[1,2]-delta.ecog[1,3]
-delta.ecog[1,4]

beta.ecog[1,4]<-3*delta.ecog[1,3]-delta.ecog[1,4]

beta.ecog[1,5]<-4*delta.ecog[1,4]

delta.stage[1,1]~dnorm(0,50)
delta.stage[1,2]~dnorm(0,166.6)
delta.stage[1,3]~dnorm(0,333.3)

beta.stage[1,1]<- -delta.stage[1,1]-delta.stage[1,2]
-delta.stage[1,3]

beta.stage[1,2]<-delta.stage[1,1]-delta.stage[1,2]
-delta.stage[1,3]

beta.stage[1,3]<-2*delta.stage[1,2]-delta.stage[1,3]

beta.stage[1,4]<-3*delta.stage[1,3]

delta.ldh[1,1]~dnorm(0,166.6)
delta.ldh[1,2]~dnorm(0,500)

beta.ldh[1,1]<- -delta.ldh[1,1]   -delta.ldh[1,2]
beta.ldh[1,2]<-  delta.ldh[1,1]   -delta.ldh[1,2]
```

```
beta.ldh[1,3]<-                    2*delta.ldh[1,2]

#-------------------------------------------------
#DYNAMIC PART
#-------------------------------------------------

rho<-0.95
rf<-1-rho*rho

#-------------------------------------------------
#DYNAMIC PART A: INTERCEPT AND CONTINUOUS COVARIATES
#-------------------------------------------------

p0.e<-66.60/rf
p.hb.e<-5102.4/rf
p.wbc.e<-277.7/rf
p.age.e<-47778.31/rf

#-----------------------------------------------
# DYNAMIC PART B: 2-LEVEL CATEGORICAL COVARIATES
#-----------------------------------------------

p.sex.e<-3.97/rf
p.albumin.e<-17.86/rf
p.ap.e<-33.41/rf
p.urea.e<-33.41/rf
p.extranod.e<-49.50/rf
p.bulk.e<-6.78/rf
p.marrow.e<-3.68/rf
p.bsy.e<-3.95/rf

#-----------------------------------------------
#DYNAMIC PART C: 3-LEVEL CATEGORICAL COVARIATES
#-----------------------------------------------
```

```
p.ecog.e[1]<-50/rf
p.ecog.e[2]<-166.6/rf
p.ecog.e[3]<-333.3/rf
p.ecog.e[4]<-500/rf

p.stage.e[1]<-50/rf
p.stage.e[2]<-166.6/rf
p.stage.e[3]<-333.3/rf

p.ldh.e[1]<-166.6/rf
p.ldh.e[2]<-500/rf

#-------------------------------------------
# DYNAMIC PART D: ITERATIVE STEPS
#-------------------------------------------

for (j in 2:10)
{
beta0m[j]<- -1.5 + rho*(beta0[j-1]+1.5)
beta0[j]~dnorm(beta0m[j],p0.e)

beta.agem[j]<-0.04+rho*(beta.age[j-1]-0.04)
beta.age[j]~dnorm(beta.agem[j],p.age.e)

beta.hbm[j]<-0.02+rho*(beta.hb[j-1]-0.02)
beta.hb[j]~dnorm(beta.hbm[j],p.hb.e)

beta.wbcm[j]<-0.08 + rho*(beta.wbc[j-1]-0.08)
beta.wbc[j]~dnorm(beta.wbcm[j],p.wbc.e)

delta.sexm[j]<-rho*(delta.sex[j-1])
delta.sex[j]~dnorm(delta.sexm[j],p.sex.e)
beta.sex[j,1]<- -delta.sex[j]
beta.sex[j,2]<-  delta.sex[j]
```

```
delta.albuminm[j]<- rho*(delta.albumin[j-1])
delta.albumin[j]~dnorm(delta.albuminm[j],p.albumin.e)
beta.albumin[j,1]<- -delta.albumin[j]
beta.albumin[j,2]<-  delta.albumin[j]

delta.apm[j]<- rho*(delta.ap[j-1])
delta.ap[j]~dnorm(delta.apm[j],p.ap.e)
beta.ap[j,1]<- -delta.ap[j]
beta.ap[j,2]<-  delta.ap[j]

delta.bsym[j]<- rho*(delta.bsy[j-1])
delta.bsy[j]~dnorm(delta.bsym[j],p.bsy.e)
beta.bsy[j,1]<- -delta.bsy[j]
beta.bsy[j,2]<-  delta.bsy[j]

delta.bulkm[j]<- rho*(delta.bulk[j-1])
delta.bulk[j]~dnorm(delta.bulkm[j],p.bulk.e)
beta.bulk[j,1]<- -delta.bulk[j]
beta.bulk[j,2]<-  delta.bulk[j]

delta.extranodm[j]<- rho*(delta.extranod[j-1])
delta.extranod[j]~dnorm(delta.extranodm[j],p.extranod.e)
beta.extranod[j,1]<- -delta.extranod[j]
beta.extranod[j,2]<-  delta.extranod[j]

delta.uream[j]<- rho*(delta.urea[j-1])
delta.urea[j]~dnorm(delta.uream[j],p.urea.e)
beta.urea[j,1]<- -delta.urea[j]
beta.urea[j,2]<-  delta.urea[j]

delta.marrowm[j]<- rho*(delta.marrow[j-1])
delta.marrow[j]~dnorm(delta.marrowm[j],p.marrow.e)
beta.marrow[j,1]<- -delta.marrow[j]
beta.marrow[j,2]<-  delta.marrow[j]
```

```
delta.ecogm[j,1]<-rho*(delta.ecog[j-1,1])
delta.ecog[j,1]~dnorm(delta.ecogm[j,1],p.ecog.e[1])

delta.ecogm[j,2]<-rho*(delta.ecog[j-1,2])
delta.ecog[j,2]~dnorm(delta.ecogm[j,2],p.ecog.e[2])

delta.ecogm[j,3]<-rho*(delta.ecog[j-1,3])
delta.ecog[j,3]~dnorm(delta.ecogm[j,3],p.ecog.e[3])

delta.ecogm[j,4]<-rho*(delta.ecog[j-1,4])
delta.ecog[j,4]~dnorm(delta.ecogm[j,4],p.ecog.e[4])

beta.ecog[j,1]<--delta.ecog[j,1]-delta.ecog[j,2]-delta.ecog
[j,3]-delta.ecog[j,4]

beta.ecog[j,2]<-delta.ecog[j,1]-delta.ecog[j,2]-delta.ecog
[j,3]-delta.ecog[j,4]

beta.ecog[j,3]<-2*delta.ecog[j,2]-delta.ecog[j,3]-delta.ecog
[j,4]
beta.ecog[j,4]<-3*delta.ecog[j,3]-delta.ecog[j,4]
beta.ecog[j,5]<-4*delta.ecog[j,4]

delta.stagem[j,1]<-rho*(delta.stage[j-1,1])
delta.stage[j,1]~dnorm(delta.stagem[j,1],p.stage.e[1])

delta.stagem[j,2]<-rho*(delta.stage[j-1,2])
delta.stage[j,2]~dnorm(delta.stagem[j,2],p.stage.e[2])

delta.stagem[j,3]<-rho*(delta.stage[j-1,3])
delta.stage[j,3]~dnorm(delta.stagem[j,3],p.stage.e[3])

beta.stage[j,1]<--delta.stage[j,1]-delta.stage[j,2]-delta.
stage[j,3]
```

```
beta.stage[j,2]<-delta.stage[j,1]-delta.stage[j,2]-delta.
stage[j,3]

beta.stage[j,3]<-2*  delta.stage[j,2]-delta.stage[j,3]
beta.stage[j,4]<-3*delta.stage[j,3]

delta.ldhm[j,1]<-rho*(delta.ldh[j-1,1])
delta.ldh[j,1]~dnorm(delta.ldhm[j,1],p.ldh.e[1])

delta.ldhm[j,2]<-rho*(delta.ldh[j-1,2])
delta.ldh[j,2]~dnorm(delta.ldhm[j,2],p.ldh.e[2])

beta.ldh[j,1]<- -delta.ldh[j,1]   -delta.ldh[j,2]
beta.ldh[j,2]<-  delta.ldh[j,1]   -delta.ldh[j,2]
beta.ldh[j,3]<-                  2*delta.ldh[j,2]
}
}
"

