Bayesian approaches of Markov models embedded in unbalanced panel data

by
Christoffel Joseph Brand Muller

Dissertation presented for the degree of
Doctor of Philosophy
in the Faculty of Economic and Management Sciences
at Stellenbosch University

Promoter: Prof Paul J Mostert
Faculty of Economic and Management Sciences

December 2012

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: ..... Date: 5 November 2012

Summary

Multi-state models are used in this dissertation to model panel data, also known as longitudinal or cross-sectional time-series data. These are data sets which include units that are observed across two or more points in time. These models have been used extensively in medical studies where the disease states of patients are recorded over time.

A theoretical overview of the current multi-state Markov models when applied to panel data is presented and based on this theory, a simulation procedure is developed to generate panel data sets for given Markov models. Through the use of this procedure a simulation study is undertaken to investigate the properties of the standard likelihood approach when fitting Markov models and then to assess its shortcomings. One of the main shortcomings highlighted by the simulation study, is the unstable estimates obtained by the standard likelihood models, especially when fitted to small data sets.

A Bayesian approach is introduced to develop multi-state models that can overcome these unstable estimates by incorporating prior knowledge into the modelling process. Two Bayesian techniques are developed and presented, and their properties are assessed through the use of extensive simulation studies.

Firstly, Bayesian multi-state models are developed by specifying prior distributions for the transition rates, constructing a likelihood using standard Markov theory and then obtaining the posterior distributions of the transition rates. A selected few priors are used in these models. Secondly, Bayesian multi-state imputation techniques are presented that make use of suitable prior information to impute missing observations in the panel data sets. Once imputed, standard likelihood-based Markov models are fitted to the imputed data sets to estimate the transition rates. Two different Bayesian imputation techniques are presented. The first approach makes use of the Dirichlet distribution and imputes the unknown states at all time points with missing observations. The second approach uses a Dirichlet process to estimate the time at which a transition occurred between two known observations and then a state is imputed at that estimated transition time.

The simulation studies show that these Bayesian methods resulted in more stable results, even when small samples are available.

Opsomming

Meerstadium-modelle word in hierdie verhandeling gebruik om paneeldata, ook bekend as longitudinale of deursnee tydreeksdata, te modelleer. Hierdie is datastelle wat eenhede insluit wat oor twee of meer punte in tyd waargeneem word. Hierdie tipe modelle word dikwels in mediese studies gebruik indien verskillende stadiums van 'n siekte oor tyd waargeneem word. 'n Teoretiese oorsig van die huidige meerstadium Markov-modelle toegepas op paneeldata word gegee. Gebaseer op hierdie teorie word 'n simulasieprosedure ontwikkel om paneeldatastelle te simuleer vir gegewe Markov-modelle. Hierdie prosedure word dan gebruik in 'n simulasie-studie om die eienskappe van die standaard aanneemlikheidsbenadering tot die pas van Markov modelle te ondersoek en dan enige tekortkominge hieruit te beoordeel. Een van die hoof tekortkominge wat uitgewys word deur die simulasiestudie, is die onstabiele beramings wat verkry word indien dit gepas word op veral klein datastelle.

'n Bayes-benadering tot die modellering van meerstadiumpaneeldata word ontwikkel om hierdie onstabiliteit te oorkom deur a priori-inligting in die modelleringsproses te inkorporeer. Twee Bayes-tegnieke word ontwikkel en aangebied, en hulle eienskappe word ondersoek deur 'n omvattende simulasiestudie.

Eerstens word Bayes-meerstadium-modelle ontwikkel deur a priori-verdelings vir die oorgangs-koerse te spesifiseer en dan die aanneemlikheidsfunksie te konstrueer deur van standaard Markov-teorie gebruik te maak en die a posteriori-verdelings van die oorgangskoerse te bepaal. 'n Gekose aantal a priori-verdelings word gebruik in hierdie modelle. Tweedens word Bayes-meerstadium invul tegnieke voorgestel wat gebruik maak van a priori-inligting om ontbrekende waardes in die paneeldatastelle in te vul of te imputeer. Nadat die waardes ge-imputeer is, word standaard Markov-modelle gepas op die ge-imputeerde datastel om die oorgangskoerse te beraam. Twee verskillende Bayes-meerstadium imputasie tegnieke word bespreek. Die eerste tegniek maak gebruik van 'n Dirichletverdeling om die ontbrekende stadium te imputeer by alle tydspunte met 'n ontbrekende waarneming. Die tweede benadering gebruik 'n Dirichlet-proses om die oorgangstyd tussen twee waarnemings te beraam en dan die ontbrekende stadium te imputeer op daardie beraamde oorgangstyd.

Die simulasiestudies toon dat die Bayes-metodes resultate oplewer wat meer stabiel is, selfs wanneer klein datastelle beskikbaar is.

Acknowledgments

I wish to express my gratitude to:

- Bianca, my chemical sister inspiration. Her dedication, her commitment and her friendship carried me through the difficult times, inspired me through the time I was stumped and enriched my life through the lonely times.
- My parents and family for their support, interest and encouragement.
- Prof Paul Mostert, for his guidance, support and enthusiasm throughout this study.
- Prof Willie Conradie for his friendship and guidance in forming my academic career.
- The van der Merwe's for providing me with a bit of normality when weird Greek letters and R code was all I had for weeks on end.

In addition my sincere thanks go out to all the persons not mentioned above who helped me and supported my efforts towards completion of this study.

Soli Deo Gloria

Table of Contents

Abbreviations

AH-F	Aguirre-Hernández and Farewell
B-MSM	Bayesian multi-state model
B-MSI	Bayesian multi-state imputation
DIC	Deviance information criterion
DP	Dirichlet process
HPD	Highest posterior density
LP-Jef	A B-MSM fitted using the limiting probabilities in the likelihood using the Jeffreys prior as prior for the limiting probabilities.
LP-MDI	A B-MSM fitted using the limiting probabilities in the likelihood using the MDI prior as prior for the limiting probabilities.
MCMC	Markov chain Monte Carlo
MDI	Maximal data information
MedSE	Median square error
M-H	Metropolis-Hastings
MLE	Maximum likelihood estimate
MSE	Mean square error
MSM	Multi-state model
PMP	Probability matching priors
TP-Cov	A B-MSM fitted using the transition probabilities in the likelihood with covariates included in the model.
TP-NoCov	A B-MSM fitted using the transition probabilities in the likelihood with no covariates in the model.

Notation

$p_{ij}(t^*, t, F_{t^*})$	The transition probability of a multi-state process, i.e., the probability that a process currently in state i at time t^*, will be in state j at time t, given the history F_{t^*}.
$p_{ij}(t)$	The transition probability of a time homogeneous Markov process.
$\lambda_{ij}(t, F_t)$	The transition rate of a multi-state process, i.e., the instantaneous hazard rate of progressing from state i to state j at time t, given the history F_t.
P_j	The limiting probability, $\lim_{t\to\infty} p_{ij}(t)$, that a Markov chain will be in state j.
$Q(\boldsymbol{\lambda})$	Transition intensity matrix of a multi-state process.
$P(t)$	Transition probability matrix of a multi-state process at time t.
$S(t_{il})$	The state occupied by individual i at time point l.
$L(\boldsymbol{\lambda}\|data)$	The likelihood function.
$\pi(\boldsymbol{\lambda})$	The prior distribution of the parameters $\boldsymbol{\lambda}$.
$f(x\|\boldsymbol{\lambda})$	The density function of x, given the parameters $\boldsymbol{\lambda}$.
$f(x, \boldsymbol{\lambda})$	The joint density function of x and $\boldsymbol{\lambda}$.
$\pi(\boldsymbol{\lambda}\|data)$	The posterior distribution of the parameters $\boldsymbol{\lambda}$, given the data.
$m(data)$	The normalising constant of the posterior distribution.
$Lo(\lambda, \delta_i)$	The loss function for decision δ_i given λ is the true parameter value.
$\boldsymbol{\lambda}_{-i}$	The vector $\boldsymbol{\lambda}$, with the i-th element removed.
$multi(n, \boldsymbol{\pi})$	Multinomial distribution with parameter n and probability vector $\boldsymbol{\pi}$.
$\exp(\lambda_0)$	Exponential distribution, $f(\lambda) = \lambda_0 e^{-\lambda_0 \lambda}$.

$dis\ unif(1,4)$	Uniform distribution (discrete) with possible values from 1 to 4.
$unif(0,1)$	Uniform distribution (continuous) between 0 and 1.
$MVN(\boldsymbol{\mu}, \Sigma)$	Multi-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix Σ.
$nor(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2.
$Dirichlet(\alpha_1, ..., \alpha_k)$	Dirichlet distribution with parameter vector $(\alpha_1, ..., \alpha_k)'$.
$DP(\gamma, F)$	Dirichlet process with base function F and weight parameter γ.
$T_{rs}(t)$	The monotonic transition function that captures the likelihood of being in state $S(t) = s$ after time t given state $S(0) = r$ at time 0.
$\mathcal{W}(\kappa, \alpha)$	Weibull distribution with scale parameter, κ, and shape parameter, α.

1

Introduction

In this chapter, the background behind the research is presented with a brief overview of the scope and major contributions of this study. The chapter concludes with a brief outline of the dissertation.

1.1 Background and description of problem

The initial starting point of the research was a paediatric HIV panel data set presented while doing statistical consultation. Clinicians and researchers at the Faculty of Medicine and Health Sciences of Stellenbosch University, South Africa had panel data for infants born with human immunodeficiency virus (HIV) that needed statistical analysis and understanding.

Panel data, also known as longitudinal or cross-sectional time-series data, are data sets which include units that are observed across two or more time points (Hsiao, 2003, p. 1). An extract from a panel data set is shown in Table 1.1. The observations per patient are the panels under study, and the variable of interest is the state each patient is in at each visit ($S(t_i)$ denotes the state observed at visit i and time point t_i.). This type of data is referred to as multi-state panel data.

Table 1.1: An example of the structure of panel data.

Patient	**Treatment**	$\mathbf{t}_1$	$\mathbf{S(t}_1\mathbf{)}$	$\mathbf{t}_2$	$\mathbf{S(t}_2\mathbf{)}$	$\mathbf{t}_3$	$\mathbf{S(t}_3\mathbf{)}$	...	...	$\mathbf{t}_9$	$\mathbf{S(t}_9\mathbf{)}$	$\mathbf{t}_{10}$	$\mathbf{S(t}_{10}\mathbf{)}$
1	1	1	1	2	1								
2	0	2	1	4	1	6	3						
14	1	1	2	10	1	12	2	⋯	⋯	24	3	26	3
20	1	5	3	6	3	7	2	⋯	⋯	14	3		

Table 1.1 can be used to highlight some of the complexities and difficulties associated with panel data:

- The data is often unbalanced. Here patient 1 has two observations, patient 2 has three, patient 14 has ten and patient 20 has nine observations.
- The unbalanced design introduces a missingness problem into the analysis. This needs to be taken into account when modelling these types of data.
- The fact that each panel has its own observational times, means that, although a large number of patients may be included in the study, only a limited number of observations are usable at each distinct time point.

These complexities and difficulties, although not unique to panel data, combine to give researchers unique challenges that need to be overcome in the modelling process.

1.2 Contribution of this dissertation

This dissertation seeks to develop models that can be used to incorporate prior expertise into multi-state models. As such, the following contributions to the field of multi-state models are made in this dissertation:

1) It gives an overview of the current Markov process approach used to model multi-state data, and a simulation method, based on the Markov process, is developed to generate panel data sets for given Markov models and data scenarios (see Section 2.1).
2) An extensive simulation study is undertaken to assess the maximum likelihood method of fitting Markov models to panel data (see Section 2.2). The simulation study highlights the shortcomings of the Markov models when they are fitted to small data sets or when complex models with covariates are fitted. By simulating and investigating different underlying models, data scenarios and covariate scenarios; it is shown at which point the Markov process becomes unstable and unable to provide suitable parameter estimates (see Section 2.3.3).
3) Bayesian multi-state models are developed where prior information is directly incorporated into the multi-state models. Two different Bayesian models are developed, one where the likelihood is constructed using the limiting probabilities of the Markov process (see Section 4.1), and the other where the transition rates are incorporated into the likelihood through the transition probabilities (see Section 4.2). Through an extensive simulation study it

is shown that even for very small data sets these models provide good estimates for the parameters of the multi-state model. These estimates are comparable, and at times better, than those found through the frequentist or classical use of the Markov model (see Section 4.3).

4) Bayesian multi-state imputation techniques are developed. These techniques make use of prior information to impute missing observations in the panel data sets. Two different imputation techniques are developed, one where all missing observations are imputed (see Section 5.1), and the other where only the transition point between two known observations is imputed (see Section 5.2). A second extensive simulation study shows that, as with the Bayesian multi-state models, the Bayesian imputation techniques provide good estimates for the parameters of the multi-state model (see Section 5.4).

5) The Bayesian imputation methods developed are fitted to published multi-state data sets (see Sections 4.4 and 5.5), and the results of the Bayesian models compare to those of the standard frequentist Markov models.

1.3 Outline of the dissertation

In Chapter 2, multi-state models are introduced as the models of choice when analysing panel data. Firstly, the Markov process theory behind the models, the maximum likelihood estimation procedure, the different types of multi-state models and how multi-state models can be interpreted from a survival analysis point of view, are presented. Secondly, model assessment tools that are untilised to assess multi-state models are discussed and explained. The chapter concludes with the presentation of a simulation process that can generate panel data sets for known/given transition rates. The simulation process is used to investigate the properties of the Markov model when fitted to multi-state data under varying transition matrices and data size scenarios. This simulation process is used extensively in the remainder of the dissertation to simulate panel data sets that are used to assess the proposed Bayesian models of Chapters 4 and 5.

In Chapter 3, an overview is given of the underlying Bayesian principles and techniques that are used in Chapters 4 and 5 for the Bayesian multi-state models. Firstly, the various non-informative priors (the MDI, the Jeffreys and the probability matching priors), the conjugate

priors, and subjective priors are introduced and briefly discussed. Secondly, methods used to summarise and to assess the posterior distribution, such as, posterior intervals and decision making and model fit criteria, are presented and discussed. Thirdly, the two most often used MCMC (Markov chain Monte Carlo) simulation techniques utilised to sample variates from intractable posterior distributions, the Gibbs sampler and the Metropolis-Hastings algorithm, are presented. The chapter concludes with two Bayesian methods that are often used in survival analysis and will be used for the Bayesian multi-state imputation techniques of Chapter 5: the Dirichlet process prior and the Gibbs sampler within the Dirichlet process.

In Chapter 4, two Bayesian multi-state models are developed and assessed. Firstly, a Bayesian multi-state model is introduced where the likelihood is expressed in terms of the limiting probabilities of a Markov process. Prior distributions, namely the MDI and the Jeffreys priors, are assumed for the limiting probabilities and a Metropolis-Hastings algorithm is used to sample variates from the posterior distribution. Secondly, a Bayesian multi-state model is presented where the transition rates are directly modelled in the likelihood and subjective priors are placed on the transition rates in the likelihood. The exponential distribution is used as a prior distribution for transition rates. A model is also developed that allows for the incorporation of covariates into the multi-state model. The chapter concludes with a simulation study based on a variety of generated data sets developed at the end of Chapter 2. Various models and data scenarios are used and Bayesian multi-state models are fitted to the panel data sets. The results are compared to the known population parameters used to generate the data sets.

In Chapter 5, two Bayesian multi-state imputation techniques are developed and assessed. Prior information is not directly used in the modelling process, but rather used to impute the missing observations in the data set. Once the data has been imputed/augmented using the prior information, a multi-state model is fitted to the imputed data set. The posterior distribution is generated by repeatedly imputing time points of transition states and fitting a multi-state model to each imputed data set. Firstly, a model is presented that uses prior probability vectors obtained from experts to impute all missing observations in the data set. A multinomial distribution is used to sample the unknown observations with underlying probabilities from a Dirichlet distribution. Secondly, a Dirichlet process is used to estimate the

unknown transition time point between two known observations. Prior information about the transition process is incorporated in the Dirichlet process by means of prior transition functions that govern the imputation process. Three different transition functions are discussed. The chapter concludes with a simulation study where the proposed Bayesian multi-state imputation techniques are used in conjunction with the data generation procedure of Chapter 2. The effect of different models, data scenarios and prior assumptions are investigated and discussed.

The dissertation concludes with a final chapter that presents the major results of the research and suggests areas of future research.

2

Multi-State Modelling

In many medical studies, especially when the interest is on the progression of a disease, longitudinal data is analysed and utilised to study the patterns of the disease. Disease progression can be quantified by a disease state, or stage, at given time points or at given time intervals, together with important and relevant markers of the disease that are measured at each of these time points or intervals. These disease markers are sometimes subjectively interpreted by clinicians and in some cases they are defined and interpreted by organisations such as the World Health Organisation (WHO), as in the case of HIV/AIDS staging.

Modelling the relationship between time and disease state is an important aspect of the study of disease progression. Patients occupy different states at observed discrete time points, which can be seen as *panel data* (Kalbfleisch and Lawless, 1985) or can also generally be referred to as interval-censored data.

In this chapter, multi-state modelling will be introduced as the model of choice when analysing panel data. The outline of this chapter is as follows:

- In Section 2.1, the Markov process theory behind multi-state models is introduced and discussed. The maximum likelihood method to estimate the parameters of the model, how covariates are incorporated into the Markov model, different types of multi-state models, estimation problems that may arise when fitting multi-state models and how multi-state models can be interpreted from a survival analysis point of view are all discussed.
- In Section 2.2, the criteria to assess the assumptions, the fit and the effect of covariates in multi-state models are discussed.
- In Section 2.3, a procedure is developed that can be used to simulate multi-state data sets for given parameters. A simulation study is undertaken to investigate the performance of this procedure and to assess how the size of the panel data set being modelled influences the estimation procedure.

2.1 Multi-State models: the Markov model

Assume k disease states are defined, $S = \{1, ..., k\}$, and that individuals move independently between these states. A multi-state process on these states is governed by a continuous time stochastic process $X(t)$ which takes values in S and is characterised through the transition probabilities between different states

$$p_{ij}(t^*, t, F_{t^*}) = P\left(X(t) = j | X(t^*) = i, F_{t^*}\right), \tag{2.1}$$

for $j, i \in S$, $t^* \leq t$ and F_{t^*} the history (or filtration) of the process up to time t^*, or through transition intensities or rates

$$\lambda_{ij}(t, F_t) = \lim_{\Delta t \to 0} \frac{P(X(t + \Delta t) = j | X(t) = i, F_t)}{\Delta t}, \tag{2.2}$$

representing the instantaneous hazard of progressing to state j (Ross, 2003, p. 362).
Another way of looking at the transition rates is that in a continuous-time Markov model, a single period of occupancy (or sojourn time) in state i has an exponential distribution, with rate given by $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$. The remaining, $k-1$ transition intensities $(\lambda_{i1}, ..., \lambda_{i(i-1)}, \lambda_{i(i+1)}, ..., \lambda_{ik})$ are proportional to the probabilities governing the next state after i to which the individual makes a transition. The probability that the individual occupies state j immediately after state i is $-\lambda_{ij}/\lambda_{ii}$ (Jackson, 2011). To illustrate the interpretation of transition rates, assume the following two transition intensity matrices for an arbitrary 4-state model

$$Q = \begin{bmatrix} -0.1 & 0.1 & 0 & 0 \\ 0.1 & -0.6 & 0.5 & 0 \\ 0 & 0.7 & -1.0 & 0.3 \\ 0 & 0 & 0.5 & -0.5 \end{bmatrix} \tag{2.3}$$

and

$$Q = \begin{bmatrix} -0.01 & 0.01 & 0 & 0 \\ 0.01 & -0.06 & 0.05 & 0 \\ 0 & 0.07 & -0.10 & 0.03 \\ 0 & 0 & 0.05 & -0.05 \end{bmatrix}. \tag{2.4}$$

The first transition intensity matrix (2.3) provides us with the following information about its underlying multi-state process:

- A patient currently in state 1 can only make a transition to state 2. The mean sojourn time for state 1, i.e. the time that a patient spends in state 1 before transitioning to state 2, is $(0.1)^{-1} = 10$ units of time. If observation times are measured in days then this would be 10 days, while if the observation times are measured in months this would be 10 months.
- A patient currently in state 2 can make a transition to state 1 or to state 3. The probability that the transition is made to state 1 is $\frac{-0.1}{-(0.1+0.5)} = 0.167$ and the probability that the transition is made to state 3 is $\frac{-0.5}{-(0.1+0.5)} = 0.833$. The mean sojourn time for state 2, i.e. the mean time that a patient spends in state 2 before transitioning to state 1 or to state 3, is $(0.1 + 0.5)^{-1} = 1.67$ units of time.
- A patient currently in state 3 can make a transition to state 2 or to state 4. The probability that the transition is made to state 2 is $\frac{-0.7}{-(0.7+0.3)} = 0.7$ and the probability that the transition is made to state 4 is $\frac{-0.3}{-(0.7+0.3)} = 0.3$. The mean sojourn time for state 3, i.e. the mean time that a patient spends in state 3 before transitioning to state 2 or to state 4, is $(0.7+0.3)^{-1} = 1$ unit of time.
- A patient currently in state 4 can only make a transition to state 3. The mean sojourn time for state 4, i.e. the time that a patient spends in state 4 before transitioning to state 3, is $(0.5)^{-1} = 2$ units of time.

This information can be summarised in the following sojourn/probability (S/P) matrix

$$\begin{bmatrix} \mathbf{10} & 1.0 & 0 & 0 \\ 0.167 & \mathbf{1.67} & 0.833 & 0 \\ 0 & 0.70 & \mathbf{1} & 0.30 \\ 0 & 0 & 1.0 & \mathbf{2} \end{bmatrix}, \tag{2.5}$$

where the diagonal values represent the sojourn time for each state and the off-diagonal values represent the conditional transition probabilities (Jackson, 2011).

The S/P matrix associated with (2.4) is given by

$$\begin{bmatrix} \mathbf{100} & 1.0 & 0 & 0 \\ 0.167 & \mathbf{16.67} & 0.833 & 0 \\ 0 & 0.70 & \mathbf{10} & 0.30 \\ 0 & 0 & 1.0 & \mathbf{20} \end{bmatrix}. \tag{2.6}$$

Comparing (2.5) and (2.6) it can be seen that, although the conditional probabilities for the two underlying multi-state processes are identical, the sojourn times for the second process is

10 times longer than for the first process.

Different model assumptions can be made about the dependence of the transition rates (2.2) on time (Meira-Machado *et al.*, 2009):

- Markov assumption: Under this assumption the intensities only depend on the history of the process through the current state, and (2.1) and (2.2) can be simplified as

$$\begin{aligned} p_{ij}(t^*, t, F_{t^*}) &= p_{ij}(t^*, t) = P\left(X(t) = j | X(t^*) = i\right), \text{ and} \\ \lambda_{ij}(t, F_t) &= \lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P(X(t, t+\Delta t) = j | X(t) = i)}{\Delta t}. \end{aligned}$$

- Time homogeneous assumption: Under this assumption the intensities are assumed constant over time, and (2.1) and (2.2) become

$$p_{ij}(t^*, t, F_{t^*}) = p_{ij}(0, t - t^*) = P\left(X(t - t^*) = j | X(0) = i\right) = p_{ij}(t - t^*), \text{ and} \quad (2.7)$$

$$\lambda_{ij}(t, F_t) = \lambda_{ij} = \lim_{\Delta t \to 0} \frac{P(X(\Delta t) = j | X(0) = i)}{\Delta t}. \quad (2.8)$$

  The focus of this dissertation is on time homogeneous Markov models and these are the models that will be discussed in the next section.

- Semi-Markov assumption: Under this assumption the intensities not only depend on the current state, but also on the entry time into the current state, and (2.1) and (2.2) are written as

$$\begin{aligned} p_{ij}(t^*, t, F_{t^*}) &= p_{ij}(t^*, t, t_i) = P\left(X(t) = j | X(t^*) = i, t_i\right), \text{ and} \\ \lambda_{ij}(t, F_t) &= \lambda_{ij}(t_i) = \lim_{\Delta t \to 0} \frac{P(X(t, t+\Delta t) = j | X(t) = i, t_i)}{\Delta t}, \end{aligned}$$

  with $t_i < t$, the time state i was entered.

2.1.1 Limiting probabilities of Markov models

The probability that a continuous-time Markov chain will be in state j at time t often converges to a limiting value which is independent of the initial state (Ross, 2003, p. 368). These limiting probabilities, defined by

$$P_j \equiv \lim_{t \to \infty} p_{ij}(t), \quad j, i \in S,$$

with p_{ij} defined by (2.7), are useful in describing the long-term expected states of the process under study. If, for example, the progression of HIV is being investigated with a multi-state model, the limiting probabilities can be used to give an indication of the spread of the different HIV states in the infected population in 10 or 15 years time. Knowing what percentage of the infected population will be in the different states of the disease can then be used to plan treatment or healthcare facilities. If most infected individuals are expected to remain in state 1 in the long-term and state 1 only requires basic home care, then it would not be necessary for government to build expensive healthcare facilities for these infected individuals. If, on the other hand, most infected individuals are expected to be in state 3 and state 3 requires extensive medical care, then government would need to plan for the future expansion of healthcare facilities for these infected individuals.

Kolmogorov's forward equations

$$p'_{ij}(t) = \sum_{h \neq j} \lambda_{hj} p_{ih}(t) - \upsilon_j p_{ij}(t), \tag{2.9}$$

with $\upsilon_j = \sum_h \lambda_{jh}$ the rate at which the process makes a transition when in state j and h the states that can be visited from j, can be used to derive equations for the limiting probabilities, P_j (Ross, 2003, p. 369). Letting t approach ∞ in (2.9) and assuming the limit and summation can be interchanged gives

$$\begin{aligned} \lim_{t\to\infty} p'_{ij}(t) &= \lim_{t\to\infty} \left[\sum_{h \neq j} \lambda_{hj} p_{ih}(t) - \upsilon_j p_{ij}(t) \right] \\ &= \sum_{h \neq j} \lambda_{hj} P_h - \upsilon_j P_j. \end{aligned}$$

As $p_{ij}(t)$ is a bounded function, it follows that if $p'_{ij}(t)$ converges, then

$$\lim_{t\to\infty} p'_{ij}(t) = 0,$$

giving (Ross, 2003, p. 369)

$$\sum_{k \neq j} \lambda_{kj} P_k = \upsilon_j P_j, \text{ for all states } j.$$

This set of equations, along with the fact that

$$\sum_j P_j = 1$$

can be used to solve the limiting probabilities for a continuous-time Markov process. The solutions for the $P_j's$ are in general non-trivial and are dependent on the structure of the multi-state process under study. See Section 4.1 for how these equations are used to solve the limiting probabilities in a three-state Markov model.

2.1.2 Time homogeneous Markov models

Let Q be the $(k \times k)$ matrix of transition intensities with entries as defined in (2.8) for $i \neq j$, $\boldsymbol{\lambda}$ be a vector of length h, the number of independent parameters[1] (*See footnote 1*), and $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$ for $i = 1, ..., k$,

$$Q(\boldsymbol{\lambda}) = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k1} & \lambda_{k2} & \dots & \lambda_{kk} \end{bmatrix}, \tag{2.10}$$

and $P(t)$ be the $(k \times k)$ transition probability matrix with entries as defined in (2.7),

$$P(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \dots & p_{1k}(t) \\ p_{21}(t) & p_{22}(t) & \dots & p_{2k}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1}(t) & p_{k2}(t) & \dots & p_{kk}(t) \end{bmatrix}. \tag{2.11}$$

The Kolmogorov equations state that (Ross, 2003, pp.363-364)

$$\frac{d}{dt} P(t) = P(t)Q,$$

and they yield unique solutions for $P(t)$,

$$P(t) = e^{Qt} = \sum_{r=0}^{\infty} \frac{(Qt)^r}{r!}, \tag{2.12}$$

conditional on $P(0) = I$.

Although (2.12) is a direct solution for the transition probabilities in terms of the transition intensities, the solutions are complicated functions of the intensities and it is only practical to calculate them for very simple models with a small number of transition parameters. For example, assume a 3-state 2-parameter progressive model where patients can only move forward through the states and where the last state is an absorbing state, i.e. once entered the patients

[1] At most $h = k^2 - k$, but in general h is smaller, for example, in (2.22) and (2.25) $h = 3$, in (2.23) $h = 6$ and in (2.24) $h = 4$.

cannot leave that state. The transition matrix for this model is given by

$$Q(\boldsymbol{\lambda}) = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ 0 & -\lambda_{23} & \lambda_{23} \\ 0 & 0 & 0 \end{bmatrix},$$

with $\boldsymbol{\lambda} = (\lambda_{12}, \lambda_{23})$ and the solution to, for example $p_{13}(t)$, the probability that a patient starting in state 1 at time 0 will be in state 3 at time t, using (2.12) and $Q(\boldsymbol{\lambda})$ is given by

$$p_{13}(t) = \frac{1}{\lambda_{12} - \lambda_{23}} \left(\lambda_{12} - \lambda_{23} - \lambda_{12} e^{-t\lambda_{23}} + \lambda_{23} e^{-t\lambda_{12}}\right). \tag{2.13}$$

Solving (2.12) without the need to directly express the transition probabilities as functions of the transition rates can be accomplished with a canonical decomposition of Q (Kalbfleisch and Lawless, 1985). Let $d_1, ..., d_k$ be the distinct eigenvalues of Q and A a $k \times k$ matrix with jth column the right eigenvector corresponding to d_j, then

$$Q = ADA^{-1}, \tag{2.14}$$

where $D = diag(d_1, ..., d_k)$, and

$$P(t) = A \text{ diag}(e^{d_1 t}, ..., e^{d_k t}) A^{-1}. \tag{2.15}$$

The derivatives of the transition probabilities, required in the next section to estimate maximum likelihood estimates of parameters, are calculated in a similar way to (2.15). The matrix with entries $\partial p_{ij}(t; \boldsymbol{\lambda})/\partial \lambda_u$ is obtained as

$$\frac{\partial P(t)}{\partial \lambda_u} = A V_u A^{-1}, \quad u = 1, ..., h, \tag{2.16}$$

with h the number of independent transition rates, and V_u a $k \times k$ matrix with (i, j) entry

$$\begin{aligned} g_{ij}^{(u)}(e^{d_i t} - e^{d_j t})/(d_i - d_j), \quad & i \neq j, \\ g_{ii}^{(u)} t e^{d_i t}, \quad & i = j, \end{aligned}$$

and $g_{ij}^{(u)}$ the (i, j) entry in $G^{(u)} = A^{-1}(\partial Q/\partial \lambda_u)A$ (Kalbfleisch and Lawless, 1985).
The transition probabilities may be complicated functions of the transition intensities but, once the transition intensities are known, it is a trivial numerical exercise to calculate the transition probabilities. In the next section the estimation of the transition intensities and, by extension, the transition probabilities are investigated.

2.1.2.1 Maximum likelihood estimation

(*See footnote* 2)It is assumed that a sample of n individuals is observed and that the data for individual i consists of a series of time points $(t_{i1}, ..., t_{in_i})$ and the corresponding states at these time points, $(S(t_{i1}), ..., S(t_{in_i}))$, where $S(t_{in_i}) \in S$ as in Section 2.1. An individual's contribution to the likelihood is his or her path through the different states (Jackson *et al.*, 2003). In general, consider two states, $S(t_l)$ and $S(t_{l+1}) \in S$, observed at times t_l and t_{l+1}. The contribution of these two states to the likelihood is

$$L_{S(t_l),S(t_{l+1})} = p_{S(t_l)S(t_{l+1})}(t_{l+1} - t_l|\boldsymbol{\lambda}),$$

the $S(t_l)^{th}$ row and $S(t_{l+1})^{th}$ column of (2.11) evaluated at $t = t_{l+1} - t_l$.

The full-likelihood is the product of all such terms over all individuals and all observation times. Let $t_0 < t_1 < ... < t_m$ denote the unique observation times in the sample and let $n_{S(t_{l-1})S(t_l)l}$ denote the number of individuals in state $S(t_{l-1})$ at t_{l-1} and in state $S(t_l)$ at t_l, then the likelihood and log-likelihood functions are defined as (Kalbfleisch and Lawless, 1985)

$$L(\boldsymbol{\lambda}|data) = \prod_{l=1}^{m} \left\{ \prod_{S(t_{l-1}),S(t_l)=1}^{k} p_{S(t_{l-1})S(t_l)}(t_{l-1} - t_l|\boldsymbol{\lambda})^{n_{S(t_{l-1})S(t_l)l}} \right\}, \text{ and} \quad (2.17)$$

$$\log L(\boldsymbol{\lambda}|data) = \sum_{l=1}^{m} \sum_{S(t_{l-1}),S(t_l)=1}^{k} n_{S(t_{l-1})S(t_l)l} \log p_{S(t_{l-1})S(t_l)}(t_l - t_{l-1}|\boldsymbol{\lambda}), \quad (2.18)$$

where $\boldsymbol{\lambda}$ is defined as the vector of h independent unknown transition intensities in (2.10). It can be noted here that in a random sample of n patients, it normally happens that some of these $n_{S(t_{l-1})S(t_l)l}$ values are zero, since certain patients may have unobserved disease states at certain time points.

The likelihood function in (2.17) can be viewed as the general form for any multi-state model. Depending on the type of data observed in a study, this general form is extended or altered based on the data under study. The general form needs to be altered if (Jackson, 2005):

– The data under study includes a death state.

  In studies where the final state is death, it is common to know the time of death but the state just before death is not always known. If $S(t_{l+1}) = D$, with D defined as the death

[2] The notation used for different states will be adjusted in subsequent sections to accomodate the state individual i occupies at certain time points.

state, then the contribution to the likelihood is summed over the unknown states $S(.)$ on the day before death

$$L_{S(t_l),S(t_{l+1})} = \sum_{S(.)\neq D} p_{S(t_l)m}(t_{l+1} - t_l)\lambda_{mD},$$

assuming a time unit of days. The sum is taken over all possible states m which can be visited between $S(t_l)$ and D.

– The transition times are exactly observed.

If the times are exact transition times between the states, with no transitions between the observation times, then the contribution to the likelihood is

$$L_{S(t_l),S(t_{l+1})} = p_{S(t_l)S(t_l)}(t_{l+1} - t_l)\lambda_{S(t_l)S(t_{l+1})},$$

since the individual stays in state $S(t_l)$ in the interval t_l to t_{l+1} with a known transition at time t_{l+1}.

– There is censoring present in the data.

If, at the end of a study, it is known that a patient is alive but the state of the patient is unknown, or if it is known that a patient has left the study but the state in which the patient left the study is not known, that observation has to be treated as a censored observation. The contribution of a censored observation to the likelihood is

$$L_{S(t_l),S(t_{l+1})} = \sum_{S(.)\in C} p_{S(t_l)S(.)}(t_{l+1} - t_l),$$

with C defined as the known subset of states that the patient could have entered before being censored.

Various algorithms have been proposed to maximise (2.18) with regards to the unknown parameters in $\boldsymbol{\lambda}$. Here a quasi-Newton (or scoring) procedure, proposed by Kalbfleisch and Lawless (1985), is implemented to obtain the maximum likelihood estimates (MLE's) of $\boldsymbol{\lambda}$ and estimates of the asymptotic covariance matrix. Let $w_l = t_l - t_{l-1}$, $l = 1, ..., m$, then from (2.18)

the first and second derivatives of the log likelihood are given as

$$R_u(\boldsymbol{\lambda}) = \frac{\partial \log L(\boldsymbol{\lambda}|data)}{\partial \lambda_u} = \sum_{l=1}^{m} \sum_{S(t_{l-1}),S(t_l)=1}^{k} n_{S(t_{l-1})S(t_l)l} \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)/\partial \lambda_u}{p_{S(t_{l-1})S(t_l)}(w_l)}, \quad u = 1, ..., h, \tag{2.19}$$

and

$$\frac{\partial^2 \log L(\boldsymbol{\lambda}|data)}{\partial \lambda_u \partial \lambda_v} = \sum_{l=1}^{m} \sum_{S(t_{l-1}),S(t_l)=1}^{k} n_{S(t_{l-1})S(t_l)l} \times \left\{ \frac{\partial^2 p_{S(t_{l-1})S(t_l)}(w_l)/\partial \lambda_u \partial \lambda_v}{p_{S(t_{l-1})S(t_l)}(w_l)} - \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)/\partial \lambda_u \partial p_{S(t_{l-1})S(t_l)}(w_l)/\partial \lambda_v}{p^2_{S(t_{l-1})S(t_l)}(w_l)} \right\}.$$

Instead of directly using a Newton-Raphson algorithm and thus evaluating the first and second derivatives, a scoring device is used were the second derivatives are replaced by estimates of their expectations. This gives an algorithm that only requires the first derivatives of the log-likelihood (Kalbfleisch and Lawless, 1985).

Let $N_{S(t_{l-1})}(t_{l-1}) = \sum_{S(t_l)=1}^{k} n_{S(t_{l-1})S(t_l)l}$ denote the number of individuals in state $S(t_{l-1})$ at time t_{l-1}. Taking the expectation of $n_{S(t_{l-1})S(t_l)l}$ conditional on $N_{S(t_{l-1})}(t_{l-1})$ and noting that $\sum_{S(t_{l-1}),S(t_l)=1}^{k} \partial^2 p_{S(t_{l-1})S(t_l)}(w)/\partial \lambda_u \partial \lambda_v = 0$, gives (Kalbfleisch and Lawless, 1985)

$$E\left(-\frac{\partial^2 \log L(\boldsymbol{\lambda}|data)}{\partial \lambda_u \partial \lambda_v}\right) = \sum_{l=1}^{m} \sum_{S(t_{l-1}),S(t_l)=1}^{k} \frac{E\left\{N_{S(t_{l-1})}(t_{l-1})\right\}}{p_{S(t_{l-1})S(t_l)}(w_l)} \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)}{\partial \lambda_u} \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)}{\partial \lambda_v}.$$

This can be estimated by (Kalbfleisch and Lawless, 1985)

$$M_{uv}(\boldsymbol{\lambda}) = \sum_{l=1}^{m} \sum_{S(t_{l-1}),S(t_l)=1}^{k} \frac{N_i(t_{l-1})}{p_{S(t_{l-1})S(t_l)}(w_l)} \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)}{\partial \lambda_u} \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)}{\partial \lambda_v}. \tag{2.20}$$

The $p_{S(t_{l-1})S(t_l)}(w_l)$ and $\frac{\partial p_{S(t_{l-1})S(t_l)}(w_l)}{\partial \lambda_u}$ terms in (2.19) and (2.20) are computed using (2.15) and (2.16).

To obtain an estimate of $\boldsymbol{\lambda}$ using (2.19) and (2.20), let $\boldsymbol{\lambda}_0$ be an initial estimate of $\boldsymbol{\lambda}$, $R(\boldsymbol{\lambda})$ be the $h \times 1$ vector $(R_u(\boldsymbol{\lambda}))$ and $M(\boldsymbol{\lambda})$ the $h \times h$ matrix $(M_{uv}(\boldsymbol{\lambda}))$. An updated estimate $\boldsymbol{\lambda}_1$ is obtained as

$$\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_0 + M(\boldsymbol{\lambda}_0)^{-1} R(\boldsymbol{\lambda}_0),$$

where it is assumed that $M(\boldsymbol{\lambda}_0)$ is nonsingular. This process is repeated with $\boldsymbol{\lambda}_1$ replacing $\boldsymbol{\lambda}_0$, and with a good initial estimate, this produces $\hat{\boldsymbol{\lambda}}$ upon convergence (Kalbfleisch and Lawless,

1985).

Kay (1986) extends the procedure for censored data, while Gentleman *et al.* (1994) and Jackson (2005) implemented this procedure in the S and R programming languages respectively.

2.1.2.2 Incorporating covariates

In many situations the interest is not only in the progression of patients through the different disease states but also on how covariates influence this progression. To assess the effect of covariates, they are incorporated into the model by assuming that the transition intensities are functions of the covariates of interest and are of the form

$$\lambda_{S(t_{l-1})S(t_l)}(\mathbf{z}) = e^{\mathbf{z}'\boldsymbol{\beta}_{S(t_{l-1})S(t_l)}}, \quad S(t_{l-1}) \neq S(t_l),$$

with $\mathbf{z}$ a $(c \times 1)$ vector of c covariates and $\boldsymbol{\beta}_{S(t_{l-1})S(t_l)}$ their corresponding vector of regression coefficients.

Marshall and Jones (1995) described a proportional hazards type formulation for the transition intensities where the intensities in (2.8) are replaced by

$$\lambda_{S(t_{l-1})S(t_l)}(\mathbf{z}) = \lambda_{S(t_{l-1})S(t_l)} e^{\mathbf{z}'\boldsymbol{\beta}_{S(t_{l-1})S(t_l)}}, \tag{2.21}$$

with $\lambda_{S(t_{l-1})S(t_l)}$ the baseline transition rate, $\mathbf{z}$ the $(c \times 1)$ vector of c covariates and $\boldsymbol{\beta}_{S(t_{l-1})S(t_l)}$ their corresponding vector of regression coefficients. In the presence of time varying covariates (2.21) becomes

$$\lambda_{S(t_{l-1})S(t_l)}(\mathbf{z}(t)) = \lambda_{S(t_{l-1})S(t_l)} e^{\mathbf{z(t)}'\boldsymbol{\beta}_{S(t_{l-1})S(t_l)}}.$$

The quasi-Newton MLE algorithm discussed earlier in Section 2.1.2.1 can be extended to estimate the coefficients of the covariates (Kalbfleisch and Lawless, 1985). A different canonical decomposition of $Q(\mathbf{z})$ in (2.14) is now required for each of the r distinct covariate vectors in the sample. Let $\mathbf{z}_b = (z_{1b}, ..., z_{cb})$,

$$Q_b = Q(\mathbf{z}_b) = (\lambda_{S(t_{l-1})S(t_l)}(\mathbf{z}_b)), \quad b = 1, ..., r,$$

and $n^{(b)}_{S(t_{l-1})S(t_l)l}$ be the number of individuals with covariate values $\mathbf{z}_b$ that are in state $S(t_{l-1})$

at t_{l-1} and state $S(t_l)$ and t_l. The log-likelihood with covariates included in the model is

$$\log L(\boldsymbol{\lambda}, \boldsymbol{\beta}|data) = \sum_{b=1}^{r}\sum_{l=1}^{m}\sum_{S(t_{l-1}),S(t_l)=1}^{k} n^{(b)}_{S(t_{l-1})S(t_l)l} \log p_{S(t_{l-1})S(t_l)}(w_l; \mathbf{z}_b|\boldsymbol{\lambda}, \boldsymbol{\beta}),$$

with

$$P_b(t) = e^{Q_b t} = \left(p_{S(t_{l-1})S(t_l)}(t; \mathbf{z}_b)\right).$$

The score vector (2.19) now involves the sum of r terms, one for each distinct covariate vector,

$$R(\boldsymbol{\theta}) = \sum_{b=1}^{r} R^{(b)}(\boldsymbol{\theta}),$$

with $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\beta})$, the baseline transition intensities and regression parameters of the covariates that have to be estimated. Each $R^{(h)}(\boldsymbol{\theta})$ is a vector of length d, with $d = h + c$, the total number of parameters to be estimated in $\boldsymbol{\theta}$,

$$R^{(b)}_u(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta}|data)}{\partial \theta_u} = \sum_{l=1}^{m}\sum_{S(t_{l-1}),S(t_l)=1}^{k} n^{(b)}_{S(t_{l-1})S(t_l)l} \frac{\partial p_{S(t_{l-1})S(t_l)}(w_l; \boldsymbol{z}_b)/\partial\theta_u}{p_{S(t_{l-1})S(t_l)}(w_l; \mathbf{z}_b)}, \quad u = 1, ..., d,$$

is calculated using equations (2.15) and (2.16). Similarly the Fisher scoring matrix $M(\boldsymbol{\theta})$ in the presence of covariates

$$M(\boldsymbol{\theta}) = \sum_{b=1}^{r} M^{(b)}(\boldsymbol{\theta}),$$

it calculated using (2.20) for each b and equations (2.15) and (2.16). As the derivatives in (2.20) are now with respect to each element in $\boldsymbol{\theta}$, a separate diagonalisation is required of each Q_b (Kalbfleisch and Lawless, 1985).

2.1.2.3 Estimation problems

As with all numerical estimation techniques, the quasi-Newton procedure discussed in Section 2.1.2.1 can run into optimisation difficulty when implemented. Most of these problems are related to the shape of the likelihood function and the information available about the parameters (Kalbfleisch and Lawless, 1985). Kalbfleisch and Lawless (1985), Kay (1986), Gentleman *et al.* (1994) and Jackson (2005 and 2011) all describe situations where the optimisation procedure may fail to produce the correct estimates of the parameters. These include situations where:

– the transition intensities are of vastly different orders of magnitude.

– the times w_l between successive observations are large.
– over-complex models - be it over-complex transition matrices or including too many covariates in a model - are applied with insufficient data.

The end result for these situations is that the optimisation algorithm fails to find the maximum of the log-likelihood, or even fails to evaluate the likelihood (Jackson, 2011).

Measures that can be taken to try and overcome the estimation problem include:

– Parameterising the model by writing $\lambda_{ij} = \exp(a_{ij}), i \neq j$. This is due to the fact that the parameters a_{ij} can take any real value whereas $\lambda_{ij} \geq 0$ (Kalbfleisch and Lawless, 1985).
– Calculate the initial estimate, $\boldsymbol{\lambda}_0$, by examining the transition counts n_{ijl} in the data set and use several different initial values when fitting the proposed model (Jackson, 2011).
– Use a modified steplength procedure which provides better convergence properties than a standard quasi-Newton approach (Gentleman *et al.*, 1994).
– If there are too few observations to estimate a transition rate, states can be merged to increase the number of transition counts between states (Kay, 1986).

Based on these measures, the best course of action before fitting a multi-state model to a data set is to investigate the pairs of transition counts in the data. If it is found that there are too few transitions in general or too few transitions between specific states, it can be an indication that the maximum likelihood technique may not be able to find suitable parameter estimates. In Chapter 5 Bayesian techniques are developed to overcome this problem associated with small panel data sets.

2.1.3 Model structures

The types of transitions allowed in a model have implications for inferences about the model. Although most multi-state models are uniquely defined for the specific data under study, the following 4 models form the building blocks for most other multi-state models and these are the ones that will be considered in this dissertation:

– Progressive model

  The progressive model is the simplest multi-state model. It is a unidirectional model in that patients can only move forward to the next state. The final state in any progressive model is an absorbing state, typically death, and is such that once an individual enters this final

state, he or she can never leave that state. Figure 2.1 illustrates a typical 4-state progressive model with transition intensity matrix (2.22).

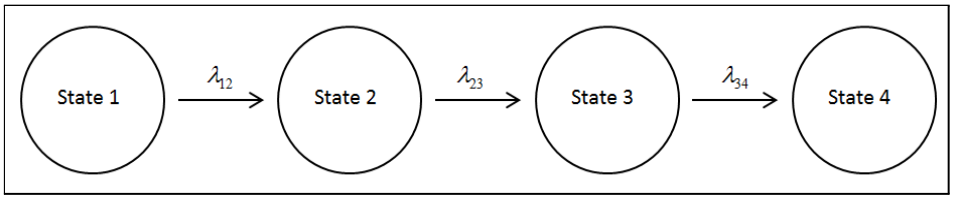

Figure 2.1: 4-State Progressive Model

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 & 0 \\ 0 & -\lambda_{23} & \lambda_{23} & 0 \\ 0 & 0 & -\lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.22}$$

Meira-Machado *et al.* (2008) investigated the effect of covariates on the recurrence and death of cancer patients using a 3-state ("Alive and Disease Free", "Alive and Recurrence", "Dead") progressive model. They compared their 3-state model with a traditional Cox model and showed that, while the two approaches had similar results, the 3-state model did highlight associations that were not evident when using the traditional Cox model.

Longini *et al.* (1989) fitted a 5-state model to HIV data to assess the waiting times of patients in the various stages of the HIV infection. Using this multi-state model they provided one of the most complete statistical descriptions at that time of the natural history of HIV infection.

– Recurrent model

Recurrent models do not have absorbing states and over time individuals move repeatedly between the different states. Figure 2.2 illustrates a 3-state recurring model with transition intensity matrix (2.23).

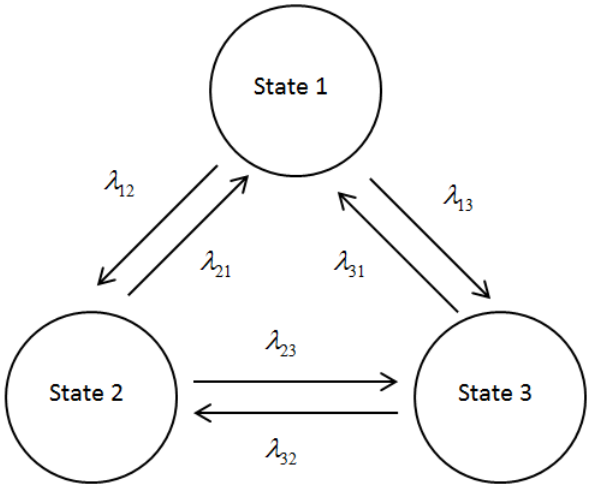

Figure 2.2: 3-State Recurring Model

$$Q = \begin{bmatrix} -(\lambda_{12} + \lambda_{12}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31} + \lambda_{32}) \end{bmatrix} \tag{2.23}$$

Marshall and Jones (1995) fitted a 4-state modification model with three transient states and a final absorbing state to diabetic retinopathy. Patients are allowed to move freely between the first three transient states (grade I, grades II-III, grades IV-V), but once the disease has progressed past grade V the patients can no longer move backwards and they enter the absorbing final state (grade VI). They provide estimates of the effects of the important covariates on the disease's progression and also calculate estimated survival curves for the probability of remaining free of state 4 (grade VI retinopathy) for subjects starting in one of the three transient states.

– Illness-death model

An illness-death model typically consists of three states: healthy, ill and death. It is similar to the recurrent model, but with one state, death, being an absorbing state. Figure 2.3 illustrates a 3-state illness-death model with transition intensity matrix (2.24).

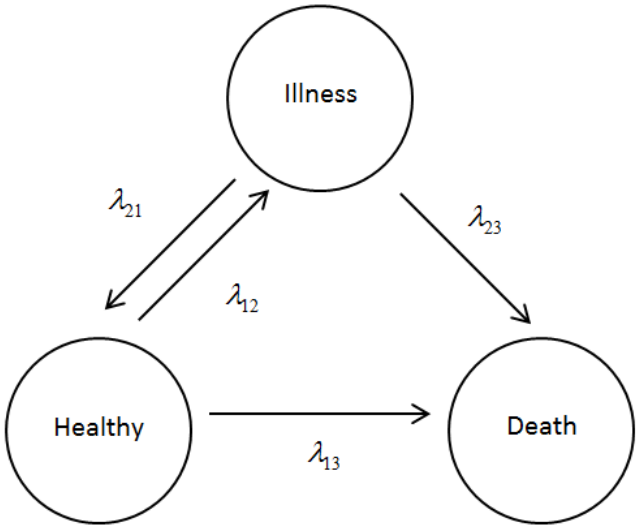

Figure 2.3: Illness-Death Model

$$Q = \begin{bmatrix} -(\lambda_{12} + \lambda_{12}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.24}$$

Pérez-Ocón *et al.* (1998) used an illness-death model in their analysis of 300 patients who had surgical treatment for breast cancer. Their healthy state is defined as a patient with no relapse after surgery, while illness is defined as having a relapse after surgery. After calculating and comparing the transition intensities for patients transitioning from healthy to death (λ_{13}) and from healthy to illness (λ_{23}), they could conclude that the most important marker in the survival time to breast cancer is the relapse time.

– Competing risk model

The competing risk model has several absorbing states, where, for example different causes of death are investigated simultaneously. Figure 2.4 illustrates a 4-state competing risk model with transition intensity matrix (2.25).

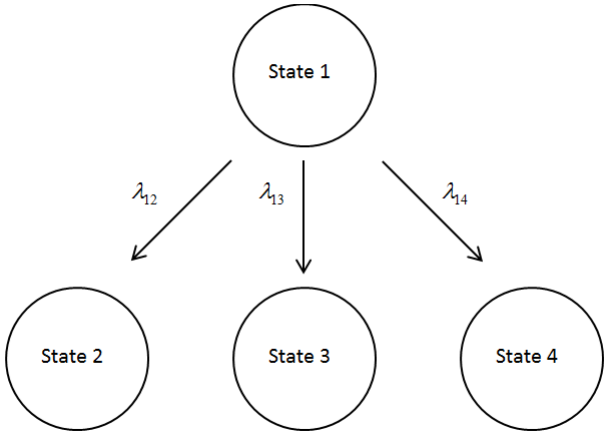

Figure 2.4: 4-State Competing Risk Model

$$Q = \begin{bmatrix} -(\lambda_{12}+\lambda_{13}+\lambda_{14}) & \lambda_{12} & \lambda_{13} & \lambda_{14} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.25}$$

Andersen *et al.* (2002) illustrated the use of the competing risk multi-state model by examining mortality after acute myocardial infarction. They followed 5983 patients who survived an acute myocardial infarction to ascertain if they died from sudden cardiovascular disease (S-CVD), non-sudden cardiovascular disease (NS-CVD) or non-cardiovascular disease (Non-CVD). From their multi-state model they were able to conclude that age was associated with an increased risk of mortality and that male gender was associated with an increased risk of S-CVD.

2.1.4 Multi-state survival models

Multi-state models play an important role in modelling disease progression and survival. When used in the survival context it is necessary to translate the transition rates in (2.10) into the fundamental survival analysis quantities; the survival function and the hazard rate.

Generally when modelling disease progression in a survival analysis context, the final state in the multi-state model is an absorbing state (typically death) and it is important to know how patients transition through the various states until reaching this final absorbing/death state.

While the transition matrix does provide all the necessary information about the multi-state process, the transition rates are in general not values that are easy to interpret. In the survival context the transition probabilities and hazard rates are the statistics of choice when trying to make sense of the underlying multi-state model. Using the transition probabilities, it is possible to generate survival plots that give the survival curves for the transient states in the model. If covariates are included in the model, the parameter estimates of the covariate effects (the $\beta's$ in 2.21) can be used to calculate the hazard ratios (e^{β}) for each covariate in the model. The hazard ratios show what effect each covariate has on the different transition rates in the model.

To illustrate this, assume a 3-state illness death model with one binary categorical variable ($z = 0, 1$) influencing the transition rates

$$Q = \begin{bmatrix} -(\lambda_{12}(z) + \lambda_{13}(z)) & \lambda_{12}(z) = \lambda_{12}e^{z\boldsymbol{\beta}_{12}} & \lambda_{13}(z) = \lambda_{13}e^{z\boldsymbol{\beta}_{13}} \\ \lambda_{21}(z) = \lambda_{21}e^{z\boldsymbol{\beta}_{21}} & -(\lambda_{21}(z) + \lambda_{23}(z)) & \lambda_{23}(z) = \lambda_{23}e^{z\boldsymbol{\beta}_{23}} \\ 0 & 0 & 1 \end{bmatrix}.$$

Table 2.1 gives the parameter estimates after fitting a multi-state model.

Table 2.1: Parameter estimates and hazard ratios of illness-death model.

Parameter	**Estimate**	**Hazard Ratio**
λ_{12}	0.037	–
λ_{13}	0.001	–
λ_{21}	0.059	–
λ_{23}	0.003	–
β_{12}	−0.1065	$e^{-0.1065} = 0.899$
β_{13}	−0.4133	$e^{-0.4133} = 0.661$
β_{21}	0.0503	$e^{0.0503} = 1.052$
β_{23}	0.2574	$e^{0.2574} = 1.294$

To calculate the survival probabilities given that the covariate $z = 1$,

$$P(t) = \exp\left(\begin{bmatrix} -(0.037e^{-0.1065} + 0.001e^{-0.413}) & 0.037e^{-0.1065} & 0.001e^{-0.413} \\ 0.059e^{0.050} & -(0.059e^{0.050} + 0.003e^{0.257}) & 0.003e^{0.257} \\ 0 & 0 & 1 \end{bmatrix} t \right),$$

is calculated for varying values of t. For $t = 1$ (days in this example) this gives

$$P(1) = \begin{bmatrix} 0.9678 & 0.0319 & 0.0003 \\ 0.0592 & 0.9375 & 0.0033 \\ 0 & 0 & 1 \end{bmatrix},$$

and for $t = 365$ this is

$$P(365) = \begin{bmatrix} 0.4029 & 0.2102 & 0.3869 \\ 0.3896 & 0.2033 & 0.4071 \\ 0 & 0 & 1 \end{bmatrix}.$$

These matrices show that if a patient is in state 1 at the beginning of the study, there is a $1 - 0.0003 = 0.9997$ probability that the patient will survive 1 day and a $1 - 0.3869 = 0.6131$ probability that the patient will survive 1 year. If a patient is in state 2 at the beginning of the study, there is $1 - 0.003 = 0.9967$ probability that the patient will be alive after 1 day and a $1 - 0.4071 = 0.5929$ probability that the patient will be alive after 1 year. A survival plot is now generated by computing $P(t)$ for different values of t and then plotting the probabilities of not being in the last (death) state (Figure 2.5).