#--------------------------------------
#Running JAGS model (100000 iterations)
#--------------------------------------

snlgjag<-jags.model(textConnection(modelstring),data=list
(t=dat2$t,t.cen=dat2$t.cen,is.cen=dat2$is.cen,perj=dat2$
perj,patients=dat2$patients,age=dat2$age,sex=dat2$sex,
albumin=dat2$albumin,ap=dat2$ap,bsy=dat2$bsy,bulk=dat2
$bulk,ecog=dat2$ecog,extranod=dat2$extranod,hb=dat2$hb,
ldh=dat2$ldh,marrow=dat2$marrow,stage=dat2$stage,urea=
dat2$urea,wbc=dat2$wbc),n.chains=2,inits=pinits)
update(snlgjag,n.iter=1000)
output<-coda.samples(model=snlgjag,variable.names=c("beta0",
"beta.sex","beta.age","beta.ldh","beta.stage","beta.ecog",
"beta.albumin","beta.ap","beta.urea","beta.hb","beta.wbc",
"beta.extranod","beta.marrow","beta.bulk","beta.bsy","alpha.
```

```
weibull"),n.iter=100000,thin=1)
summary(output)
autocorr.plot(output)
traceplot(output)
plot(output)
crosscorr(output)
```

A.9 An example of user-written Metropolis-Hastings sampler code for Weibull Survival Model

```
----------------------------------------
METROPOLIS HASTINGS CODE FOR WEIBULL MCMC
----------------------------------------

setwd("D:/Piston/Non Hodgkin Lymphoma/MH code")
mydata1<-read.csv("mydata4.csv",header=TRUE)
data=list(t=mydata1$t,died=mydata1$died,age=mydata1$age,
sex=mydata1$sex,albumin=mydata1$albumin,extranod=mydata1
$extranod,bulk=mydata1$bulk,marrow=mydata1$marrow,bsy=
mydata1$bsy,stage=mydata1$stage,ecog=mydata1$ecog,hb
=mydata1$hb1,wbc=mydata1$wbc1,urea=mydata1$urea,ldh=
mydata1$ldh,ap=mydata1$ap)
var<-list(var.betaprop=c(0.000009,0.000036,0.000625,0.0009,
0.0009,0.000625,0.000625,0.0009,0.000625,0.002025,
0.0000000225,0.0009,0.000625,0.000625,0.000625,0.0009,
0.0009,0.0009,0.0009,0.000625),alphaprop=1,a=10)
prior<-list(priormean=c(0.024,0.034,0.05,-0.112,0.203,0.091,
-0.064,0.036,0.039,-0.112,-0.005,0.091,0.091,-0.112,-0.112,
-0.112,-0.112,-0.112,-0.112,-0.112),
priorvar=c(0.00014,0.0003,0.006,0.003,0.004,0.002,0.006,
0.005,0.005,0.003,0.000007,0.002,0.002,0.003,0.003,0.003,
0.003,0.003,0.003,0.003),aalpha=9,balpha=9)
```

```
start<-list(beta=c(0.03,0.04,0.0,-0.05,0.1,0.07,-0.12,0.1,-0.03,
-0.2,-0.005,0.15,0.15,0.03,0.05,-0.05,-0.1,0.0,0.05,-0.05),
alpha=0.8)
weibullmcmc<-function(data,prior,start,niter,var,show=TRUE)
{
  t=data$t
  died=data$died
  sex=data$sex
  age=data$age
  albumin=data$albumin
  ap=data$ap
  urea=data$urea
  bulk=data$bulk
  bsy=data$bsy
  extranod=data$extranod
  marrow=data$marrow
  hb=data$hb
  wbc=data$wbc
  ldh=data$ldh
  ecog=data$ecog
  stage=data$stage
  n<-length(t)

  var.beta<-prior$priorvar

  mean.beta<-prior$priormean

  var.betaprop<-var$var.betaprop

  alphaprop=var$alphaprop
  a=var$a

  sdvar.betaprop=sqrt(var.betaprop)

  sd.alphaprop=sqrt(alphaprop)
```

```
beta<-start$beta
beta<-as.matrix(beta)
write(dim(beta),file="")

alpha=start$alpha

write(alpha,file="")
nD=sum(died)

k<-20
X<- matrix(nrow=n,ncol=k)
{
  X[,1]<-rep(1,n)
  X[,2]<-age
  X[,3]<-wbc
  X[,4]<-ifelse(sex==1,-1,1)
  X[,5]<-ifelse(albumin==1,-1,1)
  X[,6]<-ifelse(ap==1,-1,1)
  X[,7]<-ifelse(urea==1,-1,1)
  X[,8]<-ifelse(bulk==1,-1,1)
  X[,9]<-ifelse(bsy==1,-1,1)
  X[,10]<-ifelse(extranod==1,-1,1)
  X[,11]<-ifelse(marrow==1,-1,1)
  X[,12]<-hb
  X[,13]<-ifelse(ldh==1,-1,1)

  for (j in 1:4)
  {
    X[,13+j]<-ifelse(ecog>j,1,-1)
  }

  for (j in 1:3)
  {
    X[,17+j]<-ifelse(stage>j,1,-1)
```

```
    }
  }

result<-matrix(nrow=niter,ncol=21)
naccept<-0
for(i in 1:niter)
{
eta<-X%*%beta
eta<-eta[,1]
lambda<-exp(eta)
loglike<-sum(log(lambda[died==1]))+(alpha - 1)*sum(log(t
[died==1]))+(nD*log(alpha)) - sum(lambda*(t^alpha))
logprior<--1/2*sum(((beta-mean.beta)^2)/var.beta)
betaprop<-rnorm(20,beta,sdvar.betaprop)
alphaprop<-rgamma(1,a,a/alpha)
eta.prop<-X%*%betaprop
lambda.prop<-exp(eta.prop)
logprior.prop<--1/2*sum(((betaprop-mean.beta)^2)/var.beta)
loglike.prop<-sum(log(lambda.prop[died==1]))+(alphaprop - 1)*
sum(log(t[died==1]))+(nD*log(alphaprop))- sum(lambda.prop*t
^alphaprop)
logpost<-loglike+logprior
logpost.prop<-loglike.prop+logprior.prop
A<-exp(logpost.prop-logpost)
alphaden<- (alpha/alphaprop)^(2*a-1)*exp(-a*(alpha/alphaprop
-alphaprop/alpha))
aprob<-min(1,A*alphaden)

u=runif(1)
if(u<aprob)
{
beta<-betaprop
alpha<-alphaprop
naccept<-naccept+1
}
```

```
result[i,]<-c(beta,alpha)
}
if(show==TRUE)
message(paste("Acceptance=",naccept/niter))
return(result)
}
output=weibullmcmc(data,prior,start,100000,var,show=TRUE)
summary(output)
par(mfrow=c(2,2))
plot(acf(output[5000:100000,2]),main="beta.age")
plot(ts(output[5000:100000,2]),main="beta.age")
plot(density(output[5000:100000,2]),main="beta.age")
plot(acf(output[5000:100000,3]),main="beta.wbc")
plot(ts(output[5000:100000,3]),main="beta.wbc")
plot(density(output[5000:100000,3]),main="beta.wbc")
plot(acf(output[5000:100000,4]),main="beta.sex1")
plot(ts(output[5000:100000,4]),main="beta.sex1")
plot(density(output[5000:100000,4]),main="beta.sex1")
plot(acf(output[5000:100000,5]),main="beta.albumin")
plot(ts(output[5000:100000,5]),main="beta.albumin")
plot(density(output[5000:100000,5]),main="beta.albumin")
plot(acf(output[5000:100000,6]),main="beta.ap")
plot(ts(output[5000:100000,6]),main="beta.ap")
plot(density(output[5000:100000,6]),main="beta.ap")
plot(acf(output[5000:100000,7]),main="beta.urea")
plot(ts(output[5000:100000,7]),main="beta.urea")
plot(density(output[5000:100000,7]),main="beta.urea")
plot(acf(output[5000:100000,8]),main="beta.bulk")
plot(ts(output[5000:100000,8]),main="beta.bulk")
plot(density(output[5000:100000,8]),main="beta.bulk")
plot(acf(output[5000:100000,9]),main="beta.bsy")
plot(ts(output[5000:100000,9]),main="beta.bsy")
plot(density(output[5000:100000,9]),main="beta.bsy")
plot(acf(output[5000:100000,10]),main="beta.extranod")
plot(ts(output[5000:100000,10]),main="beta.extranod")
```

```
plot(density(output[5000:100000,10]),main="beta.extranod")
plot(acf(output[5000:100000,11]),main="beta.marrow")
plot(ts(output[5000:100000,11]),main="beta.marrow")
plot(density(output[5000:100000,11]),main="beta.marrow")
plot(acf(output[5000:100000,12]),main="beta.hb")
plot(ts(output[5000:100000,12]),main="beta.hb")
plot(density(output[5000:100000,12]),main="beta.hb")
plot(acf(output[5000:100000,13]),main="beta.ldh")
plot(ts(output[5000:100000,13]),main="beta.ldh")
plot(density(output[5000:100000,13]),main="beta.ld
```

A.10 User-written INLA code for intercept-only simple linear regression example

```
#-----------------------------------------------
#Create the data (iid normally distributed y_i)
#-----------------------------------------------
y <- c(1.2697,7.7637,2.2532,3.4557,4.1776,6.4320,-3.6623,
7.7567,5.9032,7.2671,-2.3447,8.0160,3.5013,2.8495,0.6467,
3.2371,5.8573,-3.3749,4.1507,4.3092,11.7327,2.6174,9.4942,
-2.7639,-1.5859,3.6986,2.4544,-0.3294,0.2329,5.2846)
n<-length(y)
ybar <- mean(y)

#-------------------------------------------------
#Define the parameters of the prior distributions
#-------------------------------------------------
mu0 <- -3
sigma2_0 <- 4
a<- 1.6
b <- 0.4
```

```
#------
#R-INLA
#------
library(INLA)
formula <- y ~ 1
inla.output <-inla(formula, data = data.frame(y = y),
control.family = list(hyper = list(prec =list(prior =
"loggamma", param = c(a, b)))),control.fixed = list(
mean.intercept = mu0,prec.intercept = 1/sigma2_0),
control.inla=list(strategy="gaussian",int.strategy=
"grid"))
summary(inla.output)
summary(inla.output$marginals.hyperpar[[1]])
inla.output.precision<-inla.output$marginals.
hyperpar[[1]]
inla.output.precision.matrix<-as.matrix(inla.
output.precision)
psi.grid.new<-inla.output.precision.matrix[,1]
theta.grid.new<-inla.output$marginals.fixed$
`(Intercept)`[,1]
lines(inla.output$marginals.fixed$"(Intercept)")

#-------------------------------------------------------------
#Select H grid points for the hyperparameter psi (precision)
#-------------------------------------------------------------

H <- 75
tau.grid <-psi.grid.new
hprior <- dgamma(tau.grid,shape=a,rate=b)
#(#prior for hyperparameter #tau (Precision))

#---------------------------------
#Compute quantities
#---------------------------------
```

```
theta.n<- sigma2.n <- lik <- num <- den <- prior <- c()
for (h in 1:H) {
theta.n[h] <-(tau.grid[h]*n*ybar+mu0/sigma2_0)/(tau.grid
[h]*n + 1/sigma2_0)
sigma2.n[h] <- 1 / (n*tau.grid[h] + 1/sigma2_0)
prior[h] <- dnorm(theta.n[h], mu0, sd=sqrt(sigma2_0))
lik[h] <- prod(dnorm(y, theta.n[h], sd=1/sqrt(tau.grid[h])))
num[h] <- hprior[h] * prior[h] * lik[h]
den[h] <- dnorm(theta.n[h], theta.n[h], sd=sqrt(sigma2.n[h]))
                }

#--------------------------------------
#Unnormalised marginal posterior for tau
#--------------------------------------

post.tau <- num/den

#-------------------------
#Normalise the tau density (precision)
#-------------------------
f.tau <- approxfun(tau.grid, post.tau,yleft=min(tau.grid), yright=max(tau.grid))
const <- integrate(f.tau, min(tau.grid), max(tau.grid))
post.tau<- post.tau/const$value
post.tau
tau.plot.new<-cbind(tau.grid,post.tau)
plot(tau.plot.new)

#-------------------------------------------
#Select J grid points for the parameter theta (intercept)
#-------------------------------------------
J <- 75
theta.grid<-theta.grid.new

#-----------------------------------------------
#Full conditional distributions theta | psi,y
```

```
#-----------------------------------------------

full.cond.theta <- matrix(NA,J,H)
for (j in 1:J) {
for (h in 1:H) {
full.cond.theta[j,h] <- dnorm(theta.grid[j], theta.n[h],
sd=sqrt(sigma2.n[h]))
}
}

#-----------------------------------------
#Weighted joint posterior for theta and psi
#-----------------------------------------
Delta <- 1/sum(post.tau)
joint.post.theta.tau <- matrix(NA,J,H)
for (h in 1:H)
{joint.post.theta.tau[,h] <-full.cond.theta[,h]*post.tau[h]*
Delta
}

#-------------------------------------------------------------
#Integrate out psi to obtain the marginal posterior of theta
#-------------------------------------------------------------
marg.post.theta <- rowSums(joint.post.theta.tau)

#---------------------------------
#Normalizing the density for theta
#---------------------------------
f.theta <- approxfun(theta.grid,marg.post.theta,yleft=
min(theta.grid),yright=max(theta.grid))
const <- integrate(f.theta,min(theta.grid),max(theta.grid))
marg.post.theta <- marg.post.theta/const$value
marg.post.theta.plot<-as.matrix(marg.post.theta)
marg.post.theta.plot
marg.post.theta.plot.final<-cbind(theta.grid,
```

```
marg.post.theta.plot)
marg.post.theta.plot.final
plot(marg.post.theta.plot.final,pch=17,col="red",type="o")
library(Hmisc)

#-----------
#PLOT RESULTS
#-----------

#INTERCEPT

plot(inla.output$marginals.fixed$`(Intercept)`,pch=1,
col="blue",cex=2,type="o",lwd=2,main=expression
(paste("Posterior of",phantom(x),theta,phantom(x),
"obtained from user-written code vs R-INLA")),xlab=
expression(paste(phantom(x),mu,phantom(x))))
par(new=TRUE)
plot(marg.post.theta.plot.final,pch=17,col="red",
type="o",lty=3,lwd=4,axes=FALSE,ylab="",xlab="")
legend("topright",c("R-INLA","Own code"),col=
c("blue","red"),lty=c(1,3),lwd=c(2,4),bty="n")