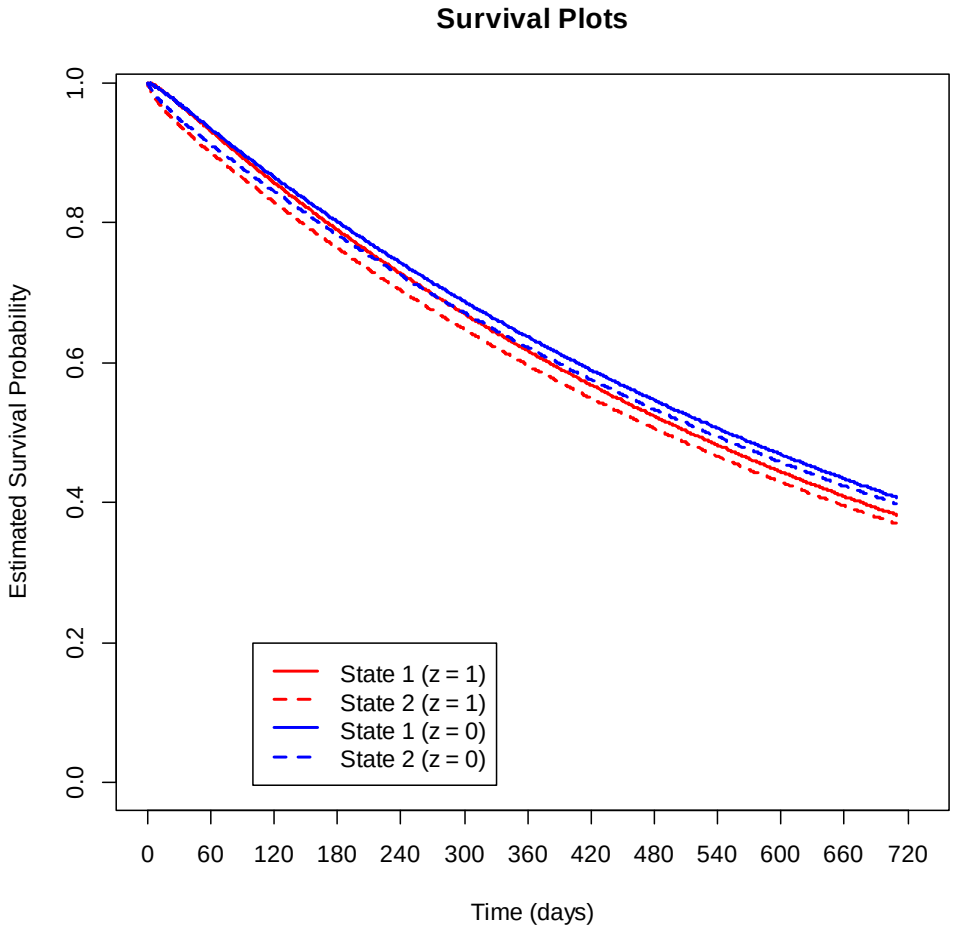

Figure 2.5: Survival probabilities based on table 2.1.

2.2 Assessing multi-state models

As with any statistical model it is important to assess and further investigate a multi-state model once it has been fitted to data. Statistical software can almost always generate parameter estimates, but in this case it is important to know if these are reliable estimates and if they provide useful insight into the data under study. In this section the following three areas that need to be assessed when fitting a multi-state model will be investigated:

- The assumptions of the model.
- The fit of the model.
- The effect of covariates in the model.

2.2.1 Assessing the assumptions of the model

The key assumptions to be investigated in the multi-state models presented in Section 2.1.2 are the Markov assumption and the assumption of homogeneity of the transition intensities across patients and across time. As these assumptions are fundamental in the creation of the multi-state model, it is important to validate and assess them.

- The Markov assumption

  The Markov assumption, that the future evolution of the process only depends on the current state and not on the past states, is a fundamental assumption for the above mentioned multi-state models. Unfortunately, as exact transition times are rarely observed, it is difficult to test this assumption explicitly. Kay (1986) proposed using interpolation to estimate exact transition times and then using these times to create a complete data set. Tests can then be performed on this complete data set to assess the Markov assumptions. For example, assume a 3-state model with recurrent transient states 1 and 2 and an absorbing death state 3. Let t be the time spent in state 2 in a previous transition from state 1. Fitting a model $\lambda_{23}(t) = \lambda_{23}\exp(\beta t)$ and testing $H_0 : \beta = 0$ would assess the assumption that the transition rate from state 2 to death is unaffected by the previous sojourn time (Kay, 1986). This same procedure can be used to assess other Markov assumptions. The accuracy of any conclusions, however, depends on the accuracy with which the exact transition times can be determined through interpolation.
- Homogeneity of the transition intensities across patients

The homogeneity of the intensities across the subject population can be tested by including covariate effects in the model. Suppose that the study population can be divided into two groups using a binary covariate x within a recurrent 3-state model. Let

$$\lambda_{ij}(x) = \lambda_{ij} e^{\beta_{ij} x}$$

with $x = 0, 1$, and $i, j = 1, 2, 3$, be the transition intensities in the model. Using a likelihood ratio test and testing $H_0 : \beta_{ij} = 0$, the hypothesis that the transition intensities differ with regard to the two groups in the study population can be tested. If no significant difference is found, the assumption of homogeneity of the transition intensities across the two groups under study has been validated.

– Homogeneity of the transition intensities across time

Faddy (1976) and Kay (1986) proposed fitting piecewise constant transition intensities and using a likelihood ratio test to test the assumption of constant intensities across time. Kalbfleisch and Lawless (1985) extended this idea by proposing a parametric time-dependent model using

$$\lambda_{ij}(t) = \lambda_{ij} e^{-\beta t}$$

as time-dependent transition intensities in the model. Testing $H_0 : \beta = 0$ can be used to assess the homogeneity of the intensities across time.

Gentleman *at al.* (1994) used a local score test to examine departures from homogeneity by considering

$$H_0 : \lambda_{ij}(t) = \lambda_{ij}$$

versus

$$H_A : \lambda_{ij}(t) = \lambda_{ij} t^{\gamma - 1} \text{ (power)}$$
$$\text{or} \qquad H_A : \lambda_{ij}(t) = \lambda_{ij} + t\gamma \text{ (linear)}.$$

The test statistic is the ratio of the partial derivative of the log-likelihood with respect to γ, evaluated at $(\hat{\theta}, \gamma = 1)$ (power) or $(\hat{\theta}, \gamma = 0)$ (linear). Under H_0 the test statistic has approximately a $N(0, 1)$ distribution. The advantage of this method is that only the time-homogeneous model needs to be fitted to the data (Titman and Sharples, 2010).

2.2.2 Assessing the fit of the model

Once the underlying assumptions are validated and the multi-state model is fitted, it is important to know if the estimated transition intensities adequately explain the data and the process under study. To this end informal model diagnostic tools as well as goodness-of-fit tests, can be used to assess the fit of the multi-state model.

– Informal diagnostic tools

As a multi-state model can be viewed as a combination of different simple survival models, one of the simplest informal methods to assess a model's fit is to use the Kaplan-Meier product limit estimate (Titman and Sharples, 2010). If the time of entry into a specific state is known exactly, plotting and comparing the empirical survival curve and the curve implied by the fitted survival model should give a good indication of the goodness-of-fit of that model. Unfortunately, as with many graphical techniques, determining whether an observed difference is significant is not straightforward. Pérez-Ocón *et al.* (1998 and 2001) and Titman and Sharples (2010) use a test of Hollander and Proschan (1976) to formally compare the fitted Markov curve and the Kaplan-Meier estimate.

Gentleman *at al.* (1994) proposed using the observed prevalence and expected transition counts to assess the goodness-of-fit of a multi-state model. The observed prevalence count for state u, $O_u^P(t)$, is the number of individuals in state u at time t, and the expected count, $E_u^P(t)$, is the product of the total number of individuals under observation at time t and the transition probability $\hat{P}_{1u}(t)$; assuming that all individuals are in state 1 at time 0. The observed transition counts, $O_{uv}^T(t_1, t_2)$, are the number of individuals observed in state u at time t_1 and in state v at time t_2. The expected transition counts, $E_{uv}^T(t_1, t_2)$, are the product of the number of individuals at risk in state u at time t_1 and the appropriate transition probability $\hat{P}_{uv}(t_2 - t_1)$. By investigating the matrix of observed minus expected values, or a scaled version such as

$$M_{uv} = \frac{(O_{uv} - E_{uv})^2}{E_{uv}}$$

departures from the fitted model can be detected. A large value of M_{uv} would indicate a poor fit, but due to the ad hoc interpolation of the observed states and the dependence between the rows of the tables, a formal test to determine if the deviances observed are

statistically significant is not possible (Titman and Sharples, 2010).

– Goodness-of-fit tests

Aguirre-Hernández and Farewell (2002) proposed a more generalised goodness-of-fit statistic based on partitioning the data using the observed transition points. Let C and R denote the number of categories the data is partitioned into based on the values of the covariates and the response variable respectively. If, for example, a 3-state recurrent Markov-model is analysed and we are not interested in covariates, then $C = 1$ and $R = 9$. As transition rates may depend on the length of time between transitions and also on the time at which the transition was observed, the data is further divided into H classes based on the length of the study-time (i.e. observations early on in the study are grouped together and observations later in the study are grouped together) and L_h intervals based on the quantiles of the length of the time intervals in category h $(h = 1, 2, ..., H)$. In studies where the time at which a transition was observed is unimportant $H = 1$ is used which also leads to the simplification $L_1 = L$ (Aguirre-Hernández and Farewell, 2002).

Let $E_{h,l,r,c}$ be the expected number of transitions in cell (h, l, r, c), calculated as the sum of the estimated transition probabilities in categories h, l, r, and c, and $O_{h,l,r,c}$ be the total number of observed transitions in categories h, l, r, and c (Aguirre-Hernández and Farewell, 2002). The AH/F (Aguirre-Hernández and Farewell) goodness-of-fit statistic is given by

$$T = \sum_{h=1}^{H} \sum_{l=1}^{L} \sum_{r=1}^{R} \sum_{c=1}^{C} \frac{(O_{h,l,r,c} - E_{h,l,r,c})^2}{E_{h,l,r,c}}.$$

For models without covariates the statistic is approximately χ^2 distributed with $(C - |\theta|)$ (the number of independent cells from the resulting contingency table minus the number of unknown parameters fitted from the data) degrees of freedom, but in the presence of covariates the exact distribution is intractable and a bootstrap procedure is required to determine significance (Aguirre-Hernández and Farewell, 2002).

The AH/F statistic is not suitable for data in which the time of entry of the absorbing state is known exactly or when the data under study include censored observations (Aguirre-Hernández and Farewell, 2002). Titman and Sharples (2007) proposed a modified goodness-of-fit statistic that can accommodate exact death times and censored observations. The modified method for exact death times involves imputing estimated times at which the next

observation would have taken place, had the patient survived. The resulting statistic has a null distribution with a mean roughly equal to $C - |\theta|$, but with a smaller variance than χ^2, and requires the use of the bootstrap obtain a more accurate p-value. In general cases, the null distribution of the statistic can be estimated by the parametric bootstrap procedure of repeatedly sampling from the fitted model, refitting the model and recomputing the test statistic, resulting in an accurate p-value (Jackson, 2011). Censoring in the data can be accommodated in two ways. Firstly, a separate category in the contingency table can be created for all censored observations and the AH/F statistic can then be used on the modified contingency table. As the number of categories that can be created are limited by the number of observations, creating a censored category may limit the use of other relevant categories. To this end a second approach is to include both censored and non-censored observations in the same category. The AH/F statistic is still appropriate under this second approach, as long as the expected transition probabilities for transitions to non-absorbing states are reweighted by the probability of not being censored (Titman and Sharples, 2007). This goodness-of-fit statistic is used in Section 5.5 to assess the fit of the proposed Bayesian techniques to model multi-state data.

2.2.3 Assessing the effect of covariates in the model

When covariates are included in a model the interest lies in knowing how these influence the flow of the patients in the study. One covariate may retard disease progression, that is it decreases the probability of a patient moving to a higher disease state, while another covariate may reverse disease progression, i.e. it increases the probability of a patient moving to a lower disease state. It is important to know if these covariate effects are significant and if they can be generalised to the population. To this end the effects of the covariates in the model need to be assessed for statistical significance. In the multi-state model setup this is done by using likelihood ratio and Wald tests (Marshall and Jones, 1995).

2.3 Simulating a panel data set

The multi-state models, structures and estimation techniques introduced in this chapter will form the building blocks for the remainder of this dissertation. To this end it will be important

to be able to simulate panel data with known transition rates and model structures that can then be used for further simulation studies. A simulation program was developed that is capable of simulating panel data with given transition rates (The R code used to simulate panel data is provided in the Appendix A.1.). In this section this simulation process is described and the process is assessed for correctness.

2.3.1 Simulation process and methodology

Any multi-state model is defined by its transition intensity matrix, Q (2.10), which in turn is used to calculate the transition probability matrix, $P(t)$ (2.11). Although $P(t)$ is a complicated function of Q, (2.12) can be used to quickly and easily calculate $P(t)$ for a given Q. This process, by calculating $P(t)$ using (2.12) with a given Q, will be used to simulate panel data in this dissertation.

Define the following quantities that will be used in, and that forms part of, the simulation process:

- Let $\mathbf{t}_{Pos} = (t_0, ..., t_m)$ be a vector of possible observation times for all patients with $t_0 = 0$. If, for example, $\mathbf{t}_{Pos} = (0, 1, 2, ..., 23, 24)$ with t measured in months, then this indicates that patients are observed over a two year period with observations taking place every month. If $\mathbf{t}_{Pos} = (0, 2, 4, ..., 34, 36)$ with t measured in months, patients are observed over a three year period with observations taking place every second month.
- Let n be the number of patients in the study.
- Let Q be the $(k \times k)$ transition intensity matrix for the multi-state process being simulated.
- Let $\boldsymbol{\beta}_{Cov}$ be the $(c \times 1)$ vector of known covariate effects if covariates are present in the data. The influence of the covariates on the baseline transition intensities in Q is modelled using (2.21).
- Let $o^{\%Mis}$ be the maximum percentage of missing observations per patient and $o_i^{\%Mis}$ the actual percentage of missing observations for patient i, with $o_i^{\%Mis} \leq o^{\%Mis}$. In an ideal world each patient will be observed at each of the m observation times defined by $\mathbf{t}$. Unfortunately, as patients leave studies early or miss prescribed visits, this rarely happens in practice. Decreasing or increasing $o^{\%Mis}$ leads to data sets with more or fewer missing values.

- Let $o_i = (1 - o_i^{\%Mis}) \times m$ be the actual number of observations for patient i, $i = 1, ..., n$.
- Let $\mathbf{t}_i = (t_{i0}, t_{i1}, ..., t_{io_i})$ be the vector of actual observation times for patient i, $i = 1, ..., n$, and $\mathbf{t}_i \subset \mathbf{t}_{Pos}$.
- Let $S(t_{il})$ be the simulated state of patient i at time l, with $i = 1, ..., n$ and $l = 0, 1, ..., o_i$.

Once $\mathbf{t}_{Pos}$, n, Q and $\boldsymbol{\beta}_{Cov}$ are defined the simulation process proceeds as follows:

1) Generate $o_i^{\%Mis}$, the actual percentage of missing observations for patient i, from a $unif(0, o^{\%Mis})$ distribution and calculate $o_i = (1-o_i^{\%Mis}) \times m$, the actual number of observations for patient i, $i = 1, ..., n$.
2) If covariates are to be included in the simulated data set, generate a covariate value for patient i from an appropriate distribution (if no covariates are included, skip this step).
3) Sample $\mathbf{t}_i = (t_{i0}, t_{i1}, ..., t_{io_i})$ from $\mathbf{t}_{Pos}$ for patient i. As all patients are seen at time 0, $t_{i0} = 0$.
4) Generate the initial state, $S(t_{i0})$ at time t_{i0} for patient i, from a $dis\ unif(1, k)$ distribution. This initial value will depend on the type of model structure being simulated. If, for example, a recurring structure (2.23) is being simulated the initial state can be any one of the possible states, while if a progressive (2.22), illness-death (2.24) or competing risk (2.25) structure is simulated, the initial value is selected so as not to be one of the absorbing states.
5) Calculate the time difference between the current observation, l, $l = 0, ..., o_i$, and the next observation $l+1$, $t = t_{i(l+1)} - t_{il}$ and use (2.12) to calculate the transition probability matrix, $P(t)$, between the two observations. In the presence of covariates (2.21) and $\boldsymbol{\beta}_{Cov}$ are used to calculate $P(t)$.
6) Generate $S(t_{i(l+1)})$ from a k-dimensional multinomial distribution with parameter 1 and probability vector equal to row $S(t_{il})$ of $P(t)$.
7) Repeat steps 5 and 6 o_i times for all elements in $\mathbf{t}_i$.
8) Repeat steps 1 to 7 n times for all patients in the study.

In order to assess the simulation process, varying values of $\mathbf{t}_{Pos}, n, o^{\%Mis}$, and Q will be used in this study. These values, and the measures that will be used to assess the simulation process, are presented in the following two sections.

2.3.1.1 Different models and scenarios

The transition matrix completely defines a multi-state model being investigated. For this simulation study 6 different transition matrices, thus 6 different multi-state models, will be investigated. There is an endless number of possible multi-state models that can be investigated; the scope of this dissertation will be limited to 3- and 4-state recurring models where transitions are only allowed between adjoining states. As the data under study is assumed to be patients moving between various states of a disease, this assumption is reasonable as a patient currently in state 1 has to move through state 2 before he or she can be classified as being in state 3. The models that will be investigated are:

1) A 3-state model where the transitions between the different states are similar across all possible states. This implies that the probability of transitioning to the next higher or lower state is the same across all states. The mean time spent in state 1 is $(0.5)^{-1} = 2$ months, in state 2 it is $(1)^{-1} = 1$ month and in state 3 $(0.5)^{-1} = 2$ months.

$$Q_1 = \begin{bmatrix} -0.5 & 0.5 & 0 \\ 0.5 & -1 & 0.5 \\ 0 & 0.5 & -0.5 \end{bmatrix} \tag{2.26}$$

2) A 3-state model with transitions to a lower state assumed to be more probable than transitioning to a higher state and once in a lower state patients are less likely to transition back to higher states. The rate of transitioning to a higher state is assumed to be 0.25, compared to the rate of transitioning to a lower state of 0.75. The mean time spent in state 1 is $(0.25)^{-1} = 4$ months, in state 2 it is $(1)^{-1} = 1$ month and in state 3 $(0.75)^{-1} = 1.33$ months.

$$Q_2 = \begin{bmatrix} -0.25 & 0.25 & 0 \\ 0.75 & -1 & 0.25 \\ 0 & 0.75 & -0.75 \end{bmatrix} \tag{2.27}$$

3) A 3-state model that is the opposite of the model 2. Under this model patients are more likely to move to higher states and once in a higher state they spend more time in that state before moving to the next state. The rate of transitioning to a higher state is assumed to be 0.75, compared to the rate of transitioning to a lower state of 0.25. The mean time spent in state 1 is $(0.75)^{-1} = 1.33$ months, in state 2 it is $(1)^{-1} = 1$ month and in state 3

$(0.25)^{-1} = 4$ months.

$$Q_3 = \begin{bmatrix} -0.75 & 0.75 & 0 \\ 0.25 & -1 & 0.75 \\ 0 & 0.25 & -0.25 \end{bmatrix} \tag{2.28}$$

4) A 4-state model where the transitions between the different states are similar across all possible states. This implies that the probability of transitioning to the next higher or lower state is the same across all states. The mean time spent in states 1 or 4 is $(0.5)^{-1} = 2$ months, and in states 2 or 3 it is $(1)^{-1} = 1$ month.

$$Q_4 = \begin{bmatrix} -0.5 & 0.5 & 0 & 0 \\ 0.5 & -1 & 0.5 & 0 \\ 0 & 0.5 & -1 & 0.5 \\ 0 & 0 & 0.5 & -0.5 \end{bmatrix} \tag{2.29}$$

5) A 4-state model with transitions to a lower state assumed to be more probable than transitioning to a higher state and once in a lower state patients are less likely to transition back to higher states. The rate of transitioning to a higher state is assumed to be 0.25, compared to the rate of transitioning to a lower state of 0.75. The mean time spent in state 1 is $(0.25)^{-1} = 4$ months, in states 2 or 3 it is $(1)^{-1} = 1$ month and in state 4 it is $(0.75)^{-1} = 1.33$ months.

$$Q_5 = \begin{bmatrix} -0.25 & 0.25 & 0 & 0 \\ 0.75 & -1 & 0.25 & 0 \\ 0 & 0.75 & -1 & 0.25 \\ 0 & 0 & 0.75 & -0.75 \end{bmatrix} \tag{2.30}$$

6) A 4-state model that is the opposite of the model 5. Under this model patients are more likely to move to higher states and once in a higher state they spend more time in that state before moving to the next state. The rate of transitioning to a higher state is assumed to be 0.75, compared to the rate of transitioning to a lower state of 0.25. The mean time spent in state 1 is $(0.75)^{-1} = 1.33$ months, in states 2 or 3 it is $(1)^{-1} = 1$ month and in state 4 it is $(0.25)^{-1} = 4$ months.

$$Q_6 = \begin{bmatrix} -0.75 & 0.75 & 0 & 0 \\ 0.25 & -1 & 0.75 & 0 \\ 0 & 0.25 & -1 & 0.75 \\ 0 & 0 & 0.25 & -0.25 \end{bmatrix} \tag{2.31}$$

To investigate the effect of the sample size and the number of observations per patient, the 6 different data scenarios presented in Table 2.2 are simulated for each Q.

Table 2.2: Data scenarios used in the simulation process.

Scenario	Sample Size	Max. % Missingness
Sc1	25	10%
Sc2	25	50%
Sc3	50	10%
Sc4	50	50%
Sc5	75	10%
Sc6	75	50%

Two different covariate models will be investigated:

A) A model with one categorical variable. The effect of the categorical covariate on the transition rates (the β_{ij}^{Cat} values in (2.21)) is assumed to be -0.7 for all transition rates. This has the effect of the transition rates being halved ($e^{-0.7} = 0.497$) when the covariate is present. This can be thought of as a patient receiving a specific treatment that retards the rate at which the patient moves to a next state.

B) A model with one continuous and one categorical variable. The effect of the categorical covariate is kept at $\beta_{ij}^{Cat} = -0.7$ for all transition rates, and the effect of the continuous variable is set at $\beta_{ij}^{Con} = 0.01$. This has the effect that for every 1 unit increase in the continuous variable, the transition rates increase by 1% ($e^{1\times 0.01} = 1.01$). This can be thought of as for every year that a patient is older, the transition rates increase by 1% and the patient makes a quicker transition to a next state.

2.3.1.2 Assessing the simulation process

To assess if a data set is representative of the specified population, and thus if the simulation process is simulating the data correctly, repeated data sets are generated. For each one of the possible $Q - Sc - covariate$ combinations 5000 data sets are generated, a multi-state model, using the methods described in Section 2.1.2.1, is fitted to each generated data set and the parameter estimates for each model is stored. The distribution of the aggregated results of the 5000 fitted models are then investigated to ascertain if they are consistent with the known population values used to simulate the data set.

The mean square error (MSE) will be the main statistic used to assess the performance of

the simulation process. A smaller MSE will indicate that the estimates of the transition rates can be viewed as representative of the population under study, while a large MSE indicates a departure from the population values. As the MSE can be influenced by extreme values, the median square error (MedSE) will be presented in cases where extreme MSE values are observed. The use of the MedSE will be clearly highlighted in the results.

The bias, another indicator as to the performance of the simulation process, and defined as

$$Bias(\hat{\theta}) = \sqrt{MSE - var(\hat{\theta})},$$

will be presented for a select number of simulation runs.

The mean, median and standard deviation of the aggregated results are also investigated for each simulation run. Large MSE's and standard deviations will indicate that there are extreme values present in the aggregated results. This points to a measure of instability in the parameter estimation. If this is observed it will be important to ascertain if this is due to the simulation process described here or due to the estimation process employed in Section 2.1.2.1.

2.3.2 Illustrating the simulation process

To illustrate the process described in the previous section, assume a sample of only size two, $n = 2$, from a population with the following transition intensity matrix:

$$Q = \begin{bmatrix} -0.1 & 0.1 & 0 & 0 \\ 0.1 & -0.6 & 0.5 & 0 \\ 0 & 0.7 & -1.0 & 0.3 \\ 0 & 0 & 0.5 & -0.5 \end{bmatrix},$$

and transition probability matrix

$$P(t) = \exp\left(\begin{bmatrix} -0.1 & 0.1 & 0 & 0 \\ 0.1 & -0.6 & 0.5 & 0 \\ 0 & 0.7 & -1.0 & 0.3 \\ 0 & 0 & 0.5 & -0.5 \end{bmatrix} t\right) = \begin{bmatrix} 0.909 & 0.075 & 0.015 & 0.001 \\ 0.075 & 0.640 & 0.246 & 0.039 \\ 0.021 & 0.345 & 0.479 & 0.155 \\ 0.003 & 0.091 & 0.259 & 0.647 \end{bmatrix}^t,$$

is required. Assume these individuals will be followed-up monthly over a two year period, that the maximum percentage of missing observations for any one patient over this period is 90% and that no covariates are measured. This gives $\mathbf{t}_{Pos} = (0, 1, ..., 24)$ and $o^{\%Mis} = 0.90$.

The simulation process for the individuals now follows as (refer to Section 2.3.1):

1) Randomly generate $o_1^{\%Mis}$ from a $unif(0, 0.90)$ distribution, $o_1^{\%Mis} = 0.88$, and calculate o_1, $o_1 = (1 - 0.88) \times 24 = 2.88 \rightarrow 3$.
2) Skip this step as no covariates are included in the data.
3) Randomly sample 3 observation points from $\mathbf{t}_{Pos} = (0, 1, ..., 24) \rightarrow \mathbf{t}_1 = (0, 10, 16, 21)$.
4) Randomly generate $S(0)$ from a $dis\ unif(1, 4)$ distribution $\rightarrow S(0) = 2$.
5) $t = 10 - 0 = 10$, giving

$$P(10) = e^{10Q} = \begin{bmatrix} 0.493 & 0.259 & 0.163 & 0.085 \\ \mathbf{0.259} & \mathbf{0.339} & \mathbf{0.249} & \mathbf{0.153} \\ 0.229 & 0.348 & 0.260 & 0.163 \\ 0.197 & 0.356 & 0.271 & 0.176 \end{bmatrix}.$$

6) Randomly generate $S(10)$ from a $multi(1, [0.259, 0.339, 0.249, 0.153])$ distribution $\rightarrow S(10) = 1$.
7) Repeat steps 5 and 6 to generate $S(16) = 1$ and $S(21) = 3$ with

$$P(16 - 10) = e^{6Q} = \begin{bmatrix} \mathbf{0.616} & \mathbf{0.216} & \mathbf{0.118} & \mathbf{0.050} \\ 0.216 & 0.365 & 0.265 & 0.154 \\ 0.165 & 0.372 & 0.283 & 0.180 \\ 0.117 & 0.360 & 0.300 & 0.223 \end{bmatrix}$$

and

$$P(21 - 16) = e^{5Q} = \begin{bmatrix} \mathbf{0.659} & \mathbf{0.199} & \mathbf{0.102} & \mathbf{0.040} \\ 0.199 & 0.379 & 0.272 & 0.150 \\ 0.143 & 0.380 & 0.292 & 0.185 \\ 0.091 & 0.350 & 0.309 & 0.250 \end{bmatrix},$$

giving

$$\mathbf{S}_1 = (2, 1, 1, 3)$$

8) Repeat steps 1 to 7 for individual 2 giving

$$\begin{aligned} o_2 &= 9 \\ \mathbf{t}_2 &= (0, 2, 4, 6, 8, 13, 14, 16, 18, 24) \\ \mathbf{S}_2 &= (2, 2, 3, 2, 3, 1, 1, 1, 1, 3). \end{aligned}$$

The simulated data set for these two individuals is presented in Table 2.3. The process described in Section 2.1.2.1 is now used to estimate the transition rates associated with this generated data set and the statistics described in the previous section are used to assess if this data is representative of the population parameters used to generate the data.

Table 2.3: Generated multi-state data set

Patient	Time (in months)	State
1	0	2
1	10	1
1	16	1
1	21	3
2	0	2
2	2	2
2	4	3
2	6	2
2	8	3
2	13	1
2	14	1
2	16	1
2	18	1
2	24	3

2.3.3 Simulation results

The mean, median, standard deviation and MSE (or MedSE) for each simulation run of 5000 repetitions are given in Tables 2.4 to 2.17. The population parameters used to simulate the data are given in the first row of each table. (See Appendix A.1 for the R functions and programs used to generate these results.)

The results can be summarised as follows:

- With the exception of model Q_2 and data scenario 2, if 3-state models with no covariates are simulated, the MSE's remain less than 0.08 across all models and data scenarios (Tables 2.4 to 2.6). The bias, here only shown for model Q_3 (Table 2.10), remains small for all models expect model Q_2. This indicates that the process developed to simulate multi-state models using known parameters is generating data sets that are consistent with the population parameters.
    - Data scenario 2 has the least number of observations (only 25 patients with each patient having up to 50% missing observations), hence it is to be expected that the estimators from this data set may at times be inconsistent with the parameters used to generate the data. This is especially true when covariates are introduced into the modelling process and the complexity of the models also increased. The MSE's for this scenario is as high as 400 when no covariates are modelled (Table 2.5), 450 when

1 covariate is included (Table 2.13) and jumps to over 7750 (Table 2.15) when two covariates are included.