#PRECISION

plot(inla.output$marginals.hyperpar[[1]],pch=1,col=
"green",cex=2,type="o",lwd=2,main=expression
(paste("Posterior of",phantom(x),tau,phantom(x),
"obtained from user-written code vs R-INLA")),
xlab=expression(paste(phantom(x),tau,phantom(x))))
par(new=TRUE)
plot(tau.plot.new,pch=17,col="purple",type="o",lty=3,
lwd=4,ylab="",xlab="",axes=FALSE)
legend("topright",c("R-INLA","Own code"),col=c("green",
"purple"),lty=c(1,3),lwd=c(2,4),bty="n")
```

```
#---------------------
#SUMMARY R-INLA OUTPUT
#---------------------
summary(inla.output$marginals.fixed$`(Intercept)`)
summary(inla.output$marginals.hyperpar$`Precision
for the Gaussian observations`)

#-----------------------
#SUMMARY INLA (OWN-CODE)
#-----------------------
#POSTERIOR SAMPLES FOR INTERCEPT
summary(marg.post.theta.plot.final)
#POSTERIOR SAMPLES FOR PRECISION
summary(tau.plot.new)
```

References

Aaldriks, A. A., E. J. Giltay, J. W. R. Nortier, L. G. M. van der Geest, B. C. Tanis, P. Ypma, S. le Cessie, and E. Maartense (2015). Prognostic significance of geriatric assessment in combination with laboratory parameters in elderly patients with aggressive non-Hodgkin lymphoma. *Leukemia & Lymphoma*, 56(4), pp. 927–935.

Aalen, O. O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, 3(3), pp. 227–243.

Aalen, O. O. and E. Husebeye (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine*, 10(8), pp. 1227–1240.

Ades, P. E., G. Lu, and J. P. T. Higgins (2005). The interpretation of random-effects meta-analysis in decision models. *Medical Decision Making*, 25(6), pp. 646–654.

Advani, R. H., H. Chen, T. M. Habermann, V. A. Morrison, E. A. Weller, R. I. Fisher, B. A. Peterson, et al. (2010). Comparison of conventional prognostic indices in patients older than 60 years with diffuse large B-cell lymphoma treated with R-CHOP in the US Intergroup Study (ECOG 4494, CALGB 9793): consideration of age greater than 70 years in an elderly prognostic index (E-IPI). *British Journal of Haematology*, 151(2), pp. 143–151.

Aitkin, M. and D. Clayton (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 29(2), pp. 156–163.

Alvaro-Meca, A., R. Akerkar, M. Alvarez-Bartolome, R. Gil-Prieto, H. Rue, and A. G. de Miguel (2013). Factors involved in health-related transitions after curative resection for pancreatic cancer. 10-years experience: a multi state model. *Cancer Epidemiology*, 37(1), pp. 91–96.

Armero, C., S. Cabras, M. E. Castellanos, S. Perra, A. Quiros, M. J. Oruezabal, and J. Sanchez-Rubio (2016). Bayesian analysis of a disability model for lung cancer survival. *Statistical methods in Medical Research*, 25(1), pp. 336–351.

Ashby, D. (2006). Bayesian statistics in medicine: a 25 year review. *Statistics in Medicine*, 25(21), pp. 3589–3631.

Azzalini, A. and A. Capitano (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 61(3), pp. 579–602.

Banerjee, S., M. M. Wall, and B. P. Carlin (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(1), pp. 123–142.

Berkey, C. S., D. C. Hoaglin, A. Antczak-Buckoms, F. Mosteller, and G. A. Colditz (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, 17(2), pp. 2537–2550.

Besag, J. and C. Kooperberg (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4), pp. 733–746.

Blangiardo, M. and M. Cameletti (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA* (1st ed.). Chichester, West Sussex, UK: Wiley.

Blangiardo, M., M. Cameletti, G. Baio, and H. Rue (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 4, pp. 33–49.

Bollen, C. W., C. S. Uiterwaal, and A. J. van Vught (2007). Meta-regression analysis of high-frequency ventilation vs conventional ventilation in infant respiratory distress syndrome. *Intensive Care Medicine*, 33(4), pp. 680–688.

Bormann, I. (2012). Digitizeit (version 2.0). Braunschweig, Germany. https://www.digitizeit.de.

Box, G. E. P. and K. Wilson (1951). On the experimental attainment of optimum condition. *Journal of the Royal Statistical Society Series B (Methodological)*, 13(1), pp. 1–45.

Bujkiewicz, S., J. R. Thompson, A. J. Sutton, N. J. Cooper, M. J. Harrison, D. P. M. Symmons, and K. R. Abrams (2013). Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Statistics in Medicine*, 32(22), pp. 3926–3943.

Burden, R. L., D. J. Faires, and A. M. Burden (2015). *Numerical Analysis* (10th ed.). Boston, MA, USA: Cengage Learning.

Carbone, P. P., H. S. Kaplan, K. Mushoff, D. W. Smithers, and M. Tubiana (1971). Report of the committee on Hodgkin's disease staging classification. *Cancer Research*, 31(11), pp. 1860–1861.

Carella, A. M., C. A. de Souza, S. Luminari, L. Marcheselli, A. Chiappella, A. di Rocco, M. Cesaretti, et al. (2013). Prognostic role of gender in diffuse large B-cell lymphoma treated with rituximab containing regimens: a Fondazione Italia Linfomi / Grupo de Estudos em Molestias Onco-Hematologicas retrospective study. *Leukemia and Lymphoma*, 54(1), pp. 53–57.

Carson, K. R., N. L. Bartlett, J. R. McDonald, S. Luo, A. Zeringue, J. Liu, Q. Fu, S. H. Chang, and G. A. Golditz (2012). Increased body mass index is associated with improved survival in United States veterans with diffuse large B-cell lymphoma. *Blood*, 30(26), pp. 3217–3222.

Carson, S. and J. M. Flemming (2014). Seal encounters at sea: A contemporary spatial approach using R-INLA. *Ecological Modelling*, 291(10), pp. 175–181.

Castillo, J. J., R. R. Ingham, J. L. Reagan, M. Furman, S. Dalia, and J. Mitri (2014). Obesity is associated with increased relative risk of diffuse large B-cell lymphoma: a meta-analysis of observational studies. *Clinical lymphoma, myeloma & leukemia*, 14(2), pp. 122–130.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction* (6th ed.). Boca-Raton, Florida, USA: Chapman & Hall (CRC).

Chen, L. P., S. J. Lin, and M. S. Yun (2012). Prognostic value of platelet count in diffuse large B-cell lymphoma. *Clinical Lymphoma, Myeloma & Leukemia*, 12(1), pp. 32–37.

Christina, T., S. Dalia, C. M. Bello, L. Sokol, E. M. Sotomayor, B. D. Shah, and J. C. Chavez (2013). The revised international prognostic index (R-IPI) score is predictive of survival in aggressive transformed lymphomas (TL) from low grade non-Hodgkin lymphomas in the chemoimmunotherapy era. *Blood*, 122(21), pp. 4275.

Chung, R., R. Lai, P. Wei, J. Lee, J. Hanson, A. R. Belch, A. R. Turner, and T. Reiman (2007). Concordant but not discordant bone marrow involvement in diffuse large B-cell lymphoma predicts a poor clinical outcome independent of the International Prognostic Index. *Blood*, 110(4), pp. 1278–1282.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1), pp. 141–151.

Collett, D. (2015). *Modelling Survival Data in Medical Research* (3rd ed.). Boca-Raton, Florida, USA: CRC Press.

Colvert, R. E. and T. J. Boardman (1976). Estimation in the piece-wise constant hazard rate model. *Communications in Statistics-Theory and Methods*, 5(11), pp. 1013–1029.

Consul, J. I. (2016). *Flexible Bayesian modeling of covariate effects on survival*. Ph.D. thesis, School of Mathematics and Statistics, Newcastle University.