- For model Q_2 the rate of transitioning to a lower state is 3 times more than the rate of transitioning to a higher state (0.75 vs. 0.25). This means that when the multi-state data set is generated it is more likely to include $2 \rightarrow 1$ and $3 \rightarrow 2$ transitions than $1 \rightarrow 2$ and $2 \rightarrow 3$ transitions. Given this, as well as the fact that under scenario 2 there are only 25 patients and that each patient can have up to 50% missing observations, it is very likely that most data sets generated under this model and scenario will include a very limited number of $1 \rightarrow 2$ and $2 \rightarrow 3$ transitions. With only a limited number of transitions in the likelihood and therefore the log-likelihood (2.17 and 2.18), it is to be expected that the scoring procedure used to estimate the parameters (2.19 and 2.20) of the multi-state model will be very unstable. In most cases the procedure will not be able to converge to a global maximum value, but will rather find local maxima for the parameters. This fact is clearly illustrated by the large standard deviations across the 5000 repetitions of models that have the large MSE's.

– As the complexity of the models are increased by the introduction of more states (Tables 2.7 to 2.9) and covariates (Tables 2.11 to 2.16), the size of the data sets being used becomes more important to ensure consistent estimates. This is due to the fact that as the number of possible transitions increases and each transition is influenced by covariates, more observations are required to correctly estimate the transition rates and the effects of the covariates.

  - Increasing the number of states from 3 to 4, increases the possible number of transitions from 4 to 6 (tables 2.7 to 2.9). For Q_4, the model where a patient is equally likely to transition to a higher or a lower state, large MSE's are only observed in the two scenarios with 25 patients (Table 2.7). With the increase in complexity, high MSE's are observed even in scenarios with 50 and 75 patients (Tables 2.8 and 2.9).
  - When modelling with one covariate, data scenarios 1 and 2 (both with only 25 patients) have MSE's that at times are as high as 957 (Table 2.13). For scenarios 3 and 4 (both with 50 patients) this decreases to a maximum MSE of 0.60 and for scenarios 5 and 6, the maximum MSE is 0.36 (Table 2.13).

- Data scenarios 5 and 6 (both with 75 patients) are the only scenarios where almost all the MSE's are below 0.75 (with the exception of Q_2 and scenario 6 (Table 2.15) where there is a MSE of 4.792 for λ_{32}) when including two covariates in the model. The small bias for both of these scenarios is evident from Table 2.17 (shown for models Q_1 and Q_3).

These results show that the simulation technique does simulate data sets that are representative of the underlying multi-state model. The maximum likelihood estimation techniques do provide reliable estimates of the population parameters, as long as the size of the data sets are sufficiently large for the models being fitted. If, for example, 3-state models with no covariates are being fitted, any one of the 6 data scenarios can be used with the required model matrix (Q-matrix). As the complexity of the models increased to 4-state models and models with covariates, even the data scenarios with 75 patients did not have enough information to provide reliable estimates. There are just too few transitions in the data set to provide enough information for the scoring procedure to find reliable parameter estimates. This is consistent with findings by Kalbfleisch and Lawless (1985), Kay (1986), Gentleman *et al.* (1994) and Jackson (2005 and 2011), that were also noted in Section 2.1.2.3 where some of the procedures that can be used to overcome this problem were mentioned.

Table 2.4: Simulating Q_1 in the presence of no covariates.

Q1	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}		λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.5	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Sc1	**Mean**	0.448	0.418	0.412	0.435	**Sc2**	0.472	0.454	0.457	0.480
	Med	0.435	0.409	0.404	0.423		0.447	0.431	0.423	0.456
	SD	0.101	0.095	0.092	0.097		0.155	0.154	0.158	0.168
	MSE	0.012	0.015	0.016	0.013		0.024	0.026	0.026	0.028
Sc3	**Mean**	0.428	0.406	0.408	0.434	**Sc4**	0.456	0.438	0.443	0.456
	Med	0.422	0.400	0.405	0.429		0.442	0.426	0.427	0.440
	SD	0.069	0.065	0.060	0.065		0.099	0.094	0.102	0.107
	MSE	0.009	0.013	0.012	0.008		0.011	0.012	0.013	0.013
Sc5	**Mean**	0.428	0.405	0.404	0.430	**Sc6**	0.451	0.433	0.438	0.453
	Med	0.424	0.402	0.403	0.427		0.442	0.428	0.430	0.445
	SD	0.054	0.051	0.050	0.055		0.078	0.074	0.077	0.079
	MSE	0.008	0.011	0.011	0.007		0.008	0.010	0.009	0.008

Table 2.5: Simulating Q_2 in the presence of no covariates.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}		λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.25	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Sc1	**Mean**	0.234	0.701	0.182	0.686	**Sc2**	68.8	103.3	66.9	33.8
	Med	0.229	0.683	0.174	0.643		0.232	0.693	0.209	0.677
	SD	0.051	0.141	0.066	0.242		13.0	7.250	19.65	9.87
	MSE	0.002	0.022	0.009	0.062		289.8	106.0	403.7	101.1
Sc3	**Mean**	0.230	0.697	0.179	0.657	**Sc4**	0.239	0.713	0.209	0.708
	Med	0.226	0.690	0.175	0.638		0.233	0.691	0.200	0.672
	SD	0.035	0.100	0.047	0.154		0.054	0.148	0.071	0.217
	MSE	0.001	0.012	0.007	0.032		0.003	0.023	0.006	0.048
Sc5	**Mean**	0.228	0.687	0.178	0.655	**Sc6**	0.237	0.711	0.209	0.693
	Med	0.226	0.681	0.175	0.646		0.233	0.697	0.201	0.666
	SD	0.030	0.079	0.037	0.117		0.045	0.126	0.060	0.173
	MSE	0.001	0.010	0.006	0.022		0.002	0.017	0.005	0.033

Table 2.6: Simulating Q_3 in the presence of no covariates.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}		λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.75	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Sc1	**Mean**	0.713	0.187	0.694	0.231	**Sc2**	0.771	0.226	0.749	0.249
	Med	0.656	0.178	0.679	0.228		0.682	0.200	0.704	0.234
	SD	0.266	0.071	0.142	0.050		0.404	0.120	0.198	0.049
	MSE	0.072	0.009	0.023	0.002		0.164	0.015	0.067	0.008
Sc3	**Mean**	0.661	0.182	0.692	0.229	**Sc4**	0.716	0.212	0.711	0.236
	Med	0.642	0.180	0.682	0.227		0.675	0.205	0.689	0.229
	SD	0.147	0.045	0.100	0.036		0.238	0.074	0.155	0.054
	MSE	0.029	0.006	0.013	0.001		0.057	0.006	0.025	0.003
Sc5	**Mean**	0.652	0.178	0.689	0.229	**Sc6**	0.696	0.211	0.710	0.237
	Med	0.640	0.174	0.684	0.228		0.668	0.204	0.692	0.233
	SD	0.119	0.038	0.078	0.028		0.184	0.060	0.126	0.044
	MSE	0.023	0.006	0.009	0.001		0.036	0.005	0.017	0.002

Table 2.7: Simulating Q_4 in the presence of no covariates.

Q4	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
		0.5	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Sc1	**Mean**	0.545	0.457	0.397	0.395	0.552	0.792
	Med	0.299	0.289	0.289	0.285	0.309	0.301
	SD	1.230	0.873	0.402	0.406	1.564	4.179
	MSE	4.968	1.885	0.172	0.175	2.447	17.535
Sc2	**Mean**	0.790	0.658	0.580	0.625	0.900	1.242
	Med	0.347	0.335	0.335	0.327	0.313	0.317
	SD	1.869	1.159	1.208	1.151	4.951	7.499
	MSE	3.575	1.368	1.463	1.340	24.645	56.723
Sc3	**Mean**	0.367	0.339	0.331	0.326	0.352	0.385
	Med	0.301	0.288	0.293	0.282	0.290	0.298
	SD	0.242	0.217	0.174	0.197	0.252	0.330
	MSE	0.076	0.073	0.059	0.069	0.085	0.122
Sc4	**Mean**	0.504	0.457	0.414	0.422	0.450	0.526
	Med	0.339	0.320	0.332	0.321	0.323	0.331
	SD	0.352	0.479	0.308	0.354	0.628	0.932
	MSE	0.304	0.231	0.102	0.131	0.396	1.065
Sc5	**Mean**	0.363	0.344	0.326	0.318	0.339	0.354
	Med	0.310	0.297	0.298	0.290	0.283	0.284
	SD	0.204	0.193	0.141	0.149	0.241	0.284
	MSE	0.060	0.062	0.050	0.055	0.084	0.102
Sc6	**Mean**	0.432	0.411	0.385	0.381	0.402	0.428
	Med	0.339	0.326	0.332	0.318	0.309	0.310
	SD	0.358	0.333	0.223	0.255	0.329	0.401
	MSE	0.133	0.119	0.063	0.079	0.118	0.166

Table 2.8: Simulating Q_5 in the presence of no covariates.

Q5	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{34}	λ_{43}
		0.25	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Sc1	**Mean**	0.261	0.804	0.370	1.857	1.496	15
	Med	0.201	0.605	0.137	0.534	0.065	0.177
	SD	0.274	0.789	3.043	6.310	21.24	31.54
	MedSE	0.002**	0.012**	0.007**	0.027**	0.010**	0.060**
Sc2	**Mean**	0.350	1.034	0.582	5.144	5.410	213
	Med	0.199	0.605	0.152	0.542	0.091	0.281
	SD	0.751	2.041	1.844	69	78.63	99.87
	MSE	0.003**	0.022**	0.006**	0.034**	0.016**	0.083**
Sc3	**Mean**	0.225	0.696	0.192	0.801	0.561	5.922
	Med	0.200	0.613	0.143	0.552	0.059	0.182
	SD	0.122	0.356	0.173	1.215	5.691	54
	MedSE	0.001**	0.008**	0.007**	0.021**	0.006**	0.030**
Sc4	**Mean**	0.255	0.778	0.252	0.997	1.493	15
	Med	0.198	0.583	0.155	0.551	0.083	0.233
	SD	0.186	0.926	0.200	1.830	25	205
	MSE	0.002**	0.012**	0.004**	0.020**	0.008**	0.042**
Sc5	**Mean**	0.226	0.678	0.177	0.627	0.224	0.772
	Med	0.210	0.638	0.147	0.518	0.062	0.175
	SD	0.089	0.249	0.123	0.469	1.157	1.587
	MSE	0.008	0.067	0.020	0.235	1.842	0.958
Sc6	**Mean**	0.235	0.710	0.222	0.777	0.541	0.895
	Med	0.201	0.611	0.158	0.551	0.084	0.265
	SD	0.145	0.412	0.234	1.149	1.018	1.589
	MSE	0.021	0.171	0.056	1.560	1.719	2.535

Table 2.9: Simulating Q_6 in the presence of no covariates.

Q6	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
		0.75	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Sc1	**Mean**	588	1.061	0.980	0.322	0.691	0.243
	Med	0.368	0.074	0.517	0.126	0.554	0.183
	SD	20.548	16	3.126	2.847	0.523	0.158
	MedSE	0.058**	0.010**	0.027**	0.008**	0.012**	0.002**
Sc2	**Mean**	16	2.91	9.116	0.488	0.946	0.325
	Med	0.428	0.097	0.569	0.144	0.567	0.184
	SD	181	39	67.549	2.243	1.574	0.478
	MedSE	0.087**	0.015**	0.034**	0.006**	0.022**	0.003**
Sc3	**Mean**	1.171	0.398	0.632	0.162	0.635	0.209
	Med	0.349	0.077	0.517	0.122	0.572	0.183
	SD	6.626	5.412	0.676	0.140	0.304	0.117
	MedSE	0.029**	0.006**	0.020**	0.007**	0.007**	0.001**
Sc4	**Mean**	2.915	0.456	0.712	0.220	0.700	0.239
	Med	0.415	0.105	0.536	0.146	0.554	0.185
	SD	48	2.547	0.723	0.299	0.544	0.196
	MedSE	0.042**	0.008**	0.019**	0.004**	0.011**	0.002**
Sc5	**Mean**	0.731	0.186	0.571	0.151	0.619	0.204
	Med	0.384	0.086	0.511	0.127	0.574	0.188
	SD	1.258	0.687	0.302	0.101	0.243	0.093
	MSE	1.113	0.476	0.123	0.020	0.076	0.011
Sc6	**Mean**	0.976	0.338	0.632	0.191	0.659	0.229
	Med	0.432	0.097	0.546	0.150	0.566	0.195
	SD	2.721	1.876	0.403	0.168	0.369	0.160
	MedSE	3.252	3.525	0.176	0.032	0.144	0.026

Table 2.10: Bias for models Q_3 and Q_6.

Model	**Sc**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Q3	1	0.036	0.063	0.056	0.018		
	2	0.018	0.023	0.167	0.074		
	3	0.089	0.067	0.058	0.020		
	4	0.032	0.037	0.038	0.014		
	5	0.097	0.072	0.060	0.020		
	6	0.054	0.039	0.039	0.012		
Q6	1	53.908	0.618	0.207	1.206	0.056	0.150
	2	14.756	2.351	40.308	0.227	0.190	0.073
	3	0.366	3.436	0.116	0.088	0.115	0.041
	4	1.544	0.190	0.030	0.029	0.047	0.009
	5	1.237	0.061	0.178	0.099	0.131	0.045
	6	0.385	0.065	0.117	0.059	0.090	0.020

Table 2.11: Simulating Q_1 in the presence of 1 covariate.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$
		0.5	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Sc1	**Mean**	0.442	0.419	0.427	0.451	-0.135	-0.120	-0.145	-0.158
	Med	0.417	0.391	0.407	0.422	-0.124	-0.101	-0.158	-0.177
	SD	0.164	0.199	0.159	0.172	0.481	0.479	0.466	0.469
	MSE	0.030	0.046	0.030	0.032	0.258	0.262	0.241	0.240
Sc2	**Mean**	0.556	0.533	0.479	0.498	-0.131	-0.111	-0.098	-0.135
	Med	0.426	0.418	0.412	0.429	-0.090	-0.100	-0.096	-0.115
	SD	1.195	1.051	0.315	0.227	0.666	0.669	0.632	0.638
	MSE	1.430	1.105	0.100	0.104	0.472	0.483	0.440	0.434
Sc3	**Mean**	0.420	0.399	0.406	0.435	-0.157	-0.137	-0.154	-0.168
	Med	0.415	0.386	0.395	0.416	-0.151	-0.151	-0.141	-0.165
	SD	0.106	0.098	0.106	0.114	0.353	0.350	0.348	0.352
	MSE	0.016	0.019	0.019	0.017	0.145	0.149	0.142	0.141
Sc4	**Mean**	0.465	0.446	0.453	0.471	-0.169	-0.144	-0.135	-0.155
	Med	0.433	0.418	0.423	0.435	-0.174	-0.129	-0.139	-0.130
	SD	0.176	0.157	0.198	0.196	0.444	0.431	0.466	0.467
	MSE	0.032	0.027	0.041	0.039	0.214	0.210	0.244	0.239
Sc5	**Mean**	0.427	0.398	0.399	0.431	-0.187	-0.143	-0.150	-0.188
	Med	0.419	0.394	0.392	0.423	-0.191	-0.153	-0.155	-0.194
	SD	0.084	0.082	0.080	0.087	0.284	0.297	0.288	0.290
	MSE	0.012	0.017	0.016	0.012	0.093	0.112	0.105	0.096
Sc6	**Mean**	0.455	0.433	0.432	0.451	-0.167	-0.134	-0.129	-0.157
	Med	0.435	0.415	0.415	0.429	-0.170	-0.123	-0.114	-0.168
	SD	0.129	0.125	0.114	0.118	0.378	0.379	0.361	0.367
	MSE	0.018	0.020	0.017	0.016	0.160	0.170	0.160	0.155

Table 2.12: Simulating Q_2 in the presence of 1 covariate.

Q2	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$
		0.25	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Sc1	**Mean**	0.238	0.723	0.944	4.034	-0.240	-0.233	-0.015	-0.122
	Med	0.226	0.674	0.165	0.657	-0.255	-0.243	-0.027	-0.108
	SD	0.094	0.280	19.91	21.8	0.514	0.448	1.000	0.874
	MSE	0.009	0.079	571.6	518.3	0.268	0.205	1.081	0.795
Sc2	**Mean**	0.575	1.592	0.506	14.0	-0.242	-0.243	-0.076	-0.173
	Med	0.234	0.694	0.199	0.697	-0.203	-0.215	-0.038	-0.150
	SD	6.7	15.8	4.515	11.15	0.780	0.764	1.336	1.131
	MSE	43.4	243.7	20.4	165.2	0.611	0.586	1.833	1.295
Sc3	**Mean**	0.228	0.697	0.177	0.678	-0.222	-0.233	-0.077	-0.162
	Med	0.223	0.677	0.167	0.635	-0.226	-0.232	-0.069	-0.152
	SD	0.055	0.162	0.071	0.243	0.330	0.307	0.574	0.469
	MSE	0.003	0.029	0.010	0.064	0.114	0.098	0.378	0.239
Sc4	**Mean**	0.255	0.778	0.227	0.768	-0.255	-0.255	-0.109	-0.157
	Med	0.232	0.698	0.199	0.654	-0.248	-0.233	-0.083	-0.156
	SD	0.174	0.671	0.149	0.518	0.474	0.443	0.720	0.607
	MSE	0.030	0.451	0.022	0.382	0.226	0.198	0.554	0.388
Sc5	**Mean**	0.229	0.689	0.175	0.650	-0.244	-0.240	-0.080	-0.159
	Med	0.223	0.679	0.170	0.631	-0.247	-0.242	-0.083	-0.177
	SD	0.045	0.121	0.056	0.181	0.271	0.241	0.456	0.381
	MSE	0.002	0.018	0.008	0.043	0.076	0.061	0.256	0.165
Sc6	**Mean**	0.240	0.715	0.215	0.728	-0.229	-0.233	-0.123	-0.171
	Med	0.231	0.677	0.193	0.663	-0.239	-0.237	-0.091	-0.155
	SD	0.072	0.194	0.107	0.357	0.386	0.342	0.596	0.506
	MSE	0.005	0.038	0.012	0.128	0.154	0.121	0.387	0.272

Table 2.13: Simulating Q_3 in the presence of 1 covariate.

Q3	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$
		0.75	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Sc1	**Mean**	31.6	4.342	0.738	0.238	-0.124	-0.047	-0.230	-0.216
	Med	0.634	0.168	0.673	0.220	-0.097	-0.038	-0.223	-0.220
	SD	27.4	13.1	0.356	0.104	0.811	0.959	0.465	0.503
	MSE	957.2	173.4	0.208	0.011	0.688	0.982	0.221	0.260
Sc2	**Mean**	22.9	0.270	66.65	17.7	-0.142	-0.080	-0.268	-0.252
	Med	0.711	0.209	0.716	0.233	-0.155	-0.072	-0.237	-0.236
	SD	19.51	0.337	17.9	16.7	1.198	1.450	0.932	0.966
	MSE	450.3	0.114	399.2	325.3	1.458	2.148	0.868	0.935
Sc3	**Mean**	0.689	0.178	0.702	0.232	-0.183	-0.078	-0.238	-0.235
	Med	0.639	0.168	0.681	0.224	-0.171	-0.096	-0.240	-0.224
	SD	0.260	0.075	0.162	0.057	0.475	0.560	0.310	0.342
	MSE	0.071	0.010	0.028	0.003	0.238	0.362	0.099	0.121
Sc4	**Mean**	0.777	0.229	0.745	0.250	-0.175	-0.122	-0.246	-0.267
	Med	0.673	0.203	0.696	0.234	-0.175	-0.096	-0.249	-0.286
	SD	0.474	0.176	0.156	0.033	0.631	0.757	0.437	0.474
	MSE	0.225	0.031	0.065	0.008	0.414	0.604	0.193	0.223
Sc5	**Mean**	0.655	0.178	0.683	0.228	-0.157	-0.114	-0.233	-0.238
	Med	0.621	0.170	0.672	0.224	-0.150	-0.113	-0.228	-0.246
	SD	0.206	0.059	0.118	0.042	0.377	0.459	0.231	0.265
	MSE	0.051	0.008	0.018	0.002	0.162	0.244	0.057	0.074
Sc6	**Mean**	0.717	0.210	0.712	0.238	-0.169	-0.095	-0.230	-0.236
	Med	0.655	0.195	0.687	0.230	-0.166	-0.066	-0.227	-0.224
	SD	0.296	0.100	0.187	0.065	0.478	0.567	0.334	0.366
	MSE	0.088	0.011	0.036	0.004	0.245	0.363	0.116	0.138

Table 2.14: Simulating Q_1 in the presence of 2 covariates.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
		0.5	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Sc1	**Mean**	0.470	0.425	0.434	0.481	-0.025	0.015	0.010	-0.056	0.021	0.022	0.015	0.017
	Med	0.324	0.309	0.317	0.342	-0.004	0.026	0.020	-0.042	0.020	0.019	0.015	0.016
	SD	0.613	0.676	0.586	0.586	0.490	0.462	0.460	0.474	0.091	0.085	0.082	0.086
	MSE	0.377	0.462	0.347	0.471	0.695	0.724	0.716	0.640	0.008	0.007	0.007	0.007
Sc2	**Mean**	0.861	0.779	0.838	0.992	-0.061	-0.042	-0.023	-0.096	0.019	0.018	0.024	0.026
	Med	0.365	0.366	0.333	0.342	-0.049	-0.006	0.001	-0.085	0.017	0.017	0.023	0.021
	SD	3.142	2.397	5.088	7.061	0.562	0.563	0.534	0.568	0.124	0.120	0.125	0.128
	MSE	9.993	5.819	25.1	50.0	0.725	0.750	0.743	0.687	0.015	0.015	0.016	0.017
Sc3	**Mean**	0.388	0.362	0.360	0.384	-0.077	-0.032	-0.029	-0.073	0.017	0.016	0.018	0.019
	Med	0.341	0.313	0.314	0.338	-0.064	-0.017	-0.005	-0.044	0.018	0.018	0.018	0.019
	SD	0.235	0.230	0.215	0.223	0.380	0.398	0.397	0.386	0.057	0.054	0.054	0.055
	MSE	0.068	0.072	0.066	0.063	0.532	0.605	0.607	0.542	0.003	0.003	0.003	0.003
Sc4	**Mean**	0.480	0.446	0.451	0.480	-0.089	-0.041	-0.042	-0.083	0.020	0.022	0.020	0.020
	Med	0.360	0.333	0.338	0.353	-0.066	-0.004	-0.012	-0.058	0.019	0.024	0.018	0.021
	SD	0.606	0.649	0.453	0.566	0.444	0.423	0.426	0.434	0.080	0.078	0.082	0.083
	MSE	0.498	0.423	0.207	0.320	0.570	0.613	0.614	0.569	0.007	0.006	0.007	0.007
Sc5	**Mean**	0.426	0.391	0.387	0.421	-0.029	0.007	-0.011	-0.044	0.011	0.012	0.013	0.013
	Med	0.400	0.367	0.368	0.397	-0.020	0.022	0.004	-0.025	0.010	0.013	0.013	0.012
	SD	0.163	0.146	0.139	0.159	0.281	0.270	0.270	0.276	0.040	0.039	0.036	0.037
	MSE	0.032	0.033	0.032	0.031	0.152	0.167	0.156	0.141	0.002	0.002	0.001	0.001
Sc6	**Mean**	0.467	0.449	0.456	0.472	-0.024	-0.006	-0.028	-0.059	0.013	0.012	0.013	0.014
	Med	0.405	0.391	0.404	0.419	0.002	0.004	-0.006	-0.054	0.011	0.012	0.015	0.014
	SD	0.432	0.340	0.282	0.311	0.328	0.321	0.320	0.321	0.054	0.054	0.053	0.054
	MSE	0.188	0.118	0.081	0.097	0.184	0.189	0.176	0.161	0.003	0.003	0.003	0.003

Table 2.15: Simulating Q_2 in the presence of 2 covariates.

Q2	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$	$\boldsymbol{\beta_{12}^{Con}}$	$\boldsymbol{\beta_{21}^{Con}}$	$\boldsymbol{\beta_{23}^{Con}}$	$\boldsymbol{\beta_{32}^{Con}}$
		0.25	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Sc1	**Mean**	0.266	0.829	0.993	7.972	-0.404	-0.411	0.283	0.056	0.015	0.012	0.040	0.041
	Med	0.201	0.624	0.128	0.487	-0.384	-0.414	0.295	0.087	0.015	0.011	0.026	0.027
	SD	0.476	1.5	17.6	25.8	0.533	0.478	0.800	0.792	0.075	0.040	0.174	0.189
	MSE	0.227	2.1	309.5	921.6	0.371	0.312	1.606	1.197	0.006	0.005	0.031	0.037
Sc2	**Mean**	0.606	1.793	3.142	23.2	-0.355	-0.397	0.126	0.000	0.013	0.016	0.037	0.041
	Med	0.203	0.643	0.146	0.520	-0.322	-0.367	0.194	-0.002	0.016	0.017	0.032	0.032
	SD	4.7	14.2	13.5	27.78	0.750	0.660	0.966	0.877	0.098	0.103	0.209	0.228
	MSE	22.5	202.1	192.6	7762	0.682	0.527	1.616	1.258	0.012	0.011	0.044	0.053
Sc3	**Mean**	0.230	0.684	0.181	0.737	-0.432	-0.425	0.186	-0.038	0.015	0.015	0.031	0.032
	Med	0.203	0.618	0.128	0.508	-0.434	-0.419	0.239	0.000	0.016	0.014	0.026	0.023
	SD	0.119	0.320	0.232	1.121	0.383	0.363	0.631	0.576	0.048	0.044	0.101	0.103
	MSE	0.015	0.106	0.058	1.489	0.219	0.207	1.183	0.770	0.002	0.002	0.011	0.011
Sc4	**Mean**	0.280	0.844	0.538	1.741	-0.401	-0.422	0.132	0.008	0.013	0.013	0.037	0.036
	Med	0.217	0.634	0.141	0.482	-0.401	-0.415	0.168	0.044	0.011	0.015	0.035	0.029
	SD	0.335	1.029	7.1	14.8	0.510	0.472	0.678	0.636	0.068	0.063	0.131	0.131
	MSE	0.113	1.066	50.4	220.3	0.350	0.300	1.152	0.906	0.005	0.004	0.018	0.018
Sc5	**Mean**	0.233	0.711	0.188	0.687	-0.169	-0.174	0.063	-0.026	0.012	0.010	0.014	0.016
	Med	0.221	0.675	0.164	0.598	-0.171	-0.164	0.089	-0.015	0.012	0.010	0.012	0.014
	SD	0.077	0.227	0.121	0.447	0.273	0.256	0.423	0.387	0.034	0.023	0.062	0.057
	MSE	0.006	0.053	0.019	0.204	0.092	0.082	0.310	0.225	0.001	0.001	0.004	0.003
Sc6	**Mean**	0.262	0.774	0.264	0.979	-0.159	-0.155	0.043	-0.037	0.009	0.009	0.013	0.012
	Med	0.229	0.694	0.186	0.638	-0.144	-0.147	0.043	-0.034	0.008	0.010	0.011	0.012
	SD	0.149	0.408	0.327	2.178	0.370	0.352	0.496	0.464	0.052	0.036	0.085	0.060
	MSE	0.022	0.167	0.107	4.792	0.156	0.145	0.363	0.285	0.003	0.002	0.007	0.006

Table 2.16: Simulating Q_3 in the presence of 2 covariates.

Q3	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$	$\boldsymbol{\beta_{12}^{Con}}$	$\boldsymbol{\beta_{21}^{Con}}$	$\boldsymbol{\beta_{23}^{Con}}$	$\boldsymbol{\beta_{32}^{Con}}$
		0.75	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Sc1	**Mean**	9.537	0.833	0.822	0.269	0.054	0.236	-0.399	-0.410	0.044	0.041	0.015	0.015
	Med	0.479	0.116	0.598	0.203	0.091	0.253	-0.407	-0.418	0.032	0.034	0.016	0.014
	SD	11.2	6.8	1.3	0.316	0.788	0.802	0.501	0.563	0.189	0.172	0.069	0.073
	MSE	148.1	46.6	1.6	0.100	1.189	1.519	0.342	0.401	0.037	0.030	0.005	0.005
Sc2	**Mean**	6.840	1.386	2.715	0.763	0.021	0.091	-0.387	-0.368	0.051	0.055	0.015	0.016
	Med	0.531	0.148	0.629	0.212	0.052	0.115	-0.353	-0.320	0.035	0.039	0.013	0.008
	SD	16.54	5.448	3.658	10.1	0.910	1.017	0.659	0.697	0.222	0.224	0.106	0.113
	MSE	587.3	47.1	19.2	102.5	1.346	1.658	0.532	0.595	0.051	0.052	0.011	0.013
Sc3	**Mean**	0.849	0.217	0.690	0.233	0.005	0.209	-0.428	-0.432	0.026	0.025	0.015	0.014
	Med	0.498	0.131	0.626	0.210	0.043	0.249	-0.439	-0.426	0.018	0.023	0.014	0.014
	SD	2.680	0.879	0.315	0.119	0.568	0.610	0.364	0.389	0.099	0.093	0.043	0.048
	MSE	13.8	0.772	0.102	0.015	0.819	1.198	0.206	0.223	0.010	0.009	0.002	0.002
Sc4	**Mean**	1.356	0.323	0.823	0.273	-0.021	0.130	-0.406	-0.385	0.026	0.030	0.014	0.014
	Med	0.546	0.156	0.627	0.209	0.008	0.163	-0.392	-0.362	0.025	0.025	0.014	0.017
	SD	3.987	0.678	0.939	0.310	0.625	0.697	0.470	0.517	0.130	0.131	0.068	0.073
	MSE	16.6	0.465	0.886	0.096	0.852	1.175	0.307	0.366	0.017	0.018	0.005	0.005
Sc5	**Mean**	0.690	0.187	0.701	0.234	-0.021	0.045	-0.155	-0.155	0.015	0.014	0.011	0.010
	Med	0.593	0.160	0.661	0.224	-0.003	0.065	-0.147	-0.151	0.016	0.017	0.012	0.011
	SD	0.410	0.113	0.239	0.085	0.387	0.420	0.256	0.278	0.057	0.061	0.024	0.028
	MSE	0.172	0.017	0.059	0.008	0.227	0.295	0.086	0.098	0.003	0.004	0.001	0.001
Sc6	**Mean**	0.829	0.239	0.777	0.258	-0.060	0.010	-0.148	-0.136	0.016	0.017	0.012	0.011
	Med	0.615	0.184	0.665	0.226	-0.030	0.028	-0.141	-0.104	0.015	0.015	0.013	0.009
	SD	0.846	0.223	0.462	0.149	0.455	0.514	0.341	0.385	0.072	0.080	0.047	0.041
	MSE	0.722	0.050	0.214	0.022	0.264	0.361	0.139	0.175	0.005	0.006	0.002	0.003

Table 2.17: Bias for Q_3 in the presence of 2 covariates.

Model	**Sc**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$	$\boldsymbol{\beta_{12}^{Con}}$	$\boldsymbol{\beta_{21}^{Con}}$	$\boldsymbol{\beta_{23}^{Con}}$	$\boldsymbol{\beta_{32}^{Con}}$
Q1	1	0.022	0.072	0.064	0.357	0.675	0.715	0.710	0.644	0.011	0.011	0.005	0.006
	2	0.347	0.269	0.298	0.438	0.639	0.658	0.677	0.604	0.008	0.008	0.013	0.015
	3	0.111	0.137	0.140	0.116	0.623	0.668	0.671	0.627	0.007	0.006	0.008	0.008
	4	0.362	0.050	0.047	0.008	0.611	0.659	0.658	0.617	0.010	0.012	0.010	0.010
	5	0.073	0.109	0.112	0.079	0.270	0.307	0.289	0.255	0.001	0.002	0.002	0.002
	6	0.030	0.050	0.043	0.027	0.276	0.294	0.272	0.241	0.003	0.002	0.003	0.004
Q3	1	4.759	0.542	0.060	0.016	0.753	0.936	0.301	0.290	0.034	0.031	0.004	0.005
	2	17.713	4.170	2.411	0.401	0.720	0.790	0.312	0.331	0.040	0.045	0.004	0.006
	3	2.521	0.018	0.059	0.016	0.705	0.909	0.272	0.268	0.014	0.014	0.004	0.003
	4	0.593	0.069	0.067	0.021	0.679	0.830	0.293	0.315	0.016	0.020	0.004	0.003
	5	0.059	0.063	0.048	0.016	0.278	0.345	0.145	0.145	0.005	0.004	0.024	0.022
	6	0.074	0.008	0.022	0.007	0.240	0.310	0.151	0.164	0.005	0.006	0.002	0.031

2.4 Conclusion

In this chapter multi-state models where introduced as the model of choice when analysing panel data.

The maximum likelihood method used to estimate the parameters of the model, how covariates are incorporated into the Markov model, different types of multi-state models, estimation problems that may arise when fitting multi-state models and how multi-state models can be interpreted from a survival analysis point of view, were discussed in Section 2.1.

In Section 2.2, the criteria used to assess the assumptions, fit and covariates in a multi-state model, were discussed.

In Section 2.3, a process to simulate panel data from a given transition intensity matrix was given, and using 6 different data scenarios, six different multi-state models were simulated and investigated. This procedure will play an important role in the remainder of the dissertation, since it is the procedure that will be used to simulate multi-state data in subsequent chapters. It was shown that the simulation procedure produces data sets that are consistent with the parameters. The simulation process also showed that as the complexity of the multi-state models increases, be it by including more states or by including covariates into the modelling process, the maximum likelihood procedure used to estimate the parameters can become unstable. In general, 3-state models can be fitted and parameters can be estimated with as few as 25 patients, while 4-state models needed at least 50 patients in the data sets. For more complex models, where covariates are included, it was found that even 75 patients in the data set is not sufficient to guarantee stable parameter estimates. These conclusions are similar to those found by Kalbfleisch and Lawless (1985), Kay (1986), Gentleman *et al.* (1994) and Jackson (2005 and 2011) and their solutions to this problem were given in Section 2.1.2.3.

The remainder of the dissertation discusses a Bayesian approach to inference based on the multi-state models presented here. The Bayesian approach, which supplements information in panel data via relevant prior knowledge, will address many of the problems and difficulties noted in this chapter when fitting multi-state models. It will be shown that the Bayesian approach is especially useful when working with small panel data sets.

Bayesian models and Bayesian multi-state data imputation techniques will be introduced in subsequent chapters. Once a Bayesian multi-state model is developed, the model assessment

tools introduced in this chapter will be used to assess the fit in a practical example of panel data.

3

Bayesian Modelling

In both the frequentist and Bayesian approach to statistical inference, a model containing parameters is specified, an experiment is conducted to obtain data and this is then used to provide insight into the underlying model and its parameters.

In the frequentist or classical framework the model parameters are assumed to be unknown but fixed quantities and the experimental data is used to perform inference about these unknown parameters.

In contrast, in the Bayesian framework it is assumed that the model parameters have some underlying distribution and the experimental data is used to gather insight into these distributions.

The aim of this chapter is to provide a compact overview of the underlying Bayesian principles and techniques that will be used in the remainder of the dissertation. The outline of the chapter is as follows:

- In Section 3.1, the two central quantities to any Bayesian analysis, the prior and the posterior distributions, are discussed.
- In Section 3.2, methods used to summarise the posterior distribution and how the posterior distribution is used for decision making, prediction and model fit are investigated.
- In Section 3.3, the major Markov chain Monte Carlo (MCMC) methods most often used to simulate the posterior distribution, namely the Gibbs sampler and the Metropolis-Hastings (M-H) algorithm, are discussed.
- In Section 3.4, two Bayesian approaches often utilised when modelling survival data, the Dirichlet process prior and the use of the Gibbs sampler with this process, are discussed.

3.1 The Bayesian approach

In this section, the two elements central to any Bayesian analysis - the prior distribution and

the posterior distribution - are discussed.

3.1.1 The prior

Let $\boldsymbol{\lambda}$ be a given vector of unknown parameters in a statistical model. Assume that $\boldsymbol{\lambda}$ follows a specified random probability distribution, $\pi(\boldsymbol{\lambda})$, the so-called prior distribution. The prior distribution quantifies our belief about the parameters of interest before the experiment is performed. This prior belief can be based on a previous experiment, such as the case would be when a phase III clinical trial follows a phase II trial (Howard *et al.*, 2005). It can be based on the experimenter's own past experience, as the case would be if a doctor has been working with the same disease for the past 20 years. In other situations, no prior information may be available and this can also be reflected in a prior distribution.

Selecting the appropriate prior for the situation at hand lies at the heart of a Bayesian analysis. Various priors have been developed and investigated for different situations in the literature and in this dissertation they are grouped into the following three types:

- non-informative,
- conjugate,
- subjective.

The background for each type of prior is briefly explained. As only the non-informative maximal data information (MDI), the Jeffreys and certain subjective priors are used in subsequent chapters, they are discussed in more detail.

3.1.1.1 Non-informative priors

Non-informative priors are used when no, or very little and vague, prior knowledge about the model parameters is available. One approach is to select a prior that is approximately uniformly distributed over the domain of the parameter space of interest. Diffuse priors, where a normal distribution with a relatively large variance is selected as a prior distribution, is such an example.

A general class of non-informative priors is derived by using the sample data and the model under study in constructing the prior distribution. Popular data or model driven methods for constructing non-informative priors are the maximal data information (MDI) prior of Zellner

(1971), the Jeffreys prior, the reference prior and probability matching priors (PMP).

– The MDI prior

The MDI prior, proposed by Zellner (1971), is derived so as to maximise the average information in the data density relative to that in the prior. Another interpretation of the MDI prior is that it maximises the expectation of the logarithm of the ratio of the likelihood function to the prior density (Zellner, 1996). From both of these interpretations it should be clear that the use of the MDI prior leads to an emphasis on the information in the data density or likelihood.

Let the joint density of an observation x and a parameter λ be $f(x,\lambda) = \pi(\lambda)f(x|\lambda)$, where $\pi(\lambda)$ is the prior density and $f(x|\lambda)$ a proper data density given λ (Zellner, 1996). Then the negative entropy of $f(x,\lambda)$ relative to an uniform measure, a measure of the information in $f(x,\lambda)$, is

$$\begin{aligned} -H &= E_{(x,\lambda)}\left[\ln f(X,\lambda)\right] \\ &= \int\int f(x,\lambda)\ln f(x,\lambda)dxd\lambda \\ &= \int\int \pi(\lambda)f(x|\lambda)\ln[\pi(\lambda)f(x|\lambda)]dxd\lambda \\ &= \int\left[\int f(x|\lambda)\ln f(x|\lambda)dx\right]\pi(\lambda)d\lambda + \int \pi(\lambda)\ln\pi(\lambda)d\lambda \\ &= \int J(\lambda)\pi(\lambda)d\lambda + \int \pi(\lambda)\ln\pi(\lambda)d\lambda, \end{aligned} \tag{3.1}$$

where $J(\lambda) = \int f(x|\lambda)\ln f(x|\lambda)dx$, is the negative entropy of $f(x|\lambda)$. Observe in (3.1) that the entropy of $f(x,\lambda)$ is divided into the sum of two integrals; the first denoting the average prior information in the distribution of the data, and the second denoting the information in the prior (Zellner, 1996). Define $G[\pi(\lambda)]$ as the difference between these two integrals (Zellner, 1996)

$$G[\pi(\lambda)] = \int J(\lambda)\pi(\lambda)d\lambda - \int \pi(\lambda)\ln\pi(\lambda)d\lambda.$$

Zellner (1996) noted that the quantity $G[\pi(\lambda)]$ gives the total information provided by an experiment over and above the prior.

The MDI prior, $\pi_{MDI}(\lambda)$, is defined as the prior distribution that maximises $G[\pi(\lambda)]$, subject to $\int \pi_{MDI}(\lambda)d\lambda = 1$. The solution is

$$\pi_{MDI}(\lambda) \propto e^{J(\lambda)},$$

and

$$G[\pi_{MDI}(\lambda)] - G[\pi(\lambda)] = \int \pi(\lambda) \ln \left[\pi(\lambda)/\pi_{MDI}(\lambda)\right] d\lambda$$

is the cross-entropy of $\pi(\lambda)$ relative to $\pi_{MDI}(\lambda)$ (Zellner, 1996).

– The Jeffreys prior

Jeffreys proposes an intrinsic approach that obviates the need to take a potential invariance structure into account, while often being compatible with it when it exists (Robert, 2001, pp. 129-133). The Jeffreys non-informative prior is based on Fisher's information given by

$$I(\lambda) = -E_\lambda \left[\frac{\partial^2 \ln f(X|\lambda)}{\partial \lambda^2}\right]$$

in the one-dimensional case and in the multi-dimensional case for $\boldsymbol{\lambda} \in \mathcal{R}^k$, $I(\boldsymbol{\lambda})$ has the following elements

$$I_{ij}(\boldsymbol{\lambda}) = -E_\lambda \left[\frac{\partial^2 \ln f(X|\boldsymbol{\lambda})}{\partial \lambda_i \partial \lambda_j}\right], \quad (i, j = 1, ..., k).$$

Jeffreys' rule defines a non-informative prior distribution, also called the Jeffreys prior, for the one-dimensional case as

$$\pi_{Jef}(\lambda) \propto I^{1/2}(\lambda),$$

defined up to a normalising coefficient when π_{Jef} is proper, and

$$\pi_{Jef}(\boldsymbol{\lambda}) \propto |I(\boldsymbol{\lambda})|^{1/2}$$

in the multi-dimensional case (Robert, 2001, pp. 129-133).

It satisfies the invariant reparametrisation requirement, since, given a one-to-one transform h, we have the (Jacobian) transformation

$$I(\lambda) = I(h(\lambda))(h'(\lambda))^2,$$

which explains the exponent $\frac{1}{2}$ (Robert, 2001, pp. 129-133).

As $I(\lambda)$ is widely accepted as an indicator of the amount of information brought by the model (or the observation(s)) about λ, it seems intuitively justified that the values of λ for which $I(\lambda)$ is larger should be more likely for the prior distribution. To favour the values of λ for which $I(\lambda)$ is large is equivalent to minimising the influence of the prior distribution and is therefore as non-informative as possible (Robert, 2001, pp. 129-133).

– The Reference prior

One drawback of Jeffreys' noninformative priors is that they do not necessarily perform satisfactorily for all inferential problems, in particular when considering subvectors of interest (Gill, 2008, pp. 152-153). Bernardo (1979) proposed a modification of the Jeffreys prior that distinguishes between parameters of interest ($\boldsymbol{\lambda}$) and nuisance parameters ($\boldsymbol{\omega}$). The proposed reference prior depends not only on the sample distribution, but also on the inferential problem at hand (Robert, 2001, pp. 133-137). When $x \sim f(x|\boldsymbol{\lambda}, \boldsymbol{\omega})$ the reference prior is obtained by first defining $\pi(\boldsymbol{\omega}|\boldsymbol{\lambda})$ as the Jeffreys prior associated with $f(x|\boldsymbol{\lambda})$ when $\boldsymbol{\lambda}$ is fixed, then deriving the marginal distribution

$$\tilde{f}(x|\boldsymbol{\lambda}) = \int f(x|\boldsymbol{\lambda}, \boldsymbol{\omega})\pi(\boldsymbol{\omega}|\boldsymbol{\lambda})d\boldsymbol{\omega},$$

and computing the Jeffreys prior $\pi(\boldsymbol{\lambda})$ associated with $\tilde{f}(x|\boldsymbol{\lambda})$. The principle behind the reference prior is to eliminate the nuisance parameter by using a Jeffreys prior where the parameter of interest remains fixed (Robert, 2001, pp. 133-137). Bernardo (1979) showed that, in situations where asymptotic normality of the posterior holds and provided there are no nuisance parameters, the reference prior for $\boldsymbol{\lambda}$ is equivalent to the Jeffreys prior

$$\pi_{Ref}(\boldsymbol{\lambda}) = \pi_{Jef}(\boldsymbol{\lambda}) \propto |I(\boldsymbol{\lambda})|^{1/2}.$$

The difficulty in using the reference prior lies in defining $\pi(\boldsymbol{\omega}|\boldsymbol{\lambda})$. Defining $\pi(\boldsymbol{\omega}|\boldsymbol{\lambda})$ as the Jeffreys prior associated with $f(x|\boldsymbol{\lambda})$ when $\boldsymbol{\lambda}$ is fixed works well when it turns out to be a proper distribution, but runs into normalisation difficulties otherwise (Berger and Bernardo, 1989). Berger and Bernardo (1989) provided a general scheme that can be used to determine the reference prior.

– The Probability Matching prior

Another approach to non-informative priors is to find priors that have good frequentist properties, that is, properties that hold on the average (in x), rather than conditional on x (Robert, 2001, pp. 137-140). One such prior commonly used is the probability matching prior. For this prior the posterior probabilities of certain specified sets are exactly or approximately equal to their coverage probabilities (Sweeting, 2005). For one-sided parametric intervals let $0 < \alpha < 1$ and suppose $\pi(\lambda)$ is a positive continuous prior on Ω. Let $\lambda^{\alpha}(\pi, data)$ denote the α-quantile of the posterior distribution of λ given the data. That is, $\lambda^{\alpha}(\pi, data)$, satisfies

$$P_{\pi}(\lambda \leq \lambda^{\alpha}(\pi, data)) = \alpha.$$

A prior, $\pi(\lambda)$, is said to be an rth-order probability matching prior if it satisfies the probability matching constraint

$$P_{\lambda}(\lambda \leq \lambda^{\alpha}(\pi, data)) = \alpha + O(n^{-r/2})$$

pointwise or very weakly for every α, $0 < \alpha < 1$ (Data and Sweeting, 2005). Welch and Peers (1963) (cited in Data and Sweeting, 2005) investigated the second-order approximation and showed that this relationship holds to $O(n^{-1})$ pointwise for all α, if and only if,

$$\pi(\lambda) \propto I^{1/2}(\lambda).$$

Therefore Jeffreys' invariant prior is a second order probability matching prior with respect to one-sided parametric regions (Data and Sweeting, 2005). See Sweeting (2005), Data and Sweeting (2005) and Data and Mukerjee (2004) for more on matching and probability matching priors, including two-sided parametric intervals and matching prior in multiparameter cases.

For more on non-informative priors and for examples on how to calculate the MDI, Jeffreys, reference and matching priors for different data examples see Martz and Waller (1982, pp. 223-226), Berger and Bernardo (1989), Box and Tiao (1992, pp. 25-60), Kass and Wasserman (1996), Data and Mukerjee (2004) and Jaynes (1980).

3.1.1.2 Conjugate priors

A characteristic of a conjugate prior distribution for a given sampling distribution $f(x|\boldsymbol{\lambda})$ is

that the prior distribution, $\pi(\boldsymbol{\lambda})$, and the posterior distribution, $\pi(\boldsymbol{\lambda}|x)$, are from the same family of distributions (Hamada *et al.*, 2008, p. 47). Robert (2001, pp. 113-123) notes that the conjugate prior approach can be partly justified through the invariance reasoning. When the observation $x \sim f(z|\boldsymbol{\lambda})$ modifies $\pi(\boldsymbol{\lambda})$ into $\pi(\boldsymbol{\lambda}|x)$, the information conveyed by x about $\boldsymbol{\lambda}$ is obviously limited; therefore, it should not lead to a modification of the whole structure of $\pi(\boldsymbol{\lambda})$, but simply of its parameters (Robert, 2001, pp. 113-123). Natural conjugate priors for some common exponential families are given in Table 3.1 (Robert, 2001, pp. 113-123).

Table 3.1: Natural conjugate priors for common exponential families.

$f(x\|\boldsymbol{\lambda})$	$\pi(\boldsymbol{\lambda})$	$\pi(\boldsymbol{\lambda}\|x)$
Normal $N(\lambda, \sigma^2)$	Normal $N(\mu, \tau^2)$	$N(\frac{\sigma^2\mu+\tau^2 x}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$
Binomial $Bi(n, \lambda)$	Beta $Be(\alpha, \beta)$	$Be(\alpha + x, \beta + n - x)$
Multinomiaal $M_k(\lambda_1, ..., \lambda_k)$	Dirichlet $D(\alpha_1, ..., \alpha_k)$	$D(\alpha_1 + x_1, ..., \alpha_k + x_k)$
Gamma $G(\nu, \lambda)$	Gamma $G(\alpha, \beta)$	$G(\alpha + \nu, \beta + x)$

See Martz and Waller (1982, pp. 226-229) and Raiffa and Schaifer (1961, pp. 43-58) for more on the fundamentals of conjugate priors and the methodology for obtaining conjugate priors for different data examples.

3.1.1.3 Subjective priors

Subjective priors are based on past experience and on what a researcher thinks are likely quantities for the parameters of interest. Information about constructing a subjective prior may be obtained from a previous study or published work, a researcher's own knowledge or from interviewing experts in the specific area of research under study.

In some research studies no information or data is available from a similar previous study, but the researchers have access to individuals with knowledge on the subject of the investigation. Here the knowledge of the experts in the field needs to be transformed into a suitable prior that can be used in a Bayesian analysis. This process leads to the use of elicited priors that are based on the knowledge of individuals rather than information from previous studies. In Chapter 5 elicited priors will be used when imputing multi-state panel data (see Sections 5.1 and 5.2.2). See Gill (2008, pp. 159-174) and Spiegelhalter *et al.* (2004, Chapter 5) for more on

elicited priors and for examples on how these priors are obtained for different clinical studies.

3.1.2 The posterior

After observing the data and calculating the likelihood, the posterior distribution of λ, $\pi(\lambda|data)$, is obtained by using Bayes' theorem

$$\pi(\boldsymbol{\lambda}|data) = \frac{L(\boldsymbol{\lambda}|data)\pi(\boldsymbol{\lambda})}{\int_{\boldsymbol{\Theta}} L(\boldsymbol{\lambda}|data)\pi(\boldsymbol{\lambda})d\boldsymbol{\lambda}}. \tag{3.2}$$

It is clear from (3.2) that $\pi(\boldsymbol{\lambda}|data)$ is proportional to the likelihood multiplied by the prior,

$$\pi(\boldsymbol{\lambda}|data) \propto L(\boldsymbol{\lambda}|data)\pi(\boldsymbol{\lambda}),$$

indicating that our prior belief of the parameters, quantified by the $\pi(\boldsymbol{\lambda})$, and the effect of the observed data on the parameters, quantified by the likelihood $L(\boldsymbol{\lambda}|data)$, both play a role in determining the posterior distribution of the parameters (Robert, 2001, pp. 22-26).
The quantity $m(data) = \int_{\boldsymbol{\Theta}} L(\boldsymbol{\lambda}|data)\pi(\boldsymbol{\lambda})d\boldsymbol{\lambda}$ is the normalising constant of $\pi(\boldsymbol{\lambda}|data)$, and is often called the marginal distribution of the data or the prior predictive distribution (Ibrahim *et al.*, 2001, pp. 17-18). In addition to ensuring that the posterior distribution is a properly defined density integrating to one, the prior predictive distribution also plays a role in model comparison problems.
In most applications of Bayes' theorem it is not possible to write (3.2) in closed form and simulation techniques need to be used to obtain the posterior distribution. The most common techniques used to draw samples from the posterior distribution are discussed in Section 3.3.

3.2 Posterior analysis

The posterior distribution is central to any Bayesian analysis and inference, but having obtained the posterior distribution is not the end of a Bayesian analysis. The posterior can now be further investigated to help make decisions about the problems or questions under study and also to assess the fit and appropriateness of the models used.

3.2.1 Summarising the posterior

Basic numerical point summaries of the posterior, such as the mean, median, mode, standard deviation and quantiles, provide useful insight into the posterior distribution and the behaviour of the parameters of interest.

In addition to point summaries, interval summaries are also used to describe the posterior distribution. Two of the most popular Bayesian intervals used in posterior analysis, and that will be used in this dissertation, are:

- The Bayesian credible or probability interval: A $100(1-\alpha)\%$ credible interval is any set C such that the probability under the posterior distribution, $\pi(\boldsymbol{\lambda}|data)$, is equal to $(1-\alpha)$,

$$1-\alpha = \int_C \pi(\boldsymbol{\lambda}|data)d\boldsymbol{\lambda}.$$

  The set C is usually selected from the percentiles of the posterior distribution, but since they are not unique, various credible intervals can be calculated for a specific posterior distribution (Gill, 2008, pp. 45-46).

- The highest posterior density regions (HPD): These regions are unique and are such that no other region outside of the interval will have higher posterior density than any region inside the HPD (Gill, 2008, pp. 48-50). A $100(1-\alpha)\%$ HPD is the subset of support of the posterior distribution for some parameter, $\boldsymbol{\lambda}$, that meets the criteria:

$$C = \{\boldsymbol{\lambda} : \pi(\boldsymbol{\lambda}|data) \geq k\},$$

  where k is the largest number such that

$$1-\alpha = \int_{\boldsymbol{\lambda}:\pi(\lambda|data)>k} \pi(\boldsymbol{\lambda}|data)d\boldsymbol{\lambda}$$

  (Casella and Berger, 2001, p. 448). See Hyndman (1996) for a general introduction to HPD regions.

3.2.2 Prediction

Prediction is an important part of most modelling problems. In the Bayesian context the posterior predictive distribution is used for future observations. For a new observation vector $\mathbf{z}$, with sampling distribution $f(\mathbf{z}|\boldsymbol{\lambda}, data)$, the posterior predictive distribution is defined as

(Ibrahim *et al.*, 2001, pp. 17-18):

$$\pi(\mathbf{z}|data) = \int_{\boldsymbol{\Theta}} f(\mathbf{z}|\boldsymbol{\lambda},data)\pi(\boldsymbol{\lambda}|data)d\boldsymbol{\lambda}, \tag{3.3}$$

which is the posterior expectation of $f(\mathbf{z}|\lambda,data)$,

$$E_{\pi(\boldsymbol{\lambda}|data)}[f(\mathbf{z}|\boldsymbol{\lambda},data)].$$

Once the posterior distribution, $\pi(\boldsymbol{\lambda}|data)$, is known, (3.3) can be used to make a prediction for the new observation $\mathbf{z}$. Using MCMC methods to sample values from the posterior distribution (see Section 3.3), the posterior predictive distribution can be calculated without the need to directly solve (3.3). This is done by in turn using each sampled posterior value to calculate the predicted value of $\mathbf{z}$ and then calculating the mean across all the realisations of the posterior. The predictive distribution is used in Section 5.5 to predict the future state of patients in a Bayesian multi-state model.

3.2.3 Decision making

Decision theory concerns itself with the situation in which decision makers have to make a choice from a given set of available actions $\Delta = (\delta_1, ..., \delta_p)$ and where the loss of a given action depends upon a state of nature λ, which is unknown. In Bayesian decision theory, the decision maker combines prior knowledge of λ and stochastic information provided by an experiment, the data, in the form of a posterior distribution of λ and then chooses the decision (action) that minimises the expected loss over the posterior (Martz and Waller, 1982, pp. 190-212). This decision can take the form of an estimate that minimises a given loss function or of choosing a specific hypothesis in favour of another. The following concepts play an important role in Bayesian decision theory (See De Groot (1970), Robert (2001, Chapter 2) and Martz and Waller (1982, pp. 190-212) for a more in depth discussion on Bayesian decision making.):

- Loss function

  The loss function, $Lo(\lambda, \delta)$, is a real-valued function that quantifies the loss (or penalty) for decision δ given λ is the true parameter value. The actual determination of the loss function is often awkward in practice, as the determination of the consequences of each action for each value of λ is impossible when λ or Δ are large sets (Robert, 2001, pp. 60-65).

  Due to its symmetrical nature and the fact that under this loss the Bayes estimator is

tractable for many recognisable distributions, the squared error loss function

$$Lo(\lambda, \delta) = (\lambda - \delta)^2$$

is the most popular loss function for Bayes procedures (Moyé, 2008, pp. 208-212).

The so called $0-1$ loss function

$$Lo(\lambda, \delta) = I(\lambda = \delta),$$

with $I(.)$ the indicator function, is mainly used in the classical approach to hypothesis testing and is an example of a nonquantitative loss; the loss associated with δ is 0 if correct and 1 otherwise (Robert, 2001, pp. 80-81).

– Posterior expected loss

Given a prior and a corresponding posterior distribution, $\pi(\lambda)$ and $\pi(\lambda|data)$, the posterior expected loss for a given decision, δ, and loss function, $Lo(\lambda, \delta)$, is defined as

$$\begin{aligned} \rho(\pi, \delta|data) &= E_\pi \left[Lo(\lambda, \delta)|data\right] \\ &= \int_\Theta Lo(\lambda, \delta)\pi(\lambda|data)d\lambda. \end{aligned}$$

This is the average loss for decision δ according to the posterior distribution, conditional on the observed data (Robert, 2001, pp. 60-65).

– Bayes estimator

The Bayes estimator, $\hat{\lambda}_B$, associated with a given loss function is the estimator that minimises the expected posterior loss

$$\hat{\lambda}_B = \arg\min_\delta \rho(\pi, \delta|data).$$

Under the squared error loss function, $\hat{\lambda}_B$, is defined as

$$\hat{\lambda}_B = \arg\min_\delta E_\pi \left[(\lambda - \delta)^2|data\right].$$

This minimum is attained when $\hat{\lambda}_B$ equals the posterior mean,

$$\hat{\lambda}_B = E_\pi[\lambda|data] = \int_\Theta \lambda\pi(\lambda|data)d\lambda = \bar{\lambda}_{Post}.$$

It can be shown that under the absolute error loss function

$$L(\lambda, \delta) = |\lambda - \delta|,$$

the posterior median is the Bayes estimator and under the $0-1$ loss function the Bayes estimator is the posterior mode (Robert, 2001, pp. 60-65).

3.2.4 Model fit

Markov chain Monte Carlo (MCMC) simulation methods (see Section 3.3) have made it possible to fit complex Bayesian models to data sets. The fact that it is now possible to fit these complex models and to obtain posterior distributions for these models, does not mean that these models are necessarily appropriate for the situations at hand. In this section, the Bayes factor and information criterion, two of the most widely used methods to assess the fit of a Bayesian model, are discussed.