Cooper, N. J., J. Peters, M. C. Lai, P. Juni, S. Wandel, S. Palmer, M. Paulden, et al. (2011). How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value in Health*, 14(2), pp. 371–380.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society Series B (Methodological)*, 34(2), pp. 187–220.

Cox, M. C., I. Nofroni, L. Ruco, R. Amodeo, A. Ferrari, G. La Verde, P. Cardelli, et al. (2008). Low absolute lymphocyte count is a poor prognostic factor in diffuse large B-cell lymphoma. *Leukemia & Lymphoma*, 49(9), pp. 1745–1751.

Cseke, B. (2011). Approximate marginals in latent Gaussian models. *Journal of Machine Learning Research*, 12, pp. 417–454.

Dalia, S., J. Chavez, B. Little, C. Bello, K. Fisher, J. H. Lee, P. Chervenick, L. Sokol, E. Sotomayor, and B. Shah (2014). Serum albumin retains independent sigificance in diffuse large B-cell lymphoma in the post-rituximab era. *Annals of Hematology*, 93(8), pp. 1305–1312.

Davies, O. L. and P. L. Goldsmith (1972). *Statistical Methods in Research and Production* (4th ed.). Edinburg, UK: Oliver and Boyd.

Dellaportas, P. and G. O. Roberts (2003). An introduction to MCMC. In J. Møller (Ed.), *Spatial Statistics and Computational Methods*, Chapter 1, pp. 15. New York: Springer.

Deshpande, J. V., M. Mukhopadhyay, and U. V. Naik-Nimbalkar (1999). Testing of two sample proportional intensity assumption for non-homogeneous Poisson processes. *Journal of Statistical Planning and Inference*, 81(2), pp. 237–251.

Dias, S., A. J. Sutton, and N. J. Welton (2013). Evidence synthesis for decision making 3: Heterogeneity - subgroups, meta-regression, bias and bias-adjustment. *Medical Decision Making*, 33(5), pp. 618–640.

Dobson, A. J. and A. G. Barnett (2008). *An Introduction to Generalized Linear Models* (3rd ed.). Boca Raton, Florida, USA: Chapman and Hall.

Du, G., Y. Yang, Y. Zhang, T. Sun, W. Liu, Y. Wang, J. Li, and H. Zhang (2013). Thrombocytosis and immunohistochemical expression of connexin 43 at diagnosis predict survival in advanced non-small-cell lung cancer treated with cisplatin-based chemotherapy. *Cancer Chemotherapy and Pharmacology*, 71(4), pp. 893–904.

Eidsvik, J., S. Martino, and H. Rue (2009). Approximate Bayesian inference in spatial generalized linear mixed models. *Scandinavian Journal of Statistics*, 36(1), pp. 1–22.

Eisenhauer, E. A., P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, et al. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2), pp. 228–247.

Engelhardt, M. and L. J. Bain (1992). Statistical analysis of a Weibull process with left-censored data. In J. P. Klein and P. K. Goel (Eds.), *Survival Analysis: State of the Art. Nato Science (Series E: Applied Sciences)*, Volume 211, pp. 173–196. Dordrecht, Netherlands: Springer.

Esteve, J., E. Benhamou, M. Croasdale, and L. Raymond (1990). Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine*, 9(5), pp. 529–538.

Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (2nd ed.). Boca Raton, Florida, USA: Chapman and Hall.

Farrow, M. (2003). Practical building of subjective covariance structures for large complicated systems. *Journal of the Royal Statistical Society Series D (The Statistician)*, 52(4), pp. 553–573.

Fiala, O., M. Pesek, J. Finek, J. Racek, M. Minarik, L. Benesova, Z. Bortlicek, O. Sorejs, R. Kucera, and O. Topolcan (2016). Serum albumin is a strong predictor of survival in patients with advanced-stage non-small cell lung cancer treated with erlotinib. *Neoplasma*, 63(3), pp. 471–476.

Fiocco, M., T. Stijnen, and H. Putter (2012). Meta-analysis of time-to-event outcomes using a hazard-based approach: Comparison with other models, robustness and meta-regression. *Computational Statistics and Data Analysis*, 56(5), pp. 1028–1037.

Flowers, C. R., P. J. Shenoy, U. Borate, K. Bumpers, T. Douglas-Holland, N. King, O. W. Brawley, et al. (2013). Examining racial differences in diffuse large B-cell lymphoma presentation and survival. *Leukemia & Lymphoma*, 54(2), pp. 268–276.

Foucher, L., M. Giral, J. P. Soulillou, and J. P. Daures (2007). A semi-Markov model for multistate and interval-censored data with multiple terminal events. application in renal transplantation. *Statistics in Medicine*, 26(30), pp. 5381–5393.

Frederiksen, B. L., S. O. Dalton, M. Osler, M. Steeding-Jessen, and P. de Nully Brown (2012). Socioeconomic position, treatment, and survival of non-Hodgkin lymphoma in Denmark - a nationwide study. *British Journal of Cancer*, 106(5), pp. 988–995.

Fung, C. (May 2017). Bayesian analysis of survival data and the piecewise constant hazard model. Master's thesis, School of Mathematics and Statistics, Newcastle University, UK.

Gamerman, D. (1991). Dynamic Bayesian models for survival data. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 40(1), pp. 63–79.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). Boca-Raton, Florida, USA: Chapman & Hall.

Gamerman, D. and M. West (1987). A time series application of dynamic survival models in unemployment studies. *Journal of the Royal Statistical Society Series D (The Statistician)*, 36(2/3), pp. 269–274.

Gao, S., G. Zhu, Y. Lin, X. Fan, P. Qian, J. Zhu, and Y. Yu (2002). Tumor necrosis factor-308 polymorphism with the risk and prognosis of non-hodgkin lymphomas: a meta-analysis study. *OncoTargets and therapy*, 9, pp. 1657–1670.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), pp. 398–409.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), pp. 721–741.

Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4), pp. 473–511.

Geyer, C. J. (2011). Introduction to Markov Chain Monte Carlo. In S. Brooks, A. Gelman, G. L. Jones, and X. L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, Chapter 1, pp. 14–16. Boca Raton, Florida: Chapman Hall / CRC.

Gomez-Rubio, V. and H. Rue (2018). Markov Chain Monte Carlo with the integrated nested Laplace approximation. *Statistics and Computing*, 28(5), pp. 1033–1051.

Greenwood, M. and G. U. Yule (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of The Royal Statistical Society*, 83(2), pp. 255–279.

Grimmett, G. and D. Stirzaker (2001). *Probability And Random Processes* (3rd ed.). Oxford, UK: Oxford University Press.

Guo, X. and B. P. Carlin (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1), pp. 1–9.

Guyot, P., A. E. Aedes, M. J. N. M. Ouwens, and N. J. Welton (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology*, 12, pp. 9.

Hand, D. J., F. Daly, K. McConway, D. Lunn, and E. Ostrowski (1994). *A Handbook of Small Data Sets* (1st ed.). London, UK: Chapman & Hall Statistics Text.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), pp. 97–109.

Hayward, R. L., R. C. F. Leonard, R. J. Prescott, the members of the Scotland, and Newcastle Lymphoma Group Therapy Working Party (1991). A critical analysis of prognostic factors for survival in intermediate and high grade non-Hodgkin's lymphoma. *British Journal of Cancer*, 63(6), pp. 945–952.

He, J., D. L. McGee, and X. Niu (2010). Application of the Bayesian dynamic survival model in medicine. *Statistics in Medicine*, 29(3), pp. 347–360.

Heidelberger, P. and P. D. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6), pp. 1109–1144.

Henderson, R. and P. Oman (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 61(2), pp. 367–379.

Hespanhol, V., H. Queiroga, A. Magalhaes, A. R. Santos, M. Coelho, and A. Marques (1995). Survival predictors in advanced non-small cell lung cancer. *Lung Cancer*, 13(3), pp. 253–267.

Hirakawa, T., H. Yamaguchi, N. Yokosei, S. Gomi, K. Inokuchi, and K. Dan (2010). Importance of maintaining the relative dose intenstiy of CHOP-like regimens combined with rituximab in patients with diffuse large B-cell lymphoma. *Annals of Hematology*, 89(9), pp. 897–904.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics*, 32(3), pp. 587–597.

Holford, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, 36(2), pp. 299–305.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, 1(3), pp. 255–273.

Hougaard, P. (2013). Frailty models. In J. P. Klein, H. C. van Houwelingen, J. G. Ibrahim, and T. H. Scheike (Eds.), *Handbook of Survival Analysis*, Chapter 22, pp. 407–488. Oxford: Oxford University Press.

Hsu, C. L., K. Y. Chen, J. Y. Shih, C. C. Ho, C. H. Yang, C. J. Yu, and P. C. Yang (2012). Advanced non-small cell lung cancer in patients aged 45 years or younger: outcomes and prognostic factors. *BMC Cancer*, 13(12), pp. 241.

Ibrahim, J. and M. H. Chen (2000). Power prior distributions for regression models. *Statistical Science*, 15(1), pp. 46–60.

Ibrahim, J. G., M. H. Chen, and D. Sinha (2001). *Bayesian Survival Analysis*. Springer.

Jiang, H., P. E. Brown, H. Rue, and S. Shimakura (2014). Geostatistical survival models for environmental risk assessment with large retrospective cohorts. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 177(3), pp. 679–695.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society Series B (Methodological)*, 40(2), pp. 214–221.

Kalbfleisch, J. D. and R. Prentice (1973). Marginal likelihood based on Cox's regression and life model. *Biometrika*, 60(2), pp. 267–278.

Kim, K. H., T. Y. Park, J. Y. Lee, S. M. Lee, J. J. Yim, C. G. Yoo, Y. W. Kim, S. K. Han, and S. C. Yang (2014). Prognostic significance of initial platelet counts and fibrinogen level in advanced non-small cell lung cancer. *Journal of Korean Medical Science*, 29(4), pp. 507–511.

Kim, S., M. H. Chen, K. Dey, and D. Gamerman (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, 13(470), pp. 17–35.

Laird, N. and D. Olivier (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of American Statistical Association*, 76(374), pp. 231–240.

Leung, E. Y., H. R. Scott, and D. C. McMillan (2012). Clinical utility of the pretreatment Glasgow prognostic score in patients with advanced inoperable non-small cell lung cancer. *Journal of Thoracic Oncology*, 7(4), pp. 655–662.

Li, X., Z. Bie, Z. Zhang, Y. Li, X. Hu, W. Liu, S. Zhang, G. Cheng, and B. Ai (2015). Clinical analysis of 64 patients with lung-cancer-associated hypercalcemia. *Journal of Cancer Research and Therapeutics*, 11(suppl), pp. C275–279.

Li, X., Z. Liu, J. Cao, X. Hong, W. Jianmin, F. Chen, C. Wang, S. Zou, J. Li, and Z. Shen (2012). Rituximab in combination with CHOP chemotherapy for the treatment of diffuse large B-cell lymphoma in China: a 10-year retrospective follow-up analysis of 437 cases from Shanghai Lymphoma Research Group. *Annals of Hematology*, 91(6), pp. 837–845.

Lim, J. U., C. D. Yeo, H. S. Kang, C. K. Park, J. S. Kim, J. W. Kim, S. J. Kim, and S. H. Lee (2018). Prognostic value of platelet count and lymphocyte to monocyte ratio combination in stage IV non-small cell lung cancer with malignant pleural effusion. *PLoS One*, 13(7), pp. e0200341.

Lindqvist, B. H., E. Elvebakk, and K. Heggland (2003). The trend-renewal process for statistical analysis of repairable systems. *Technometrics*, 45(1), pp. 31–44.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, New Jersey, USA: Wiley.

Longmore, M., I. Wilkinson, A. Baldwin, and E. Wallin (2014). *Oxford Handbook of Clinical Medicine* (9th ed.). Oxford, UK: Oxford University Press.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4), pp. 325–337.

Maartense, E., H. C. Kluin-Nelemans, S. le Cessie, P. M. Kluin, S. Snijder, and E. M. Noordijk (2000). Different age limits for elderly patients with indolent and aggressive non-Hodgkin Lymphoma and the role of relative survival with increasing age. *Cancer*, 89(12), pp. 2667–2676.

Maeda, T., H. Ueoka, M. Tabata, K. Kiura, T. Shibayama, K. Gemba, N. Takigawa, A. Hiraki, H. Katayama, and M. Harada (2000). Prognostic factors in advanced non-small cell lung cancer: elevated serum levels of neuron specific enolase indicate poor prognosis. *Japanese Journal of Clinical Oncology*, 30(12), pp. 534–541.

Mandrekar, S. J., S. E. Schild, S. L. Hilman, K. L. Allen, R. S. Marks, J. A. Mailliard, and J. E. K. others (2006). A prognostic model for advanced stage non-small cell lung cancer. *Cancer*, 107(14), pp. 781–792.

Marsden, A. T. H. (1958). The geographical pathology of cancer in Malaya. *British Journal of Cancer*, 12(2), pp. 161–176.

Martino, S., R. Akerkar, and H. Rue (2011). Approximate Bayesian inference for survival models. *Scandinavian Journal of Statistics*, 38(3), pp. 514–528.

Martins, T. G. and H. Rue (2014). Extending integrated nested Laplace approximation to a class of near-Gaussian latent models. *Scandinavian Journal of Statistics*, 41(4), pp. 893–912.

Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2013). Bayesian computing with INLA: new features. *Computational Statistics and Data Analysis*, 67, pp. 68–83.

Matakidou, A., T. Eisen, and R. S. Houlston (2005). Systematic review of the relationship between family history and lung cancer risk. *British Journal of Cancer*, 93(7), pp. 825–833.

Matsunuma, R., Y. Tanbo, N. Asai, Y. Ohkuni, S. Watanabe, S. Murakami, Y. Kawaura, and K. Kasahara (2014). Prognostic factors in patients with terminal stage lung cancer. *Journal of Palliative Medicine*, 17(2), pp. 189–194.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed.). Boca Raton, Florida, USA: Chapman and Hall.

Meira-Machado, L., J. de Uña-Alvarez, C. Cadarso-Suàrez, and P. K. Andersen (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, 18(2), pp. 195–222.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), pp. 1087–1092.

Møller, M. B., N. T. Pedersen, and B. E. Christensen (2003). Factors predicting long-term survival in low-risk diffuse large B-cell lymphoma. *American Journal of Hematology*, 74(2), pp. 94–98.

Montgomery, D. C. (2017). *Design and Analysis of Experiments* (9th ed.). Hoboken, NJ, USA: Wiley.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 64(2), pp. 231–252.

Nahman, N. S., F. Middendorf, D. Bay, W. H. McElligott, S. Powell, and J. Anderson (1992). Modification of the percutaneous approach to peritoneal dialysis catheter placement under peritoneoscopic visualization: clinical results in 78 patients. *Journal of the American Society of Nephrology*, 3(1), pp. 103–107.