3.2.4.1 Bayes factor

The Bayes factor combines prior and posterior information in a ratio that provides evidence of one model specification over another. Suppose data is observed and interest lies in comparing two models, M_1 and M_2, each with its own set of parameters λ_1 and λ_2. Let $\pi_1(\lambda_1)$ and $\pi_2(\lambda_2)$ be the prior distributions for the parameter vectors λ_1 and λ_2, and $\pi(M_1)$ and $\pi(M_2)$ be the prior probability of the two models. Using Bayes' theorem the posterior odds or ratio of model 1 versus model 2 is (Bernardo and Smith, 1994, pp. 389-395):

$$\begin{aligned} \frac{\pi(M_1|data)}{\pi(M_2|data)} &= \frac{\pi(M_1)/\pi(data)}{\pi(M_2)/\pi(data)} \times \frac{\int_{\boldsymbol{\lambda}_1} L_1(\boldsymbol{\lambda}_1|data)\pi_1(\boldsymbol{\lambda}_1)d\boldsymbol{\lambda}_1}{\int_{\boldsymbol{\lambda}_2} L_2(\boldsymbol{\lambda}_2|data)\pi_2(\boldsymbol{\lambda}_2)d\boldsymbol{\lambda}_2}, \\ \text{Posterior Odds} &= \text{Prior Odds} \times \text{Bayes Factor}. \end{aligned} \tag{3.4}$$

Rearranging (3.4) gives the standard form of the Bayes factor as:

$$B(data) = \frac{\pi(M_1|data)/\pi(M_1)}{\pi(M_2|data)/\pi(M_2)}.$$

If equal prior probability is placed on the two models and the models share the same parameter space but at different hypothesised levels, the Bayes factor reduces to the common likelihood ratio. See Bernardo and Smith (1994, pp. 389-395) and Robert (2001, pp. 227-229) for a detailed discussion of the Bayes factor.

3.2.4.2 Information criteria

Due to the fact that the Bayes factor is often difficult or impossible to calculate, especially for models with large numbers of unknowns or improper priors, an alternative approach is to adopt an approximation to the Bayes factor. This is done by using information criteria.

– The Bayesian information criterion (BIC), also known as the Schwarz criterion, is given by:

$$BIC = -2\log[L(\boldsymbol{\lambda}|data)] + p\log(n)$$

where $L(\lambda|data)$ is the maximised likelihood value, p the number of explanatory variables in the model and n is the sample size (Ibrahim *et al.*, 2001, pp. 246-254).

– Spiegelhalter *et al.* (2002) introduced the deviance information criterion (DIC) as a model assessment and comparison tool. It is a Bayesian alternative to Akaike's information criterion (AIC) and it is defined as a Bayesian measure of fit or adequacy, penalised by an additional complexity term. The posterior mean difference,

$$\overline{D(\boldsymbol{\lambda})} = E_{\boldsymbol{\lambda}}\left[-2\log L(\boldsymbol{\lambda}|data)|data\right] + 2\log\left[f(data)\right]$$

is used as a measure of Bayesian model fit, with $f(data)$, the standardising term, some function of just the data. The effective number of parameters in the model, p_D, is defined as

$$p_D = \overline{D(\boldsymbol{\lambda})} - D(\bar{\boldsymbol{\lambda}}).$$

This is the "mean deviance minus the deviance of the means", with the deviance, $D(\bar{\lambda})$, defined as

$$D(\bar{\boldsymbol{\lambda}}) = -2\log\left[L(E(\boldsymbol{\lambda}|data)|data)\right] + 2\log\left[f(data)\right].$$

Combining the estimate of Bayesian fit with the effective number of parameters gives the DIC as (Spiegelhalter *et al.*, 2002):

$$\begin{aligned} DIC &= \overline{D(\boldsymbol{\lambda})} + p_D \\ &= 2\overline{D(\boldsymbol{\lambda})} - D(\bar{\boldsymbol{\lambda}}). \end{aligned}$$

Unlike the BIC that requires a maximisation over the parameter space, the DIC can easily be calculated from a MCMC run, as $\overline{D(\boldsymbol{\lambda})}$ can be approximated by taking the mean of the

simulated values of $D(\boldsymbol{\lambda})$, and $D(\bar{\boldsymbol{\lambda}})$ can be approximated by the plug-in estimate of the deviance using the means of the simulated values of $\boldsymbol{\lambda}$ (Spiegelhalter *et al.,* 2002). The DIC is used in Section 5.5 to assess the fit of the multi-state models to panel data sets, generated by the proposed Bayesian multi-state imputation techniques.

3.3 Sampling from the posterior

The posterior distribution lies at the heart of any Bayesian analysis. Unfortunately, although (3.2) and (3.3) provide elegant and straightforward equations to obtain the posterior and posterior predictive distributions, in general $m(data)$, and thus by extension $\pi(\lambda|data)$ and $\pi(\mathbf{z}|data)$, do not have analytical closed forms. To overcome this problem numerous computer intensive simulation techniques have been developed over the years to aid sampling from the posterior distribution without having to know $m(data)$. The most common techniques used recently are referred to as Markov chain Monte Carlo (MCMC) techniques. These are techniques where the results are based on repeated sampling from a Markov chain with limiting distribution equal to the posterior distribution or target distribution of interest.
Numerous MCMC techniques have been developed. The Gibbs sampler and the Metropolis-Hastings (M-H) algorithm, being two of the most widely used, will be discussed here. The M-H algorithm is utilised in Chapter 4 (see Sections 4.1 and 4.2) to generate posterior distributions for the proposed Bayesian multi-state models.

3.3.1 The Gibbs sampler

The Gibbs sampler is a technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density itself (Casella and George, 1992). At its heart, the Gibbs sampler takes advantage of hierarchical structures that exist in the distribution of interest or target distribution. For example, assume $\pi(\boldsymbol{\lambda}|x)$, the target distribution, can be written as:

$$\pi(\boldsymbol{\lambda}|x) = \int \pi_1(\boldsymbol{\lambda}|x, \boldsymbol{\theta})\pi_2(\boldsymbol{\theta}|x)d\boldsymbol{\theta}.$$

The idea is then to simulate from the joint distribution $\pi_1(\boldsymbol{\lambda}|x, \boldsymbol{\theta})\pi_2(\boldsymbol{\theta}|x)$ to recover $\pi(\boldsymbol{\lambda}|x)$ as the marginal distribution. If both distributions $\pi_1(\boldsymbol{\lambda}|x, \boldsymbol{\theta})$ and $\pi_2(\boldsymbol{\theta}|x)$ are known and can be

sampled from, the generation of $\boldsymbol{\lambda}$ from $\pi(\boldsymbol{\lambda}|x)$ is equivalent to generating $\boldsymbol{\theta}$ from $\pi_2(\boldsymbol{\theta}|x)$ and then $\boldsymbol{\lambda}$ from $\pi_1(\boldsymbol{\lambda}|x, \boldsymbol{\theta})$ iteratively (Robert, 2001, pp. 307-309).

In general assume that the target distribution is $\pi(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_d)'$ and that the full conditional distributions $\pi(\lambda_i) = \pi(\lambda_i|\boldsymbol{\lambda}_{-i}), i = 1, ..., d$ are available. The steps in the Gibbs sampler to generate a draw from $\pi(\boldsymbol{\lambda})$ are (Ibrahim *et al.*, 2001, pp. 18-22):

1) Generate an arbitrary starting point $\boldsymbol{\lambda}^{(0)} = (\lambda_1^{(0)}, ..., \lambda_d^{(0)})'$ and initialise the counter of the chain $i = 0$.
2) Obtain a new value $\boldsymbol{\lambda}^{(i+1)} = (\lambda_1^{(i+1)}, ..., \lambda_d^{(i+1)})'$ from $\boldsymbol{\lambda}^{(i)}$ as follows:
   - Generate $\lambda_1^{(i+1)} \sim \pi(\lambda_1|\lambda_2^{(i)}, ..., \lambda_d^{(i)})$
   - Generate $\lambda_2^{(i+1)} \sim \pi(\lambda_2|\lambda_1^{(i)}, \lambda_3^{(i)}, ..., \lambda_d^{(i)})$

     . . . . . .
   - Generate $\lambda_d^{(i+1)} \sim \pi(\lambda_d|\lambda_1^{(i)}, \lambda_2^{(i)}, ..., \lambda_{d-1}^{(i)})$
3) Set $i = i + 1$ and return to 2.

When convergence is reached, and after a suitable burn-in period, the resulting value $\boldsymbol{\lambda}^{(j)}$ is a draw from $\pi(\boldsymbol{\lambda})$ (Casella and George, 1992).

After reaching convergence it is important to assess if the Gibbs sampler has in fact converged to the correct distribution. To this end Zellner and Min (1995) introduced three convergences criteria for the Gibbs sampler:

- The anchored ratio convergence criterion (ARC^2)
- The difference convergence criterion (DC^2)
- The ratio convergence criterion (RC^2)

These easily evaluated criteria detect convergence failures and also convergences to incorrect values (Zellner and Min, 1995).

3.3.2 Metropolis-Hastings sampling

As in the case of the Gibbs sampler, M-H sampling is a method of obtaining a sample from a target distribution, $\pi(\boldsymbol{\lambda})$, without having to directly sample from the distribution. Whereas the Gibbs sampler relies on finding full conditional distributions, the M-H sampler relies on a candidate distribution to provide candidate values of $\pi(\boldsymbol{\lambda})$ that are accepted or rejected using a specific probability mechanism. It is this probability mechanism that lies at the heart of

the M-H sampler and that ensures that the candidate values will converge to $\pi(\boldsymbol{\lambda})$ (Chib and Greenberg, 1995).

Let $f(\boldsymbol{\lambda}^*|\boldsymbol{\lambda})$ be a candidate-generating density with $\int f(\boldsymbol{\lambda}^*|\boldsymbol{\lambda})d\boldsymbol{\lambda}^* = 1$, then the M-H algorithm can be summarised as (Ibrahim *et al.*, 2001, pp. 18-22):

1) Choose an arbitrary starting point $\boldsymbol{\lambda}^{(0)}$ and initialise the counter of the chain $i = 0$.
2) Generate a candidate point $\boldsymbol{\lambda}^*$ from $f(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^{(i)})$ and u from $unif(0, 1)$.
3) Set $\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^*$ if $u \leq \alpha(\boldsymbol{\lambda}^*, \boldsymbol{\lambda}^{(i)})$ else set $\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)}$, with $\alpha(\boldsymbol{\lambda}^*, \boldsymbol{\lambda}^{(i)})$ defined as

$$\alpha(\boldsymbol{\lambda}^*, \boldsymbol{\lambda}^{(i)}) = \min\left\{\frac{\pi(\boldsymbol{\lambda}^*)f(\boldsymbol{\lambda}^{(i)}|\boldsymbol{\lambda}^*)}{\pi(\boldsymbol{\lambda}^{(i)})f(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^{(i)})}, 1\right\}.$$

4) Set $i = i + 1$ and return to 2.

After a suitable burn-in period, i.e. the chain has passed the transient stage and the effect of $\boldsymbol{\lambda}^{(0)}$ on the generated values is negligible, $\boldsymbol{\lambda}^{(i)}$ can be regarded as a draw from $\pi(\boldsymbol{\lambda})$. This algorithm is utilised in Chapter 4 (see Sections 4.1 and 4.2) to generate posterior distributions for the proposed Bayesian multi-state models.

3.4 Process priors: Bayesian survival analysis

In general when fitting a survival model, the interest is in the survival or hazard functions. More often than not these functions are modelled non-parametrically and no parametric assumptions are made about their form or shape. Since no parametric assumptions are made, there are no parameters of interest, and thus the Bayesian approach of placing a prior distribution on a parameter of interest is no longer valid. The "parameter" of interest, the survival or hazard function, is a realisation of a stochastic process and as such a prior has to be placed on this stochastic process rather than just on a parameter in a model. Process priors, so called because they are placed over the stochastic process of interest, have developed into the priors of choice when fitting Bayesian survival models (Sinha and Dey, 1997).

In this section, the Dirichlet process prior and its associated Monte Carlo Bayesian method or Gibbs sampler - both techniques that will be employed in Chapter 5 to develop Bayesian multi-state imputation methods - are discussed.

3.4.1 Dirichlet process prior

Assume that $G(t) = P(Y > t)$, the survival function, is sampled from a Dirichlet process (DP) with a parameter function (also known as the base function) $F(t)$ and weight parameter γ, $G(t) \sim DP(\gamma, F)$.

A Dirichlet process, defined on the positive real line, has the property that for any set of intervals $A_1, ..., A_k$, which partitions the positive real line, the joint distribution of the prior probabilities $P(Y \in A_1) = W_1, ..., P(Y \in A_k) = W_k$ has a k dimensional Dirichlet distribution with parameters $[\gamma F(A_1), ..., \gamma F(A_k)]$. This property must hold for any such set of intervals and any k (Susarla and Van Ryzin, 1976).

A vector $(W_1, ..., W_k)$ has a $(k-1)$-dimensional Dirichlet distribution with parameters $(\alpha_1, ..., \alpha_k)$ if $W_i = \frac{Z_i}{\sum_{i=1}^{k} Z_i}$, where each z_i are independent gamma random variables with shape parameter α_i and scale parameter 1. The joint density function of $(W_1, ..., W_{k-1})$ is given by (Klein and Moeschberger, 2003, pp. 187-198)

$$f(w_1, ..., w_{k-1}) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \left[\prod_{i=1}^{k-1} w_i^{\alpha_i - 1}\right] \left[1 - \sum_{i=1}^{k-1} w_i\right]^{\alpha_k - 1},$$

with $\alpha = \sum_{i=1}^{k} \alpha_i$.

To assign a prior distribution to the survival function, assume that $G(t)$ follows a Dirichlet process with parameter function $F(t)$ and weight γ. The parameter function is typically written as $F(t) = \gamma F_0(t)$, where $F_0(t)$ is the prior estimate of the survival function and γ is a weight parameter that gives prior weight to $F_0(t)$. The mean and variance of this prior distribution is given by (Klein and Moeschberger, 2003, pp. 187-198)

$$\begin{aligned} E[G(t)] &= F_0(t) \text{ and} \\ var[G(t)] &= \frac{F_0(t)[1 - F_0(t)]}{\gamma + 1}. \end{aligned}$$

Ferguson (1973) showed that if $G \sim DP(\gamma, F)$ and $x_1, ..., x_n$ is a sample from G, the posterior can be written as

$$G|data \sim DP(\gamma, F + \sum_{i=1}^{n} \delta_{x_i}),$$

with δ_{x_i} a point mass giving probability 1 to x_i.That is to say that the posterior distribution

will also be a Dirichlet process with parameter the original parameter F, plus a point mass of one where events occur.

Using this posterior distribution, Susarla and Van Ryzin (1976) derived the Bayes estimator of the survival function under squared error loss as

$$\hat{G}_B(t) = \frac{\gamma F_0(t) + X_{l+1}}{\gamma F_0(0) + n} \prod_{k=1}^{l} \frac{\gamma F_0(t_k) + X_{k+1} + \eta_k}{\gamma F_0(t_k) + X_{k+1}} \tag{3.5}$$

for $t_l \leq t < t_{l+1}, l = 0, ..., m$, and X_l and η_l the number of individuals at risk and the number of censored observations respectively at time l.

In Chapter 5 a Dirichlet process prior will be assumed for the transition between two known states in a multi-state model (see Section 5.2).

3.4.2 Gibbs sampler in the Dirichlet process

Unlike the Dirichlet process prior, where closed form Bayesian estimators of the posterior survival function are calculated, the Gibbs sampling approach is used to directly simulate the posterior survival function (Klein and Moeschberger, 2003, pp. 187-198).

Let $t_0 = 0 < t_1 < ... < t_m$ be m distinct time points, d_j be the number of events in the interval $(t_{j-1}, t_j]$ and r_j the number of right-censored observations at time t_j. Let $G(t_j)$ be the survival function at time t_j, so the likelihood function is proportional to $\prod_{j=1}^{m} (G(t_{j-1}) - G(t_j))^{d_j} G(t_j)^{r_j}$.

Let $\theta_j = G(t_{j-1}) - G(t_j)$, be the probability of an event occurring in the interval $(t_{j-1}, t_j]$, $j = 1, ..., m$ and $\theta_{m+1} = G(t_m)$. For a prior distribution, assume that the joint distribution of $\boldsymbol{\theta}$ is the Dirichlet distribution with density function

$$\pi(\theta_1, ..., \theta_m) = \frac{\Gamma(\gamma)}{\prod_{j=1}^{m+1} \Gamma(\alpha_j)} \prod_{j=1}^{m+1} (\theta_j)^{\alpha_j - 1},$$

where $\alpha_j = \gamma[F_0(t_{j-1}) - F_0(t_j)]$ for $j = 1, ..., (m+1)$ with $F_0(t_0) = 1$, $F_0(t_{m+1}) = 0$, $F_0(t)$ the prior estimate of the survival function and γ a weight parameter that gives prior weight to $F_0(t)$.

The Gibbs sampler approximates the posterior distribution via a Monte Carlo simulation (Klein and Moeschberger, 2003, pp. 187-198). The censored observations are treated as unknown

parameters and the event time for each observation is simulated. These simulated event times are then combined with the known event times to simulate the parameters θ_j, $j = 1, ..., m$. The new $\boldsymbol{\theta}'s$ are used to simulate the new event times for the censored observations, and so forth. Gelfand and Smith (1990) showed that this procedure converges to a realisation of $\boldsymbol{\theta}$ drawn from the posterior distribution of $\boldsymbol{\theta}$, given the data. This process is repeated a large number of times to obtain a sample from the posterior distribution of $\boldsymbol{\theta}$ given the data.

For censored data, the Gibbs sample is generated as follows (Klein and Moeschberger, 2003, pp. 187-198):

1) Create the time intervals $(t_{j-1}, t_j]$, $j = 1, ..., (m+1)$, with $t_0 = 0$ and $t_{m+1} = \infty$, ensuring a point mass at t_j where $d_j > 0$, by creating an interval $(t_j - \delta t, t_j]$ if $d_j > 0$. This is to ensure that $\widehat{G}(t)$ has jumps at each of the event times.
2) Calculate the number of events and censored observations, d_j and r_j, for each interval.
3) Calculate initial values of α_j as $\alpha_j = \gamma[F_0(t_{j-1}) - F_0(t_j)]$ for $j = 1, ..., (m+1)$ with $F_0(t_0) = 1$, $F_0(t_{m+1}) = 0$ and $F_0(t)$ given and assumed to be the prior estimate of the survival function.
4) Generate θ_j^1 from a Dirichlet distribution with parameters $(\alpha_1, ..., \alpha_{m+1})$.
5) Initialise the counter $i = 1$.
6) At every $r_j > 0$, let $Z^i_{j+1,j}, ..., Z^i_{m+1,j}$ denote the number of observations out of the r_j that may have been events in the intervals $(t_j, t_{j+1}], ..., (t_{m-1}, t_m], (t_m, \infty)$, respectively, with $\sum_{k=j+1}^{m+1} Z^i_{k,,j} = r_j$, by generating $Z^i_{j+1,j}, ...Z^i_{m+1,j}$ from a multinomial distribution with sample size r_j and probabilities

$$\rho_k = \frac{\theta_k^i}{\sum_{h=j+1}^{m+1} \theta_h^i}, k = j+1, ..., m+1.$$

Repeat this step for all $r_j > 0$.

7) Generate a new set of $\boldsymbol{\theta}^{i+1\prime}s$ by computing

$$R_h^i = \alpha_h + d_h + \sum_{j=1}^{m} Z^i_{h,j}, h = 1, ..., m+1$$

and, then, sampling $\boldsymbol{\theta}^{i+1} = (\theta_1^{i+1}, \theta_2^{i+1}, ..., \theta_{m+1}^{i+1})$ from a Dirichlet distribution with parameters $(R_1^i, R_2^i, ..., R_{m+1}^i)$.

8) Repeat steps 6 and 7 i times, typically i is relatively small, 10 or 20, to yield a single

realisation of $\boldsymbol{\theta}^s$ and $\mathbf{R}^s$.

9) Repeat steps 4 to 8 L times, with L typically in the order of 1000 or 10000. The posterior estimate of θ_h is given by

$$\widetilde{\theta}_h = \frac{1}{L}\sum_{l=1}^{L}\frac{R_h^l}{\sum_{k=1}^{m+1}R_k^l}. \tag{3.6}$$

The simulation process described forms the building blocks for the process that will be used in Chapter 5 to estimate the unknown transition point between two known multi-state observations. In that setting the point at which the transition occurs is regarded as the event of a censored observation and it will be simulated in Section 5.2.

3.5 Conclusion

In this chapter the general theory of Bayesian modelling was introduced and discussed. The aim of this is to provide an overview of the underlying Bayesian principles and techniques that will be used in the remainder of this dissertation.

In Section 3.1, the two central quantities to any Bayesian analysis, the prior and the posterior distributions, were discussed. The MDI, Jeffreys' and subjective priors that are used in Chapters 4 and 5 were presented.

In Section 3.2, the methods used to summarise the posterior distribution and how the posterior distribution is used for decision making, prediction and model fit were given. The Bayesian intervals, prediction distribution and model fit criteria introduced in this section will be used in Chapters 4 and 5 to investigate the posterior distribution of the proposed Bayesian multi-state models and to assess the fit of these models to some practical examples.

In Section 3.3, the MCMC methods most often used to simulate the posterior distribution, namely the Gibbs sampler and the Metropolis-Hastings algorithm, were discussed. The Metropolis-Hasting algorithm will be used in Chapter 4 to generate samples from the posterior distributions of the proposed Bayesian multi-state models.

In Section 3.4, two Bayesian approaches often utilised when modelling survival data, the Dirichlet process prior and the Gibbs sampler in that setting, were discussed. The techniques behind both of these approaches form the building blocks of the Bayesian multi-state imputation methods that are developed in Chapter 5.

4

Bayesian Multi-State Models

In Chapter 2, multi-state models (MSM's) were introduced as the modelling tool of choice to model panel data and the disease progression of patients over time. As shown, these models can be used to calculate the transition rates between various states of a disease. These can in turn be used to assess the effectiveness of treatment on the natural progression of a disease. Unfortunately it is extremely rare to have a medical study in which all the patients are followed-up for the required number of observations or time periods. This means that in most studies, especially those that have long follow-up periods, there will be missing observations in the data. Depending on the number of missing observations, this can have serious consequences for a multi-state model and the estimated transition rates as was shown in Section 2.3.3.

The Bayesian methodology, introduced in Chapter 3, is used when prior information is available for the parameters of interest. This prior information is incorporated into the model building process to supplement or enhance the observed data and to obtain better estimates for parameters under study. This is especially useful in cases were there is limited data and where experts in the specific field of research can provide prior information about the processes or parameters of interest.

In this chapter, Bayesian MSM's (B-MSM's) will be discussed that use available prior information to augment the information from panel data. By augmenting the information available in the data, it is hoped that more accurate estimates of the transition rates can be provided. Two Bayesian multi-state approaches to estimate the transition rates will be discussed in this chapter:

- Firstly, in Section 4.1 a B-MSM will be introduced with the likelihood expressed in terms of the limiting probabilities of a Markov process. These time independent limiting probabilities are then used to estimate the transition rates.
- Secondly, in Section 4.2 a B-MSM will be discussed where the likelihood is expressed in terms

of the transition probabilities. The time dependent panel data is then used to estimate the transition rates.

The chapter concludes with Section 4.3 where a simulation study is performed to assess how the two proposed models fare when they are used to model multi-state data sets for various models.

4.1 Estimating the transition rates using the limiting probabilities in the likelihood

The limiting probabilities of a MSM are useful in summarising the long-term expected states of a multi-state process (see Section 2.1.1). They are calculated independently of time and are only dependent on the transition rates of the process. Due to being time independent, they are relatively simple to calculate and interpret; but due to the fact that the time of the transitions is not used in their calculation they can only be viewed as a summary statistic of the process. Although the limiting probabilities are only summary statistics of the Markov process, they are a useful starting point for developing a Bayesian MSM for the transition rates.

In this section a Bayesian model is proposed where non-informative priors are placed on the limiting probabilities of a 3-state Markov process. The likelihood and the posterior distributions are developed in terms of the limiting probabilities and the mathematical relationship between the transition rates and the limiting probabilities are used to express the limiting probabilities in terms of the transition rates.

Assume a 3-state Markov model with limiting probabilities (p_1, p_2, p_3) and transition intensity matrix

$$Q = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ \lambda_{21} & -(\lambda_{21}+\lambda_{23}) & \lambda_{23} \\ 0 & \lambda_{32} & -\lambda_{32} \end{bmatrix}, \tag{4.1}$$

with 4 possible transition, $\lambda_{12}, \lambda_{21}, \lambda_{23}$, and λ_{32}. The limiting probabilities of a process are solved by equating the rate at which people leave a state to the rate that people enter a state

(Ross, 2003, p. 370),

$$\begin{array}{rcl} \text{Leave state 1} & = & \text{Enter state 1} \\ \text{Leave state 2} & = & \text{Enter state 2} \\ \text{Leave state 3} & = & \text{Enter state 3} \end{array}.$$

Expressing these in terms of the transition rates, yield

$$\begin{array}{rcl} \lambda_{12}p_1 & = & \lambda_{21}p_2 \\ (\lambda_{21}+\lambda_{23})p_2 & = & \lambda_{12}p_1+\lambda_{32}p_3 \\ \lambda_{32}p_3 & = & \lambda_{23}p_2 \end{array}, \tag{4.2}$$

and, as the limiting probabilities are collectively exhaustive, it must also hold that

$$p_1+p_2+p_3=1. \tag{4.3}$$

Combining (4.2) and (4.3) gives

$$\begin{array}{rcl} p_1+p_2+p_3 & = & 1, \\ \lambda_{12}p_1-\dfrac{\lambda_{21}\lambda_{32}}{\lambda_{23}}p_3 & = & 0, \\ -\lambda_{12}p_1+(\lambda_{21}+\lambda_{23})p_2-\lambda_{32}p_3 & = & 0, \end{array}$$

or in matrix form

$$\begin{bmatrix} 1 & 1 & 1 \\ \lambda_{12} & 0 & -\frac{\lambda_{21}\lambda_{32}}{\lambda_{23}} \\ -\lambda_{12} & (\lambda_{21}+\lambda_{23}) & -\lambda_{32} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Using Cramer's Rule the limiting probabilities, (p_1, p_2, p_3), can be expressed as functions of the transition rates

$$\begin{array}{rcl} p_1 & = & \dfrac{\lambda_{21}\lambda_{32}}{\lambda_{21}\lambda_{32}+\lambda_{12}\lambda_{32}+\lambda_{23}\lambda_{12}}, \\ p_2 & = & \dfrac{\lambda_{12}\lambda_{32}}{\lambda_{21}\lambda_{32}+\lambda_{12}\lambda_{32}+\lambda_{23}\lambda_{12}}, \\ p_3 & = & \dfrac{\lambda_{23}\lambda_{12}}{\lambda_{21}\lambda_{32}+\lambda_{12}\lambda_{32}+\lambda_{23}\lambda_{12}}. \end{array} \tag{4.4}$$

As the limiting probabilities are independent of the observed transition times in the data, the number of transitions made by each patient, rather than the time at which the transitions are made, is used in the likelihood (Mostert *et al.*, 2004). The likelihood of patient i, if it is

assumed that patients are followed-up at regular intervals, follows the multinomial distribution,

$$l_i(p_1, p_2, p_3|data) \propto p_1^{m_{i1}} p_2^{m_{i2}} (1 - p_1 - p_2)^{n_i - m_{i1} - m_{i2}},$$

with n_i the total number of states visited by patient i and m_{ij} the number of times that patient i is in state j. For n independent patients the likelihood is given by

$$\begin{aligned} l(p_1, p_2, p_3|data) &\propto p_1^{\sum_i m_{i1}} p_2^{\sum_i m_{i2}} (1 - p_1 - p_2)^{\sum_i n_i - m_{i1} - m_{i2}} \\ &= p_1^{m_1} p_2^{m_2} (1 - p_1 - p_2)^{n - m_1 - m_2} \\ &= p_1^{m_1} p_2^{m_2} p_3^{m_3}, \end{aligned} \tag{4.5}$$

with each p_i, $i = 1, 2, 3$, defined as in (4.4).[3] *(see footnote 3)*

Two non-informative priors introduced in Section 3.1.1.1 are proposed as prior distributions for the limiting probabilities. Firstly, Jeffreys' non-informative prior, with the prior distribution based on Fisher's information matrix, will be used. Secondly, Zellner's MDI prior where the prior is chosen so as to maximise the average information in the data density relative to that in the prior distribution, will be used.