Ngo, L., S. W. Hee, L. C. Lim, M. Tao, R. Quek, S. P. Yap, E. L. Loong, et al. (2008). Prognostic factors in patients with diffuse large B-cell lymphoma: before and after the introduction of rituximab. *Leukemia & Lymphoma*, 49(3), pp. 462–469.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow (2006). *Uncertain Judgements: Elicitng Expert's Probabilities* (1st ed.). Hoboken, New Jersey, USA: Wiley-Sons Limited.

Ohshima, K., Y. Mukai, H. Shiraki, T. Suzumiya, K. Tashiro, and M. Kikuchi (1997). Clonal integration and expression of human T-cell lymphotropic virus type I in carriers detected by polymerase chain reaction and inverse PCR. *American Journal of Hematology*, 54(4), pp. 306–312.

Oken, M. M., R. H. Creech, D. C. Tormey, J. Horton, T. E. McFadden, and P. P. Carbone (1982). Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American Journal of Clinical Oncology*, 5(6), pp. 649–655.

Oki, Y., K. Yamamoto, H. Kato, Y. Kuwatsuka, H. Taji, Y. Kagami, and Y. Morishima (2008). Low absolute lymphocyte count is a poor prognostic marker in patients with diffuse large B-cell lymphoma and suggests patients' survival benefit from rituximab. *European Journal of Haematology*, 81(6), pp. 448–453.

O'Mahony, S., S. Nathan, R. Mohajer, P. Bonomi, M. Batus, M. J. Fidler, K. Wells, N. Kern, S. Sims, and D. Amin (2016). Survival prediction in ambulatory patients with stage III/IV non-small cell lung cancer using the palliative performance scale, ECOG, and lung cancer symptom scale. *The American Journal of Hospice and Palliative Care*, 33(4), pp. 374–380.

Pao, W., V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, B. Singh, et al. (2004). EGF receptor gene mutations are common in lung cancers from never smokers and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A.*, 101(36), pp. 13306–11.

Papoila, A. L., A. Riebler, A. Amaral-Turkman, R. Sao-Jao, C. Ribeiro, C. Geraldes, and A. Miranda (2014). Stomach cancer incidence in southern Portugal: A spatio-temporal analysis. *Biometrics Journal*, 56(3), pp. 403–415.

Park, M. J., J. Lee, J. Y. Hong, M. K. Choi, J. H. Yi, S. J. Lee, S. J. Oh, J. S. Ahn, K. Park, and M. J. Ahn (2009). Prognostic model to predict outcomes in nonsmall cell lung cancer patients treated with gefitinib as a salvage treatment. *Cancer*, 115(7), pp. 1518–1530.

Peyrade, F., F. Jordin, C. Thieblemont, A. Thyss, J. F. Emile, S. Castaigne, B. Coiffier, et al. (2011). Attenuated immunochemotherapy regimen (R-miniCHOP) in elderly patients older than 80 years with diffuse large B-cell lymphoma: a multicentre, single-arm phase 2 trial. *Lancet Oncology*, 12(5), pp. 460–468.

Pfreundschuh, M., A. D. Ho, E. Cavallin-Stahl, M. Wolf, R. Pettengell, I. Vasovia, A. Belch, et al. (2008). Prognostic significance of maximum tumour (bulk) diameter in young patients with good-prognosis diffuse large B-cell lymphoma treated with CHOP-like chemotherapy with or without rituximab: an exploratory analysis of the MabThera International Trial Group (MInT) study. *Lancet Oncology*, 9(5), pp. 435–444.

Plummer, M. (2016). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-6, https://CRAN.R-project.org/package=rjags.

Prairie, R. R. and B. Ostle (1961). An analysis of some relay failure data from a composite exponential population. *Technometrics*, 3(3), pp. 423–428.

Proctor, S. J. and P. R. A. Taylor (2000). A practical guide to continuous population-based data collection (PACE): a process facilitating uniformity of care and research into practice. *QJM*, 93(2), pp. 67–73.

Putter, H., M. Fiocco, and R. B. Geskus (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26(11), pp. 2389–2430.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R package version 2.0.0, http://www.R-project.org/.

Raftery, A. E. and S. Lewis (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 763–773. Oxford University Press.

Rami-Portas, R., H. Asamura, W. D. Travis, and V. W. Rusch (2017). Lung. In M. B. Amin (Ed.), *AJCC Cancer Staging Manual* (8 ed.)., Chapter 36, pp. 431–456. Switzerland: Springer.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In O. Bousquet, U. von Luxburg, and G. Ratsch (Eds.), *Advanced Lectures on Machine Learning*, Chapter 4, pp. 64. Berlin: Springer.

Ren, S. and J. E. Oakley (2014). Assurance calculations for planning clinical trials with time-to-event outcomes. *Statistics in Medicine*, 33(1), pp. 31–45.

Ren, W., L. Cunha, and C. V. Deutsch (2005). Preservation of multiple point structure when conditioning by kriging. In O. Leuangthong and C. Deutsch (Eds.), *Geostatistics Banff 2004*, pp. 643–645. Boca Raton, Florida: Springer.

Rizvi, M. A., A. M. Evens, M. S. Tallman, and S. T. Rosen (2006). T-cell non-Hodgkin lymphoma. *Blood*, 107(4), pp. 1255–1264.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7(1), pp. 110–120.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4), pp. 351–367.

Royston, P. and M. K. B. Parmar (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics In Medicine*, 21(15), pp. 2175–2179.

Rue, H. (2001). Fast sampling of Gaussian Markov random field. *Journal of Royal Statistical Society Series B (Statistical Methodology)*, 63(part 2), pp. 325–338.

Rue, H. and L. Held (2005). *Gaussian Markov Random Field: Theory and Applications* (1st ed.). Boca-Raton, Florida, USA: Chapman & Hall (CRC).

Rue, H. and S. Martino (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference*, 137(10), pp. 3177–3192.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 71(2), pp. 319–392.

Rue, H., A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, 4, pp. 395–421.

Rue, H., I. Steinsland, and S. Erland (2004). Approximating hidden Gaussian Markov random fields. *Journal of Royal Statistical Society Series B (Statistical Methodology)*, 66(part 4), pp. 877–892.

Schad, F., A. Thronicke, M. L. Steele, A. Merkle, B. Matthes, C. Grah, and H. Matthes (2018). Overall survival of stage IV non-small cell lung cancer patients treated with Viscum album L. in addition to chemotherapy, a real-world observational multicenter analysis. *PLoS One*, 13(8), pp. e0203058.

Schardt, C., M. B. Adams, T. Owens, S. Keitz, and P. Fontelo (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, 7(16), pp. 1–6.

Schwartz, L. H., S. Lètiere, E. de Vries, R. Ford, S. Gwyther, S. Mandrekar, L. Shankar, et al. (2016). RECIST 1.1 – update and clarification: From the RECIST committee. *European Journal of Cancer*, 62(2), pp. 132–137.

Scott, H. R., D. C.Mcmillan, L. M. Forrest, D. J. Brown, C. S. McArdle, and R. Milroy (2002). The systemic inflammatory response, weight loss, performance status and survival in patients with inoperable non-small cell lung cancer. *British Journal of Cancer*, 87(3), pp. 264–267.

Shaked, M. (1979). An estimator for the generalized hazard rate function. *Communications in Statistics-Theory and Methods*, 8(1), pp. 17–33.

Shankland, K. R., J. O. Armitage, and B. W. Hancock (2012). Non-Hodgkin lymphoma. *Lancet*, 380(9844), pp. 848–857.