4.1.1 The Jeffreys prior on limiting probabilities

The Jeffreys prior is a non-informative prior based on Fisher's information matrix, $I(\mathbf{P})$. Fisher's information

$$I(\mathbf{P}) = -E_{\mathbf{P}} \left[\frac{\partial \ln f(\mathbf{m}|\mathbf{P})}{\partial p_i \partial p_j} \right],$$

for the proposed three state model with probabilities $\mathbf{P} = (p_1, p_2, p_3)$ and $\ln f(\mathbf{m}|P) = C + \sum_{i=1}^{3} m_i \ln p_i$ (with C a normalising constant) is given by

$$-E_{\mathbf{P}} \left[\frac{\partial \ln f(\mathbf{m}|\mathbf{P})}{\partial p_i \partial p_j} \right] = \begin{cases} \frac{n}{p_i} & i = j \\ 0 & i \neq j \end{cases}.$$

Jeffreys' prior, defined as

$$\pi_{Jef}(\mathbf{P}) \propto I^{1/2}(\mathbf{P}),$$

[3] Although the likelihood (4.5) seems to be a very simplistic function of the parameters at hand, this can be very misleading. The likelihood in this case is a very complex function of the actual parameters λ_{12}, λ_{21}, λ_{23} and λ_{32}.

for the three state model is given by

$$\pi_{Jef}(\mathbf{P}) \propto p_1^{-\frac{1}{2}} p_2^{-\frac{1}{2}} p_3^{-\frac{1}{2}},$$

which is a Dirichlet distribution with $\alpha_i = \frac{1}{2}, i = 1, 2, 3$.
As the Dirichlet distribution is not only the Jeffreys' prior for the multinomial distribution, but also the conjugate prior for the multinomial distribution (this prior is also the one group-reference and second order probability matching prior, see Section 3.1.1.1), the posterior distribution is also a Dirichlet distribution with parameters $\alpha_i = m_i + \frac{1}{2}, i = 1, 2, 3$,

$$\pi_{Jef}(\boldsymbol{\lambda}|data) = \pi_{Jef}(\mathbf{P}|data) \propto p_1^{m_1-\frac{1}{2}} p_2^{m_2-\frac{1}{2}} p_3^{m_3-\frac{1}{2}}, \tag{4.6}$$

with each p_i defined as in (4.4) (Agresti, 2002, pp. 607-608). As it is not possible to write this posterior in closed form a Metropolis-Hastings algorithm is presented in Section 4.1.3 to sample variates from this distribution.

4.1.2 The MDI prior for limiting probabilities

The non-informative maximal data information (MDI) prior is based on the negative entropy of a model and on maximising the function $G[\pi(\theta)]$. This function gives the total information provided by an experiment over and above the prior (Zellner, 1998) (see Section 3.1.1.1).

The negative entropy, as defined by Zellner (1998), for the proposed 3-state model with probabilities $\mathbf{P} = (p_1, p_2, p_3)$ and probability mass function $f(\mathbf{m}|\mathbf{P}) \propto p_1^{m_1} p_2^{m_2} p_3^{m_3}$, is given by

$$\begin{aligned} J(\mathbf{P}) &= \int f(\mathbf{m}|P) \ln f(\mathbf{m}|P) d\mathbf{m} \\ &= p_1 \ln p_1 + p_2 \ln p_2 + p_3 \ln p_3. \end{aligned}$$

Maximising

$$\begin{aligned} G[\pi_{MDI}(\mathbf{P})] &= \int J(\mathbf{P}) \pi_{MDI}(\mathbf{P}) d\mathbf{P} \\ &\quad - \int \pi(\mathbf{P}) \ln \pi_{MDI}(\mathbf{P}) d\mathbf{P} \end{aligned}$$

subject to $\pi_{MDI}(\mathbf{P})$ being proper (Zellner, 1998), gives the MDI prior for the 3-state model as

$$\begin{aligned}\pi_{MDI}(\mathbf{P}) &\propto e^{J(\mathbf{P})} \\ &= p_1^{p_1} p_2^{p_2} p_3^{p_3}.\end{aligned}$$

Combining the MDI prior with the likelihood gives the posterior distribution as

$$\pi_{MDI}(\boldsymbol{\lambda}|data) = \pi_{MDI}(\mathbf{P}|data) \propto p_1^{m_1+p_1} p_2^{m_2+p_2} p_3^{m_3+p_3}, \tag{4.7}$$

with each p_i defined as in (4.4). It is not possible to write this posterior in closed form and in Section 4.1.3 a Metropolis-Hastings algorithm is presented to sample variates from this distribution.

4.1.3 Sampling from the posteriors

In the previous section it was shown that posteriors (4.6) and (4.7) cannot be expressed in closed form when a function of the actual parameters, $\boldsymbol{\lambda}$. A suitable MCMC technique is therefore required to sample from these posterior distributions. The Metropolis-Hastings algorithm introduced in Section 3.3 will be used for this purpose.

The parameters of interest are the rates at which transitions are made between the various states, and the proposed candidate densities that will be used for these rates are independent exponential distributions

$$\lambda|\lambda_0 \sim \exp(\lambda_0), \quad \text{i.e. } q(\lambda|\lambda_0) = \lambda_0 e^{-\lambda_0 \lambda}, \ \lambda > 0, \ \lambda_0 > 0 \ .$$

The exponential distribution is chosen as a result of its constant hazard rate property, which coincides with Markov model properties.

Since there are four transition rates in the proposed 3-state Markov model, the combined candidate density for $\boldsymbol{\lambda} = (\lambda_{12}, \lambda_{21}, \lambda_{23}, \lambda_{32})$ is given as

$$\begin{aligned}q(\boldsymbol{\lambda}|\boldsymbol{\lambda}^i) &= \lambda_{12}^i e^{-\lambda_{12}^i \lambda_{12}} \lambda_{21}^i e^{-\lambda_{21}^i \lambda_{21}} \lambda_{23}^i e^{-\lambda_{23}^i \lambda_{23}} \lambda_{32}^i e^{-\lambda_{32}^i \lambda_{32}} \\ &= \lambda_{12}^i \lambda_{21}^i \lambda_{23}^i \lambda_{32}^i e^{-\lambda_{12}^i \lambda_{12} - \lambda_{21}^i \lambda_{21} - \lambda_{23}^i \lambda_{23} - \lambda_{32}^i \lambda_{32}}.\end{aligned}$$

The Metropolis-Hastings algorithm used to generate a sample from the posterior distributions

(4.6) and (4.7) is given by:

1) Choose an arbitrary starting point $\boldsymbol{\lambda}^0 = \left(\lambda_{12}^0, \lambda_{21}^0, \lambda_{23}^0, \lambda_{32}^0\right)$, set $i = 0$.

2) Generate a candidate value $\boldsymbol{\lambda}^*$ from

$$q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^i)$$

and u from

$$unif(0,1).$$

3) Set $\lambda^{i+1} = \lambda^*$ if

$$u < a(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*)$$

and $\lambda^{i+1} = \lambda^i$ otherwise. Where,

$$a(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*) = \min\left\{1; \frac{\pi(\boldsymbol{\lambda}^*|data)q(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*)}{\pi(\boldsymbol{\lambda}^i|data)q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^i)}\right\},$$

and $\pi(.|data)$ equals posterior (4.6) when using Jeffreys' prior or posterior (4.7) when using the MDI prior.

4) Set $i = i + 1$, return to step 2, repeating until $i = 5500$, with the first 500 being used as a burn-in sample.

The simulation study and its results are shown and discussed in Section (4.3).

4.2 Using a likelihood with the transition probabilities and rates

In the previous section a Bayesian MSM was developed where the likelihood was modelled using the limiting probabilities of a Markov process and these limiting probabilities were then used to estimate the transition rates. Due to the time independent nature of the limiting probabilities these can at most been seen as summary statistics of the Markov process. A lot of useful information available in the data about the nature of the transition rates is lost if they are based solely on the limiting probabilities. In this section a second B-MSM is developed

where the likelihood is based directly on the transition rates and the observed transitions in the data. The time of each transition and the length of time spent in each state are now directly incorporated into the likelihood. A prior distribution is placed on the transition rates and the likelihood and posterior distribution are directly expressed in terms of the transition rates.

Assume a 3-state Markov model with transition intensity matrix (4.1). As shown in Section 2.1.2.1 the likelihood can be expressed in terms of the transition probabilities of the process

$$L(\boldsymbol{\lambda}|data) \propto \prod_h p_h(t), \tag{4.8}$$

with the product defined over all transitions, $h, h = 1, ..., 4$ (as there are 4 possible transitions, $\lambda_{12}, \lambda_{21}, \lambda_{23}$, and λ_{32} in (4.1)) for all patients in the study, $p_h(t)$ the h^{th} element of $P(t)$ corresponding to transition λ_h (for transition $\lambda_1 = \lambda_{12}$ this would be row 1 and column 2 in $P(t)$) and $P(t)$ defined as in (2.12).

The parameters of interest in the model are the transition rates of the process. The functional form of the transition probabilities expressed in terms of the transition rates varies depending on the transition intensity matrix that is used. Due to this, the likelihood is expressed in terms of the transition probabilities, but the underlying parameters and the parameters onto which priors are placed, are the transition rates (see (2.13) for the solution of $p_{13}(t)$ in terms of the transition rates for one specific Q). Here the exponential distribution is used as a prior for the transition rates,

$$\lambda_h \sim \exp(\theta_h).$$

The exponential distribution is chosen as a result of its constant hazard rate property, which coincides with Markov model properties.

Combining the chosen prior distributions with the likelihood expressed in terms of the transition probabilities, the posterior distribution is given as,

$$\pi(\boldsymbol{\lambda}|data) \propto L(\boldsymbol{\lambda}|data) \left[\prod_h \theta_h \exp\left(-\theta_h \lambda_h\right) \right]. \tag{4.9}$$

As it is not possible to write (4.9) in a closed form, a Metropolis-Hastings algorithm is presented in Section 4.2.2 to draw samples from this posterior distribution.

4.2.1 Incorporating covariates

Covariates play an important role in multi-state models in that they alter the transition rates for individuals under study for different covariate values; individuals on treatment A may have lower rates compared to individuals on treatment B. The main aim of this dissertation is not to derive models where priors are placed on the covariate parameters, but to establish the influence of covariates on the transition rates.

In Section 2.1.2.2 the proportional hazards type formulation of Marshall and Jones (1995) was introduced to incorporate covariates into the multi-state model. Using this formulation the transition rates are defined as:

$$\lambda_h(\mathbf{z}) = \lambda_h e^{\mathbf{z}'\boldsymbol{\beta}_h}, \tag{4.10}$$

with λ_h the baseline transition rate, $\mathbf{z}$ the $(c \times 1)$ vector of c covariates and $\boldsymbol{\beta}_h$ the $(c \times 1)$ vector of their corresponding regression coefficients.

Under the Bayesian approach, a prior distribution is placed on the baseline transition rates, λ_h, as well as, on the regression coefficients, $\boldsymbol{\beta}_h$. An exponential prior is placed on each one of the baseline transition rates,

$$\lambda_h \sim \exp(\theta_h)$$

and a normal prior is placed on each one of the regression coefficients

$$\beta_{hg} \sim nor(\mu_{hg}, \sigma^2_{hg}), \tag{4.11}$$

with $h = 1, ..., 4$, and $g = 1, ..., c$.

Combining the chosen prior distributions with the likelihood expressed in terms of the transition probabilities, the posterior distribution in the presence of covariates is given as:

$$\pi(\boldsymbol{\lambda}|data) \propto L(\boldsymbol{\lambda}) \left[\prod_h \theta_h \exp\left(-\theta_h \lambda_h\right)\right] \left[\prod_{h,g} e^{-\frac{1}{2}\left(\frac{\beta_{hg}-\mu_{hg}}{\sigma_{hg}}\right)^2}\right] \tag{4.12}$$

Due to the fact that (4.12) cannot be expressed in a closed form, a Metropolis-Hastings algorithm is presented in Section 4.2.2 to draw samples from this posterior distribution.

4.2.2 Sampling from the posteriors

As it is not possible to directly sample from the posterior distributions (4.9) or (4.12), a Metropolis-Hastings algorithm is used to generate samples from these posteriors. The multi-variate normal (MVN) distribution is used as a candidate distribution for both the transition rates and the coefficients of the covariates. Although the transition rates are constrained to being values between 0 and 1, the multi-variate normal distribution is used here due to the ease with which its location and scale can be controlled. Only candidate values that fall within the allowable $(0, 1)$ range will be considered during the sampling process.

In the absence of covariates the posterior distribution of the transition rates is generated using the following steps:

1) Choose an arbitrary starting point $\boldsymbol{\lambda}^0 = (\lambda_1, ..., \lambda_h)'$, $h = 1, ..., 4$ and set $i = 0$.

2) Generate a candidate value $\boldsymbol{\lambda}^*$ from

$$q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^i) \sim MVN(\boldsymbol{\mu} = \boldsymbol{\lambda}^i, \Sigma)$$

on the condition that the generated transition rates fall within the allowable $(0, 1)$ range, and u from

$$unif(0, 1).$$

3) Set $\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^*$ if

$$u < a(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*)$$

and $\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i$ otherwise, where

$$a(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*) = \min\left\{1; \frac{\pi(\boldsymbol{\lambda}^*|data)q(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*)}{\pi(\boldsymbol{\lambda}^i|data)q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^i)}\right\}.$$

Since the MVN distribution is symmetrical,

$$\frac{q(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*)}{q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^i)} = 1$$

and $a(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*)$ can be simplified as

$$a(\boldsymbol{\lambda}^i|\boldsymbol{\lambda}^*) = \min\left\{\frac{\pi(\boldsymbol{\lambda}^*|data)}{\pi(\boldsymbol{\lambda}^i|data)}, 1\right\}$$

4) Set $i = i + 1$, return to step 2, repeating until $i = 5500$, with the first 500 being used as a burn-in sample.

In the presence of covariates the posterior distributions of the $\lambda's$ and the $\beta's$ are generated using the following steps:

1) Choose an arbitrary starting point $\boldsymbol{\Lambda}^0 = (\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0)'$, with $\boldsymbol{\lambda}^0 = (\lambda_1, ..., \lambda_h)'$ and $\boldsymbol{\beta}^0 = \left(\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_h^0\right)' = \left(\beta_{11}^0, ..., \beta_{1g}^0, \beta_{21}^0, ..., \beta_{hg}^0\right)'$, $h = 1, ..., 4$ and $g = 1, ..., c$, and set $i = 0$.
2) Generate a candidate value $\boldsymbol{\Lambda}^* = (\boldsymbol{\lambda}^*, \boldsymbol{\beta}_h^*)$ from

$$\begin{aligned} q(\boldsymbol{\lambda}^*|\boldsymbol{\lambda}^i) &\sim MVN(\boldsymbol{\mu} = \boldsymbol{\lambda}^i, \Sigma_\lambda) \\ q(\boldsymbol{\beta}_h^*|\boldsymbol{\beta}_h^i) &\sim MVN(\boldsymbol{\mu} = \boldsymbol{\beta}_h^i, \Sigma_{\beta_h}) \end{aligned}$$

on the condition that the generated transition rates fall within the allowable $(0, 1)$ range, $h = 1, ..., 4$, and u from

$$unif(0, 1).$$

3) Set $\boldsymbol{\Lambda}^{i+1} = \boldsymbol{\Lambda}^*$ if

$$u < a(\boldsymbol{\Lambda}^i|\boldsymbol{\Lambda}^*)$$

and $\boldsymbol{\Lambda}^{i+1} = \boldsymbol{\Lambda}^i$ otherwise. Where

$$a(\boldsymbol{\Lambda}^i|\boldsymbol{\Lambda}^*) = \min\left\{1; \frac{\pi(\boldsymbol{\Lambda}^*|data)q(\boldsymbol{\Lambda}^i|\boldsymbol{\Lambda}^*)}{\pi(\boldsymbol{\Lambda}^i|data)q(\boldsymbol{\Lambda}^*|\boldsymbol{\Lambda}^i)}\right\}.$$

Since the MVN distribution is symmetrical, $a(\boldsymbol{\Lambda}^i|\boldsymbol{\Lambda}^*)$ can be simplified as

$$a(\boldsymbol{\Lambda}^i|\boldsymbol{\Lambda}^*) = \min\left\{\frac{\pi(\boldsymbol{\Lambda}^*|data)}{\pi(\boldsymbol{\Lambda}^i|data)}, 1\right\}$$

4) Set $i = i + 1$, return to step 2, repeating until $i = 5500$, with the first 500 being as a the burn-in sample.

The results of a simulation study is shown and discussed in Section 4.3.

4.3 Simulation study

Two different methods of fitting Bayesian models to multi-state data were presented in this chapter:

- Firstly the likelihood was expressed in terms of the limiting probabilities of the Markov process, and
- Secondly the likelihood was expressed in terms of the transition probabilities of the process.

Each method can be divided into two approaches, giving the following four models:

i) Using the limiting probabilities in the likelihood with Jeffreys' prior placed on the limiting probabilities (LP-Jef).

ii) Using the limiting probabilities in the likelihood with the MDI prior placed on the limiting probabilities (LP-MDI).

iii) Using the transition probabilities in the likelihood with independent exponential distribution priors placed on the transition rates with no covariates in the model (TP-NoCov).

iv) Using the transition probabilities in the likelihood with independent exponential distribution priors placed on the transition rates with covariates in the model upon which normal distribution priors are assumed (TP-Cov).

In this Section the properties of these four proposed B-MSM's are investigated using simulation studies.

4.3.1 Simulation process and methodology

The starting point for the simulation study is to generate data from a known population. This sample generated from a known population is then used to assess the effectiveness of the proposed techniques. Effectiveness here is defined as the proposed technique being able to accurately estimate the known population values. To this end, the simulation process employed in this chapter will proceed as follows:

1) Generate a multi-state data set from a specified population using the process described in Section 2.3 (see Section 4.3.1.1 for the various models that will be used.).

2) Depending on the technique being investigated, LP-Jef, LP-MDI, TP-NoCov or TP-Cov, use

the appropriate M-H algorithm presented in Sections 4.1.3 and 4.2.2 to obtain the posterior distribution of the parameters of the specific multi-state model. The posterior distribution contains the 5000 effective posterior variates of the transition rates.

3) Assess the posterior distribution in terms of its effectiveness. See Section 4.3.1.2 for the posterior statistics that will be used to assess each simulation run.

An additional tool to assess the different models, is to obtain coverage probabilities for all relevant parameters. Since non-informative priors are used for the transitions rates, the coverage probabilities will indicate if the posteriors have good frequentist properties. To calculate the coverage probabilities, steps 1 and 2 are repeated l times. The relative frequency that the credible interval and HPD region include the known population value is calculated to define the coverage probability (see Section 4.3.1.2).

4.3.1.1 Different models and scenarios

The investigation of the effect of the different model structures in the simulation process is done by using 3 different Q matrices. As the interest in this dissertation is on the disease progression of individuals, it will be assumed that individuals can only make a transition from one state to an adjoining state and that no instantaneous transitions are possible to non-adjoining states. The models that will be investigated are the same models introduced in Section 2.3.1.1. The three 3-state models, Q_1 (2.26), Q_2 (2.27) and Q_3 (2.28) will be used in this simulation procedure.

As noted in Section 2.3.3, the sample size plays an important role when fitting a multi-state model. To this end, the effect of the sample size and the number of observations per patient on the proposed techniques are also investigated. Four different data scenarios (Sc1 to Sc4 from Table 2.2) are assumed and compared across all models. Along with these four data scenarios, data with covariates included will also be investiagted. As noted in Section 4.2.1 the aim of this dissertation is not to create a model with priors placed on the covariates and as such, when introducing covariates into the modelling process, only models with one categorical covariate (covariate model A as defined in Section 2.3.1.1) are investigated.

The prior assumptions regarding the transition rates play an important role in the B-MSM's. As such it is critical to understand and investigate how the proposed models are influenced by

the use of different priors. The different priors that will be investigated are:

- The effect of using Jeffreys' versus the MDI prior when using the limiting probabilities to model the multi-state process will be compared and investigated (see Section 4.1).
- The exponential distribution is used as prior distribution for all the transition rates in Section 4.2, where the multi-state model is modelled using the transition rates. Four different priors will be used here:

Prior_1 The first prior assumes that the rates are similar across all possible transitions

$$\theta_{12} = 0.2; \quad \theta_{21} = 0.2; \quad \theta_{23} = 0.2; \quad \theta_{32} = 0.2,$$

and that the mean time spent in each state before making a transition is $(0.2)^{-1} = 5$ months in state 1, $(0.2 + 0.2)^{-1} = 2.5$ months in state 2 and $(0.2)^{-1} = 5$ months in state 3.

Prior_2 The second prior will assume that the rates are similar across all possible transitions

$$\theta_{12} = 0.8; \quad \theta_{21} = 0.8; \quad \theta_{23} = 0.8; \quad \theta_{32} = 0.8,$$

and that the mean time spent in each state before making a transition is $(0.8)^{-1} = 1.25$ months in state 1, $(0.8 + 0.8)^{-1} = 0.625$ months in state 2 and $(0.8)^{-1} = 1.25$ months in state 3.

Prior_3 The third prior will assume that the rates for transitions to higher states differ from the rates for transitions to lower states

$$\theta_{12} = 0.2; \quad \theta_{21} = 0.8; \quad \theta_{23} = 0.2; \quad \theta_{32} = 0.8.$$

A patient is assumed to be more likely to make a transition to a lower state (0.8 vs. 0.2) and once in a lower state a patient is assumed to spend more time in that state (5 months in state 1, 1 month in state 2 and 1.25 months in 3).

Prior_4 The fourth prior is the opposite of the third prior

$$\theta_{12} = 0.8; \quad \theta_{21} = 0.2; \quad \theta_{23} = 0.8; \quad \theta_{32} = 0.2.$$

A patient is assumed to be more likely to make a transition to a higher state (0.8 vs. 0.2) and once in a higher state a patient is assumed to spend more time in that state (1.25 months in state 1, 1 month in state 2 and 5 months in 3).

– The effect of the covariates (the $\beta's$ in equation 4.10) is assumed to be located around -0.7 and a flat, $\beta_{hg} \sim nor(-0.7, 1000^2)$, $h = 1, ..., 4$ and $g = 1, ..., c$, prior is used as prior distribution for all covariate effects.

Combining the different model structures, data scenarios and prior assumptions gives $3 \times 4 \times 2 = 24$ simulation studies for the approach in Section 4.1 and $3 \times 4 \times 4 = 48$ simulation studies for the approach in Section 4.2.

4.3.1.2 Assessing the simulation process

The posterior MSE will be the main statistic used to assess the performance of the proposed techniques. A smaller MSE will indicate that the estimates of the transition rates can be viewed as representative of the population under study, while a large MSE indicates a departure from the population values. As the MSE can be influenced by extreme values, the median square error (MedSE) will be presented in cases where extreme MSE values are observed. The use of the MedSE will be clearly highlighted in the results.

The bias, another indicator of to the performance of the simulation process, and defined as

$$Bias(\hat{\theta}) = \sqrt{MSE - var(\hat{\theta})},$$

will be presented in a select number of simulation runs.

The mean, median and standard deviation of the posterior distribution will also be investigated for each simulation run.

The credible interval (Cred) and highest posterior density (HPD) region is calculated for each simulation run and will be used to calculate frequentist coverage probabilities of the process. For this simulation study the coverage probabilities will be based on 1000 repetitions of the process described in Section 4.3.1. The mean length of the posterior intervals over the 1000 repetitions are also calculated to give an indication of the width of the posterior intervals.

4.3.2 Simulation results

In this section the simulation results are presented for the three different Baysian approaches when estimating transition rates:

i) Limiting probabilities in the likelihood (see Appendix A.2 for the R functions and programs used to generate these results),

ii) Transition probabilities in the likelihood - no covariates (see Appendix A.3 for the R functions and programs used to generate these results), and

iii) Transition probabilities in the likelihood - with covariates (see Appendix A.3 for the R functions and programs used to generate these results).

4.3.2.1 Limiting probabilities in the likelihood

In Tables 4.1 to 4.6, comparisons are made between the two priors, Jeffrey's and MDI, when using the limiting probabilities in the likelihood for the different data scenarios noted in the previous section. Tables 4.1 to 4.3 summarise the posterior distributions across the different scenarios, Table 4.4 shows the bias associated with a select number of scenarios and in Tables 4.5 and 4.6 the posterior coverage probabilities and mean lengths of the credible and HPD intervals with $\alpha = 0.05$ are shown for a select number of scenarios.

The results can be summarised as follows:

- The maximum MSE observed for both priors across the three models and four scenarios is 0.1058 (model Q_1 and scenario 4, Table 4.1) and the maximum bias is 0.278 (model Q_1 and scenario 4, Table 4.4). This indicates that even though this method of using limiting probabilities in the likelihood is at best a summarising technique, in that it does not take time into consideration, both these priors are performing well with regards to estimating the parameters under study.
- The sample size being used does not have a big influence on the MSE or bias values of the model. When comparing the MSE of scenario 2 with the MSE of scenario 3 (the two scenarios with the smallest and largest number of observations), the values are similar across the two priors and the 3 models.
- The coverage probabilities of the two priors are similar and again no big differences are observed when using a smaller or a larger data set (Tables 4.5 and 4.6). What is apparent is that as the sample size increases, the mean lengths of the intervals decrease (from mean lengths of around 0.6 when based on 25 individuals to mean lengths around 0.4 when based on 50 individuals), indicating that the larger sample size leads to narrower posterior intervals as expected.
- When comparing the results presented here to those in Section 2.3.3, it is clear that this

method of estimating the transition rates has significantly higher MSE values than the method presented in that section. This said, it was found that for model Q_2 and scenario 2 (Table 2.5) the estimating maximum likelihood procedure produced very unstable parameter estimates. This instability is not present when modelling the data using the limiting probabilities in the likelihood.

From these results it is clear that this method of modelling multi-state data gives reasonable parameter estimates of the population transition rates, but that there is no real difference using the non-informative priors.

Table 4.1: Simulation results of modelling multi-state data using the limiting probabilities - Q_1.

Q1		**LP-Jef**				**LP-MDI**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.5	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Sc1	**Mean**	0.6406	0.6220	0.6001	0.6558	0.6467	0.6261	0.6126	0.6668
	Med	0.6800	0.6582	0.6194	0.6882	0.6746	0.6532	0.6237	0.6917
	SD	0.1907	0.1882	0.1692	0.1724	0.1689	0.1685	0.1590	0.1570
	MSE	0.0957	0.0862	0.0705	0.0960	0.0919	0.0800	0.0713	0.0969
Sc2	**Mean**	0.5687	0.6398	0.5605	0.6697	0.5750	0.6486	0.5671	0.5744
	Med	0.5811	0.6677	0.5696	0.6041	0.5924	0.6753	0.5764	0.6124
	SD	0.1733	0.1747	0.1732	0.1704	0.1667	0.1689	0.1746	0.1709
	MSE	0.0635	0.0898	0.0585	0.1026	0.0650	0.0909	0.0609	0.1056
Sc3	**Mean**	0.6140	0.6439	0.6176	0.5411	0.6236	0.6562	0.6346	0.5571
	Med	0.6355	0.6738	0.5498	0.5759	0.6484	0.6869	0.5609	0.5898
	SD	0.1631	0.1679	0.1776	0.1803	0.1608	0.1657	0.1637	0.1675
	MSE	0.0774	0.0888	0.0808	0.0913	0.0812	0.0944	0.0868	0.0960
Sc4	**Mean**	0.6027	0.6766	0.6847	0.6560	0.6957	0.6702	0.5676	0.6384
	Med	0.6204	0.6102	0.6039	0.6879	0.6165	0.6077	0.5893	0.6691
	SD	0.1647	0.1685	0.1625	0.1681	0.1699	0.1732	0.1724	0.1797
	MSE	0.0699	0.1058	0.0698	0.0951	0.0677	0.1036	0.0658	0.0912

Table 4.2: Simulation results of modelling multi-state data using the limiting probabilities - Q_2.

Q2		**LP-Jef**				**LP-MDI**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.25	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Sc1	**Mean**	0.2435	0.7942	0.2637	0.7917	0.2427	0.7947	0.2692	0.7994
	Med	0.2474	0.8350	0.2561	0.8345	0.2461	0.8338	0.2644	0.8395
	SD	0.0623	0.1654	0.0902	0.1677	0.0601	0.1610	0.0908	0.1635
	MSE	0.0052	0.0296	0.0217	0.0301	0.0052	0.0286	0.0208	0.0300
Sc2	**Mean**	0.2507	0.7942	0.2811	0.7993	0.2488	0.7917	0.2874	0.7961
	Med	0.2485	0.8323	0.2655	0.8335	0.2450	0.8176	0.2734	0.8327
	SD	0.0698	0.1615	0.1114	0.1591	0.0680	0.1553	0.1113	0.1618
	MSE	0.0075	0.0287	0.0140	0.0279	0.0066	0.0264	0.0145	0.0284
Sc3	**Mean**	0.2717	0.7769	0.2358	0.7983	0.2765	0.7910	0.2383	0.7983
	Med	0.2756	0.8127	0.2316	0.8385	0.2801	0.8269	0.2343	0.8345
	SD	0.0707	0.1709	0.0729	0.1614	0.0672	0.1635	0.0725	0.1591
	MSE	0.0060	0.0304	0.0060	0.0285	0.0061	0.0290	0.0058	0.0278
Sc4	**Mean**	0.2837	0.7429	0.1832	0.7919	0.3099	0.8088	0.1821	0.7857
	Med	0.2974	0.7885	0.1830	0.8319	0.3126	0.8338	0.1816	0.8304
	SD	0.0828	0.1996	0.0525	0.1652	0.0621	0.1424	0.0541	0.1711
	MSE	0.0082	0.0414	0.0083	0.0293	0.0077	0.0238	0.0085	0.0308

Table 4.3: Simulation results of modelling multi-state data using the limiting probabilities - Q_3.