Sherlock, C. (2013). Optimal scaling of the random walk metropolis: general criteria for the 0.234 acceptance rule. *Journal of Applied Probability*, 50(1), pp. 1–15.

Sherlock, C. and G. Roberts (2009). Optimal scaling of the random walk metropolis on elliptically symmetric unimodal targets. *Bernoulli*, 15(3), pp. 774–798.

Siegel, R., D. Naishadham, and A. Jemal (2013). Cancer statistics, 2013. *CA Cancer J Clin.*, 63(1), pp. 11–30.

Simmons, C. P., F. Koinis, M. T. Fallon, K. C. Fearon, J. Bowden, T. S. Solheim, B. H. Gronberg, D. C. McMillan, I. Gioulbasanis, and B. J. Laird (2015). Prognosis in advanced lung cancer-a prospective study examining key clinicopathological factors. *Lung Cancer*, 88(3), pp. 304–309.

Simonato, L., A. Agudo, W. Ahrens, E. Benhamou, S. Benhamou, P. Bofetta, P. Brennan, et al. (2001). Lung cancer and cigarette smoking in Europe: an update of risk estimates and an assessment of inter-country heterogeneity. *International Journal of Cancer*, 91(6), pp. 876–887.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 88(423), pp. 979–983.

Slymen, D. J., T. P. Miller, S. M. Lippman, C. M. Spier, D. P. Kerrigan, J. A. Rybaski, C. S. Rangel, L. C. Richter, and T. M. Grogan (1990). Immunobiologic factors predictive of clinical outcome in diffuse large-cell lymphoma. *Journal of Clinical Oncology*, 8(6), pp. 3589–3631.

Smith, A. F. M., A. M. Skene, J. E. H. Shaw, and J. C. Naylor (1987). Progress with numerical and graphical methods for practical Bayesian statistics. *Journal of the Royal Statistical Society Series D (The Statistician)*, 36(2/3), pp. 75–82.

Song, F., T. A. Sheldon, A. J. Sutton, K. R. Abrams, and D. R. Jones (2001). Methods for exploring heterogeneity in meta-analysis. *Evaluation & the Health Professions*, 24(2), pp. 126–151.

Song, M. K., J. S. Chung, O. S. Yong, G. W. Lee, S. G. Kim, Y. M. Seol, H. J. Shin, Y. J. Choi, D. H. Shin, and E. Y. Yun (2010). Clinical impact of bulky mass in the patient with primary extranodal diffuse large B-cell lymphoma treated with R-CHOP therapy. *Annals of Hematology*, 89(2), pp. 985–991.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 64(4), pp. 583–639.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2014). The deviance information criterion: 12 years on. *Journal of Royal Statistical Society Series B (Statistical Methodology)*, 76(3), pp. 485–493.

Svaton, M., O. Fiola, M. Pesek, F. Bruha, P. Mukensnabl, J. Racek, M. Minarik, and Z. Bortlicek (2014). Predictive and prognostic significance of sodium levels in patients with NSCLC treated by erlotinib. *Anticancer Research*, 34(12), pp. 7461–7465.

Takahashi, H., N. Chung, M. Yokoyama, S. Tsunoda, T. Yano, K. Murayama, C. Hashimoto, K. Tamura, K. Sato, and Y. Ishigatsubo (2012). Prognostic impact of extranodal involvement in diffuse large B-cell lymphoma in the rituximab era. *Cancer*, 118(17), pp. 4166–4172.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), pp. 528–540.

Teramukai, S., T. Kitano, Y. Kishida, M. Kawahara, K. Kubota, K. Komuta, K. Minato, et al. (2009). Pretreatment neutrophil count as an independent prognostic factor in advanced non-small-cell lung cancer: an analysis of Japan Multinational Trial Organisation LC00-03. *European Journal of Cancer*, 45(11), pp. 1950–1958.

Thompson, S. G. and J. P. T. Higgins (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), pp. 1559–1573.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22(4), pp. 1701–1762.

Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393), pp. 82–86.

Troppan, K., A. Deutsch, A. Gerger, T. Stojakovic, C. Beham-Schmid, K. Wenzl, P. Neumeister, and M. Pichler (2014). The derived neutrophil to lymphocyte ratio is an independent prognostic factor in patients with diffuse large B-cell lymphoma. *British Journal of Cancer*, 110(2), pp. 369–374.

Ulas, A., F. P. Turkoz, K. Silay, S. Tukluoglu, N. Avci, B. Oksuzoglu, and N. Alkis (2014). A laboratory prognostic index model for patients with advanced non-small cell lung cancer. *PLoS One*, 9(12), pp. e114471.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3), pp. 439–454.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Behavioural and Educational Statistics*, 30(3), pp. 261–293.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), pp. 1–48.

Viechtbauer, W. (2017). *metafor: Meta-Analysis Package for R.* R package version 2.0.0, https://CRAN.R-project.org/package=metafor.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics ASME*, 18(3), pp. 293–297.

West, M. (May, 1982). *Aspects of Recursive Bayesian Estimation.* Ph.D. thesis, University of Nottingham.

West, M. and M. Berliner (1992). Modelling time-varying hazards and covariate effects. In J. P. Klein and P. K. Goel (Eds.), *Survival Analysis: State of the Art. Nato Science (Series E: Applied Sciences)*, Volume 211, pp. 47–62. Dordrecht, Netherlands: Springer.

West, M., P. J. Harrison, and H. S. Migon (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80(389), pp. 73–83.

Wienke, A. (2011). *Frailty Models in Survival Analysis* (1st ed.). Boca-Raton, Florida, USA: Chapman & Hall, CRC Biostatistics Series.

Wikle, C. K. and S. H. Holan (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (discussion). *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 71(2), pp. 384.

Wilson, K. J. and M. Farrow (2010). Bayes linear kinematics in the analysis of failure rates and failure time distributions. *Journal of Risk and Reliability*, 224(4), pp. 309–321.

Wilson, K. J. and M. Farrow (2017). Bayes linear kinematics in a dynamic Bayesian survival model. *International Journal of Approximate Reasoning*, 80, pp. 239–256.

Yan, Y., W. C. Chan, D. D. Weisenburger, J. R. Anderson, M. A. Bast, J. M. Vose, P. J. Bierman, and J. O. Armitage (1995). Clinical and prognostic significance of bone marrow involvement in patients with diffuse aggressive B-cell lymphoma. *Journal of Clinical Oncology*, 13(6), pp. 1336–1342.

Ye, W., Y. Yang, J. Wang, Z. Kadziola, N. Rajan, and S. Qin (2014). Prognostic factors for patients with advanced non-small cell lung cancer treated with gemcitabine-platinum as first-line therapy in an observational setting in China. *Thoracic Cancer*, 5(4), pp. 319–324.

Zhang, X. and Y. Ran (2015). Prognostic role of elevated platelet count in patients with lung cancer: a systematic review and meta-analysis. *International Journal of Clinical and Experimental Medicine*, 8(4), pp. 5379–5387.

Zhao, X. (June, 2010). *Bayesian survival analysis for prognostic index development with many covariates and missing data*. Ph.D. thesis, University of Newcastle Upon Tyne.

Zhao, Z., L. H. Sehn, A. W. Rademaker, L. I. Gordon, A. S. Lacase, A. Crosby-Thompson, and A. Vanderplas (2014). An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood*, 123(26), pp. 837–842.

Ziepert, M., D. Hasenclever, E. Kuhnt, B. Glass, N. Schmitz, M. Pfreundschuh, and M. Loeffler (2010). Standard International Prognostic Index remains a valid predictor of outcome for patients with aggressive $CD20^+$ B-Cell Lymphoma in the Rituximab era. *Journal of Clinical Oncology*, 28(14), pp. 2373–2380.