Q3		**LP-Jef**				**LP-MDI**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.75	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Sc1	**Mean**	0.7946	0.4124	0.7806	0.2892	0.7948	0.4149	0.8057	0.3017
	Med	0.8327	0.4032	0.8252	0.2921	0.8308	0.4094	0.8413	0.3033
	SD	0.1610	0.1271	0.1713	0.0791	0.1610	0.1264	0.1506	0.0731
	MSE	0.0282	0.0570	0.0325	0.0083	0.0281	0.0591	0.0258	0.0090
Sc2	**Mean**	0.7951	0.3003	0.7934	0.3533	0.7997	0.3036	0.7951	0.3563
	Med	0.8320	0.2852	0.8344	0.3506	0.8397	0.2906	0.8290	0.3514
	SD	0.1626	0.1119	0.1618	0.0971	0.1583	0.1069	0.1575	0.0969
	MSE	0.0286	0.0225	0.0284	0.0221	0.0277	0.0220	0.0270	0.0229
Sc3	**Mean**	0.7809	0.2686	0.7547	0.2755	0.7938	0.2764	0.7952	0.2900
	Med	0.8225	0.2705	0.7888	0.2826	0.8346	0.2754	0.8253	0.2960
	SD	0.1742	0.0772	0.1829	0.0727	0.1653	0.0772	0.1578	0.0663
	MSE	0.0318	0.0064	0.0346	0.0074	0.0296	0.0067	0.0316	0.0077
Sc4	**Mean**	0.7928	0.3269	0.8047	0.3230	0.7866	0.3283	0.7765	0.3113
	Med	0.8288	0.3220	0.8440	0.3278	0.8274	0.3273	0.8161	0.3180
	SD	0.1625	0.0955	0.1566	0.0756	0.1695	0.0999	0.1667	0.0762
	MSE	0.0283	0.0206	0.0277	0.0161	0.0302	0.0216	0.0296	0.0152

Table 4.4: Bias for a select number of models using the limiting probabilities.

Model	Scenario	Prior	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Q1	1	Jef	0.244	0.225	0.205	0.258
		MDI	0.252	0.227	0.214	0.269
	2	Jef	0.183	0.243	0.169	0.271
		MDI	0.193	0.250	0.174	0.276
	3	Jef	0.225	0.246	0.222	0.242
		MDI	0.235	0.259	0.245	0.261
	4	Jef	0.207	0.278	0.208	0.259
		MDI	0.197	0.271	0.190	0.243
Q3	1	Jef	0.047	0.202	0.056	0.045
		MDI	0.047	0.208	0.056	0.060
	2	Jef	0.047	0.100	0.047	0.112
		MDI	0.051	0.103	0.047	0.116
	3	Jef	0.038	0.021	0.033	0.046
		MDI	0.048	0.028	0.082	0.057
	4	Jef	0.043	0.107	0.057	0.102
		MDI	0.038	0.108	0.042	0.097

Table 4.5: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals of modelling multi-state data using the limiting probabilities - Q_1 Sc_1.

Prior	Interval	Statistic	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Jef	Cred	Cov %	0.9761	0.9368	0.9574	0.9580
	HPD	Cov %	0.8743	0.9820	0.9281	0.9042
MDI	Cred	Cov %	0.9880	0.8574	0.8369	0.9640
	HPD	Cov %	0.8743	0.9461	0.9641	0.8922
Jef	Cred	Mean length	0.6111	0.6117	0.6250	0.6253
	HPD	Mean length	0.5680	0.5835	0.5904	0.5812
MDI	Cred	Mean length	0.6214	0.6197	0.6215	0.6198
	HPD	Mean length	0.5718	0.5898	0.5854	0.5753

Table 4.6: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals of modelling multi-state data using the limiting probabilities - Q_3 Sc_3.

Prior	Interval	Statistic	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Jef	Cred	Cov %	0.9454	0.9376	0.9486	0.9452
	HPD	Cov %	0.9421	0.9632	0.9435	0.9387
MDI	Cred	Cov %	0.9540	0.9438	0.9353	0.9386
	HPD	Cov %	0.9456	0.9521	0.9354	0.9435
Jef	Cred	Mean length	0.4356	0.3287	0.3246	0.3214
	HPD	Mean length	0.3765	0.2896	0.3276	0.3426
MDI	Cred	Mean length	0.3745	0.3521	0.3764	0.3243
	HPD	Mean length	0.4187	0.4043	0.3876	0.3974

4.3.2.2 Transition probabilities in the likelihood - No covariates

Tables 4.7 to 4.15 summarise the results from the simulation study where B-MSM's are fitted using the transition probabilities in the likelihood with no covariates. The posterior distributions across the different scenarios are given in Tables 4.7 to 4.12, the bias for a select few scenarios are given in Table 4.13 and in Tables 4.14 and 4.15 the posterior coverage probabilities and mean lengths of the credible and HPD intervals with $\alpha = 0.05$ are shown for a few select scenarios.

The results can be summarised as follows:

- For model Q_1 (Tables 4.7, 4.8 and 4.13), there are no clear differences with regards to the MSE or bias values across the four priors. This indicates that none of these priors performs significantly different from the others. This is to be expected, as the population parameters for model Q_1 is assumed to be 0.5, and the prior values placed on the unknown parameters are either 0.2 or 0.8; both the same distance from 0.5.
- Clear differences are observed when using the 4 different priors to model models Q_2 and Q_3 (Tables 4.9 to 4.12).
    - When modelling Q_2, prior 3 generally leads to smaller MSE's than the other three priors (Tables 4.9 and 4.10). In model Q_2, the forward transition rates (the rates to move up to the next higher state) for the population are 0.25 and the backward rates (the rates to move down to the next lower state) are 0.75. Prior 3, with forward rates of 0.2 and backward rates of 0.8, is the closest prior to these population values and should thus be the best choice of prior for this model.
    - When modelling Q_3, prior 4 is generally the prior with the smallest MSE and bias values (Tables 4.11, 4.12 and 4.13). In model Q_3, the forward transition rates for the population are 0.75 and the backward rates 0.25. Prior 4, with forward rates of 0.8 and backward rates of 0.2, is the closest prior to these population values and should thus be the best choice of prior for this model.
- The effect of the four data scenarios is visible across all models and priors. The MSE values increase from scenario 1 to scenario 2 as the percentage of missing observations increases from 10% to 50%; decreases when the sample size increases to 50 for scenario 3 and increases under scenario 4 (compared to scenario 3) when the percentage of missing observations again

increases to 50%.

- The frequentist coverage probabilities are of similar order for both models Q_1 and Q_3, although the coverage does increase as the sample size is increased from 25 to 50 individuals (Tables 4.14 and 4.15). The same result is observed for the mean lengths of the posterior intervals; they decrease as the sample size is increased.
- Compared to Section 2.3.3, where the data was modelled without any prior information, it is clear that the MSE's have increased, but with this increase the results appear to be more stable. None of the extreme MSE's observed in Section 2.3.3 are present in the models considered here.

From these results it is reasonable to claim that this method of modelling multi-state data gives good all-round parameter estimates of the population transition rates. When the correct prior is used, i.e. the prior that best matches the underlying population model, the MSE's of the models decrease, indicating that correctly specifying the prior does lead to better parameter estimates. This is observed across all three models and all four data scenarios, unlike the case where no prior information is used in Section 2.3.3.

Table 4.7: Simulation results of modelling multi-state data using the transition probabilities without covariates included in the model - Q_1 Sc_1 and Sc_2.

Q1		**Sc1**				**Sc2**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.5	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.5228	0.5052	0.4363	0.4339	0.5627	0.6727	0.4803	0.5188
	Med	0.5222	0.5032	0.4230	0.4276	0.5512	0.6705	0.4693	0.5025
	SD	0.1131	0.1142	0.0996	0.0950	0.1482	0.1451	0.1446	0.1592
	MSE	0.0206	0.0205	0.0145	0.0145	0.0297	0.0635	0.0288	0.0460
Pr2	**Mean**	0.5148	0.4749	0.4329	0.5061	0.5024	0.5299	0.4797	0.5365
	Med	0.5099	0.4668	0.4141	0.4887	0.4942	0.5206	0.4601	0.5226
	SD	0.1081	0.1123	0.1212	0.1271	0.1234	0.1319	0.1404	0.1582
	MSE	0.0224	0.0267	0.0205	0.0216	0.0177	0.0240	0.0273	0.0400
Pr3	**Mean**	0.5281	0.4933	0.4018	0.4333	0.5519	0.5763	0.5258	0.5593
	Med	0.5132	0.4819	0.3940	0.4249	0.5403	0.5643	0.5118	0.5506
	SD	0.1208	0.1204	0.0835	0.0886	0.1344	0.1385	0.1557	0.1603
	MSE	0.0221	0.0200	0.0191	0.0160	0.0259	0.0386	0.0267	0.0364
Pr4	**Mean**	0.4037	0.4391	0.5461	0.5822	0.4656	0.4712	0.5765	0.6268
	Med	0.3905	0.4227	0.5312	0.5899	0.4390	0.4568	0.5637	0.6094
	SD	0.0858	0.0971	0.1473	0.1366	0.1369	0.1270	0.1445	0.1561
	MSE	0.0211	0.0161	0.0507	0.0351	0.0234	0.0364	0.0638	0.0511

Table 4.8: Simulation results of modelling multi-state data using the transition probabilities without covariates included in the model - Q_1 Sc_3 and Sc_4.

Q1		**Sc3**				**Sc4**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.5	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.3973	0.4093	0.5112	0.4866	0.493	0.485	0.527	0.499
	Med	0.3918	0.4043	0.5038	0.4784	0.485	0.477	0.517	0.486
	SD	0.0593	0.0656	0.0798	0.0808	0.090	0.091	0.110	0.116
	MSE	0.0149	0.0160	0.0086	0.0089	0.011	0.011	0.028	0.025
Pr2	**Mean**	0.4397	0.4405	0.4468	0.4336	0.4777	0.4305	0.4417	0.4520
	Med	0.4352	0.4370	0.4360	0.4258	0.4697	0.4203	0.4359	0.4459
	SD	0.0647	0.0703	0.0734	0.0732	0.0976	0.0904	0.0905	0.0940
	MSE	0.0091	0.0087	0.0120	0.0111	0.0112	0.0159	0.0161	0.0135
Pr3	**Mean**	0.4445	0.4531	0.4433	0.5040	0.5053	0.5032	0.4528	0.4845
	Med	0.4414	0.4463	0.4385	0.5011	0.4991	0.4912	0.4437	0.4714
	SD	0.0640	0.0683	0.0670	0.0745	0.1037	0.1135	0.0948	0.1027
	MSE	0.0109	0.0136	0.0141	0.0108	0.0195	0.0180	0.0136	0.0148
Pr4	**Mean**	0.4091	0.4242	0.4168	0.4226	0.4599	0.5113	0.4418	0.4590
	Med	0.4067	0.4191	0.4138	0.4165	0.4479	0.4926	0.4290	0.4407
	SD	0.0580	0.0675	0.0653	0.0696	0.1035	0.1262	0.0916	0.1001
	MSE	0.0123	0.0134	0.0169	0.0140	0.0145	0.0170	0.0157	0.0167

Table 4.9: Simulation results of modelling multi-state data using the transition probabilities without covariates included in the model - Q_2 Sc_1 and Sc_2.

Q2		**Sc1**				**Sc2**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.25	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.2354	0.7786	0.1885	0.5519	0.2172	0.7424	0.2239	0.5975
	Med	0.2314	0.7807	0.1742	0.5418	0.2113	0.7544	0.2068	0.5881
	SD	0.0468	0.1075	0.0759	0.1419	0.0602	0.1382	0.1007	0.1740
	MSE	0.0040	0.0150	0.0144	0.0618	0.0055	0.0233	0.0205	0.0635
Pr2	**Mean**	0.2632	0.7339	0.2276	0.6161	0.2084	0.6626	0.2768	0.7151
	Med	0.2582	0.7299	0.2187	0.6037	0.1994	0.6508	0.2618	0.7402
	SD	0.0552	0.1246	0.0770	0.1705	0.0562	0.1349	0.1096	0.1878
	MSE	0.0079	0.0186	0.0088	0.0541	0.0059	0.0300	0.0174	0.0477
Pr3	**Mean**	0.2522	0.7458	0.2179	0.6855	0.2473	0.7470	0.2696	0.6924
	Med	0.2475	0.7508	0.2056	0.6922	0.2417	0.7591	0.2554	0.5942
	SD	0.0546	0.1296	0.0696	0.1749	0.0653	0.1483	0.0946	0.1731
	MSE	0.0037	0.0139	0.0063	0.0379	0.0053	0.0199	0.0163	0.0611
Pr4	**Mean**	0.2838	0.7703	0.2087	0.6812	0.2917	0.8402	0.1995	0.5648
	Med	0.2801	0.7809	0.2109	0.6844	0.2866	0.8576	0.1804	0.5504
	SD	0.0582	0.1242	0.0698	0.1433	0.0658	0.1067	0.0925	0.1696
	MSE	0.0042	0.0208	0.0087	0.0664	0.0046	0.0249	0.0164	0.0615

Table 4.10: Simulation results of modelling multi-state data using the transition probabilities without covariates included in the model - Q_2 Sc_3 and Sc_4.

Q2		**Sc3**				**Sc4**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.25	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.2255	0.6096	0.1780	0.6832	0.2589	0.7647	0.2360	0.6423
	Med	0.2196	0.6004	0.1680	0.6827	0.2549	0.7695	0.2268	0.6317
	SD	0.0427	0.0925	0.0583	0.1511	0.0489	0.1070	0.0715	0.1513
	MSE	0.0033	0.0365	0.0096	0.0335	0.0077	0.0196	0.0066	0.0363
Pr2	**Mean**	0.2251	0.7008	0.1916	0.6676	0.2118	0.6609	0.2617	0.7109
	Med	0.2198	0.6931	0.1868	0.6656	0.2046	0.6508	0.2515	0.7164
	SD	0.0439	0.1049	0.0506	0.1076	0.0516	0.1173	0.0795	0.1518
	MSE	0.0032	0.0153	0.0063	0.0267	0.0053	0.0303	0.0132	0.0284
Pr3	**Mean**	0.2370	0.7206	0.2052	0.7505	0.2415	0.7330	0.2291	0.7198
	Med	0.2230	0.7181	0.2097	0.7410	0.2383	0.7273	0.2277	0.7198
	SD	0.0398	0.0864	0.0510	0.1331	0.0496	0.1169	0.0684	0.1568
	MSE	0.0025	0.0166	0.0091	0.0221	0.0043	0.0122	0.0108	0.0448
Pr4	**Mean**	0.2128	0.6240	0.2975	0.6319	0.2685	0.6606	0.3019	0.6364
	Med	0.2173	0.6103	0.2912	0.6275	0.2653	0.6539	0.2925	0.6397
	SD	0.0419	0.1012	0.0564	0.0226	0.0531	0.1246	0.0693	0.1533
	MSE	0.0044	0.0228	0.0152	0.0520	0.0073	0.0193	0.0173	0.0661

Table 4.11: Simulation results of modelling multi-state data using the transition probabilities without covariates included in the model - Q_3 Sc_1 and Sc_2.

Q3		**Sc1**				**Sc2**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.75	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.6824	0.3061	0.8152	0.2521	0.6998	0.2775	1.0544	0.3525
	Med	0.6783	0.2975	0.8107	0.2440	0.6985	0.2657	1.0316	0.3279
	SD	0.1242	0.0834	0.1569	0.0644	0.1588	0.1041	0.2950	0.1121
	MSE	0.0254	0.0151	0.0519	0.0055	0.0399	0.0152	0.2291	0.0317
Pr2	**Mean**	0.5718	0.1375	0.6495	0.2225	0.6952	0.2246	0.7488	0.3012
	Med	0.5663	0.1247	0.6405	0.2159	0.6951	0.2057	0.7650	0.2999
	SD	0.1347	0.0622	0.1238	0.0541	0.1293	0.0957	0.1772	0.0793
	MSE	0.0516	0.0198	0.0259	0.0047	0.0306	0.0131	0.0511	0.0185
Pr3	**Mean**	0.5652	0.2585	0.6841	0.2019	0.7227	0.3002	0.8492	0.2810
	Med	0.5539	0.2458	0.6715	0.1925	0.7316	0.2898	0.8522	0.2758
	SD	0.1319	0.0839	0.1250	0.0554	0.1441	0.0999	0.1967	0.0767
	MSE	0.0568	0.0107	0.0339	0.0091	0.0256	0.0421	0.1259	0.0203
Pr4	**Mean**	0.6585	0.2213	0.7309	0.2822	0.6723	0.2377	0.7466	0.2931
	Med	0.6641	0.2158	0.7121	0.2716	0.6780	0.2351	0.7347	0.2907
	SD	0.1255	0.0719	0.1528	0.0693	0.1429	0.0814	0.2034	0.0765
	MSE	0.0364	0.0085	0.0225	0.0029	0.0251	0.0108	0.0495	0.0119

Table 4.12: Simulation results of modelling multi-state data using the transition probabilities without covariates included in the model - Q_3 Sc_3 and Sc_4.

Q3		**Sc3**				**Sc4**			
	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{12}	λ_{21}	λ_{23}	λ_{32}
		0.75	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.6383	0.2073	0.9238	0.3131	0.7015	0.2099	0.7671	0.2848
	Med	0.6330	0.2004	0.9142	0.3121	0.6987	0.2042	0.7502	0.2785
	SD	0.1078	0.0563	0.1498	0.0524	0.1216	0.0622	0.1535	0.0635
	MSE	0.0264	0.0053	0.1162	0.0132	0.0201	0.0059	0.0313	0.0079
Pr2	**Mean**	0.6577	0.1876	0.6457	0.2071	0.6746	0.2187	0.7983	0.2746
	Med	0.6537	0.1813	0.6392	0.2022	0.6671	0.2082	0.7745	0.2661
	SD	0.1213	0.0540	0.0924	0.0400	0.1188	0.0672	0.1827	0.0723
	MSE	0.0266	0.0070	0.0220	0.0038	0.0227	0.0061	0.0719	0.0094
Pr3	**Mean**	0.5607	0.2135	0.8166	0.2392	0.7659	0.3086	0.7069	0.2421
	Med	0.5475	0.2027	0.8065	0.2343	0.7855	0.2951	0.6977	0.2361
	SD	0.1015	0.0639	0.1164	0.0432	0.1307	0.0876	0.1160	0.0495
	MSE	0.0512	0.0070	0.0323	0.0025	0.0217	0.0275	0.0335	0.0060
Pr4	**Mean**	0.6996	0.1937	0.7742	0.2570	0.7100	0.2467	0.7349	0.2375
	Med	0.6969	0.1863	0.7609	0.2512	0.7114	0.2471	0.7337	0.2312
	SD	0.1229	0.0584	0.1134	0.0492	0.1232	0.0750	0.0153	0.0483
	MSE	0.0226	0.0051	0.0116	0.0038	0.0208	0.0019	0.0205	0.0070

Table 4.13: Bias for a select number of models using the transition probabilities without covariates included in the model.

Model	**Scenario**	**Prior**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Q1	1	1	0.088	0.086	0.068	0.074
		2	0.103	0.119	0.126	0.074
		3	0.087	0.074	0.110	0.090
		4	0.117	0.082	0.170	0.128
	3	1	0.107	0.108	0.048	0.049
		2	0.070	0.061	0.081	0.076
		3	0.082	0.095	0.098	0.073
		4	0.095	0.094	0.112	0.096
Q3	1	1	0.100	0.090	0.165	0.037
		2	0.183	0.126	0.103	0.042
		3	0.144	0.057	0.200	0.090
		4	0.198	0.060	0.082	0.078
	3	1	0.121	0.046	0.306	0.102
		2	0.109	0.064	0.116	0.047
		3	0.087	0.060	0.137	0.037
		4	0.202	0.054	0.137	0.025

Table 4.14: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals of modelling multi-state data using the transition probabilities without covariates included in the model - $Q_1\ Sc_1\ Pr_1$.

Interval	**Statistic**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$
Cred	Cov %	0.9401	0.8802	0.9102	0.9281
HPD	Cov %	0.9042	0.8563	0.8982	0.9222
Cred	Mean length	0.3775	0.3627	0.3802	0.3784
HPD	Mean length	0.3663	0.3515	0.3706	0.3679

Table 4.15: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals of modelling multi-state data using the transition probabilities without covariates included in the model - $Q_3\ Sc_4\ Pr_1$.

Interval	**Statistic**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$
Cred	Cov %	0.9512	0.9125	0.9654	0.9358
HPD	Cov %	0.9441	0.9325	0.9425	0.9587
Cred	Mean length	0.2587	0.3198	0.2857	0.2798
HPD	Mean length	0.2967	0.2987	0.3157	0.2587

4.3.2.3 Transition probabilities in the likelihood - with covariates

In Tables 4.16 to 4.30 the results when including a covariate in the B-MSM, fitted using the transition probabilties in the likelihood, are summarised. The posterior distributions across the different scenarios are given in Tables 4.16 to 4.27, the bias for a select few scenarios are given in Table 4.28 and in Tables 4.29 and 4.30 the posterior coverage probabilities and mean lengths of the credible and HPD intervals ($\alpha = 0.05$) are shown for a select few scenarios.

The results can be summarised as follows:

- In general, similar to the models without covariates discussed in the previous section, no clear differences between the MSE's and biases are observed for model Q_1 across the four priors (Tables 4.16, 4.19 and 4.28); for model Q_2 prior 3 generally leads to smaller MSE's (Tables 4.20 and 4.23) and for model Q_3 prior 4 is generally the prior with the smallest MSE's (Tables 4.24 and 4.27).
- No big differences in the MSE's of the covariate effects are observed if the values are compared on a per scenario basis, i.e. the values for Q_1 scenario 1 (Table 4.16) are compared with Q_2 scenario 1 (Table 4.20) and Q_3 scenario 1 (Table 4.24). This is to be expected, as the prior on the covariates are the same across all models and scenarios. Big differences are however observed if the values are compared across different scenarios. For model Q_1 scenario 2, a MSE value of 3.52 is observed, while for Q_1 scenario 3, the largest MSE is found to be 0.4265 (Table 4.18). This indicates that the sample size plays a very important role when modelling these more complex models.
- The effect of the four data scenarios is even more visible across all models and priors than in the previous section. The MSE values increase from scenario 1 to scenario 2 as the percentage of missing observations increases from 10% to 50%; decreases when the sample size increases to 50 for scenario 3 and increases under scenario 4 (compared to scenario 3) when the percentage of missing observations again increases to 50%.
- The frequentist coverage probabilities are of similar order for both model Q_1 and Q_3, although the coverage does increase as the sample size is increased from 25 to 50 individuals (Tables 4.14 and 4.15). The same is observed for the mean lengths of the posterior intervals; they decrease as the sample size is increased.
- Compared to the previous section where the data was modelled without covariates, it is clear

that the MSE's have increased significantly, except for data scenario 3. This corresponds to the results found in Section 2.3.3; as the complexity of the model increases, the sample size becomes more and more important to ensure reliable parameter estimates.

As in the previous section, these results indicate that this method of modelling multi-state data gives good all-round parameter estimates of the population transition rates, as long as the sample size increases when handling more complex models.

Table 4.16: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_1 Sc_1.

Q1	**Par**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Sc1		**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.7708	0.7110	0.7377	0.7222	-0.6567	-0.4868	-0.4791	-0.9360
	Med	0.7171	0.6609	0.6513	0.6405	-0.6340	-0.4547	-0.3509	-0.7814
	SD	0.2790	0.2618	0.3062	0.2943	0.5439	0.5110	0.4952	0.5286
	MSE	0.1667	0.1379	0.2973	0.3184	0.9423	0.7739	0.8595	1.6120
Pr2	**Mean**	0.6180	0.6175	0.5287	0.5533	-0.4267	-0.5561	-0.3571	-0.0045
	Med	0.6137	0.6164	0.5210	0.5432	-0.4445	-0.5679	-0.3322	-0.0417
	SD	0.1451	0.1439	0.1426	0.1566	0.3557	0.3600	0.3960	0.3532
	MSE	0.0522	0.0559	0.0376	0.0552	0.2142	0.3260	1.2387	0.6089
Pr3	**Mean**	0.5699	0.5893	0.6288	0.6003	-0.2733	-0.1350	-0.2526	-0.2487
	Med	0.5678	0.5828	0.6059	0.5850	-0.2620	-0.1108	-0.1779	-0.1738
	SD	0.1334	0.1396	0.1689	0.1642	0.3583	0.3600	0.4821	0.5177
	MSE	0.0527	0.0454	0.0553	0.0678	0.2148	0.3263	0.4574	0.6269
Pr4	**Mean**	0.6524	0.6052	0.8471	0.9383	-0.2690	-0.0120	-1.0282	-1.0996
	Med	0.6168	0.5876	0.7687	0.8344	-0.2359	-0.0417	-0.7464	-0.8531
	SD	0.1902	0.1690	0.3707	0.4294	0.4903	0.5595	0.8070	0.8479
	MSE	0.0712	0.0538	0.5782	0.8866	0.5242	0.7490	2.1389	2.3516

Table 4.17: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_1 Sc_2.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\beta_{23}}$	$\boldsymbol{\beta_{32}}$
Sc2		**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.4843	0.5715	1.0431	1.1963	-0.1735	-0.1098	-0.6453	-0.5599
	Med	0.4485	0.5445	0.9765	1.1369	-0.1813	-0.1140	-0.5941	-0.4644
	SD	0.1775	0.1941	0.3923	0.4380	0.4664	0.4798	0.6551	0.6478
	MSE	0.0399	0.0683	0.5739	0.9740	0.3858	0.4514	1.6137	0.8542
Pr2	**Mean**	0.7337	0.7952	0.6869	0.7303	-0.7976	-0.7520	-0.0980	-0.1322
	Med	0.6931	0.7725	0.6322	0.6830	-0.8207	-0.7315	-0.0930	-0.1264
	SD	0.2551	0.2635	0.2643	0.2861	0.4613	0.4855	0.6104	0.6351
	MSE	0.1986	0.2510	0.1622	0.1953	1.1780	1.1175	0.6918	0.9656
Pr3	**Mean**	0.7956	0.7295	0.6193	0.7642	-0.7879	-0.3891	-0.0708	-0.1116
	Med	0.7495	0.7183	0.5796	0.7287	-0.8526	-0.4482	-0.0408	-0.0891
	SD	0.3019	0.2507	0.2302	0.2762	0.5639	0.6401	0.4416	0.4612
	MSE	0.3062	0.2639	0.0913	0.1772	2.4865	2.3146	0.4144	0.4211
Pr4	**Mean**	0.7060	0.6257	1.2347	1.4304	-0.3501	-0.3838	-1.1642	-1.2785
	Med	0.6562	0.5917	1.1535	1.2850	-0.3661	-0.3842	-1.0992	-1.2174
	SD	0.2650	0.2238	0.5955	0.7196	0.5347	0.6066	0.9762	0.9959
	MSE	0.1526	0.0801	1.4922	2.0869	1.1991	1.7523	3.0716	3.5200

Table 4.18: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_1 Sc_3.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\beta_{23}}$	$\boldsymbol{\beta_{32}}$
Sc3		**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.5118	0.5066	0.4833	0.5349	-0.1512	-0.1425	-0.1632	-0.2214
	Med	0.5106	0.5102	0.4883	0.5371	-0.1454	-0.1379	-0.1583	-0.2039
	SD	0.1041	0.1056	0.1012	0.1119	0.2177	0.2547	0.2976	0.3115
	MSE	0.0170	0.0303	0.0371	0.0276	0.1655	0.2878	0.6359	0.4265
Pr2	**Mean**	0.4981	0.5033	0.5143	0.5893	-0.1938	-0.2705	-0.3989	-0.4467
	Med	0.4985	0.5087	0.5199	0.5927	-0.1985	-0.2676	-0.3977	-0.4430
	SD	0.0984	0.1031	0.1032	0.1213	0.1882	0.1947	0.2346	0.2375
	MSE	0.0158	0.0305	0.0360	0.0423	0.1198	0.1346	0.0828	0.1099
Pr3	**Mean**	0.5610	0.5166	0.4906	0.5825	-0.0973	-0.0317	-0.1474	-0.1099
	Med	0.5687	0.5314	0.4847	0.5793	-0.0963	-0.0355	-0.1236	-0.0921
	SD	0.1384	0.1375	0.0948	0.1102	0.2139	0.2706	0.2901	0.3127
	MSE	0.0331	0.0342	0.0197	0.0367	0.1837	0.3267	0.3216	0.3540
Pr4	**Mean**	0.6075	0.5531	0.5229	0.5550	-0.4437	-0.3205	-0.2423	-0.3374
	Med	0.6088	0.5597	0.5222	0.5550	-0.4312	-0.3056	-0.2366	-0.3064
	SD	0.1276	0.1220	0.0953	0.1040	0.2280	0.2425	0.2838	0.2855
	MSE	0.0362	0.0284	0.0304	0.0342	0.2592	0.2464	0.2876	0.2708

Table 4.19: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_1 Sc_4.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\beta_{23}}$	$\boldsymbol{\beta_{32}}$
Sc4		**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	1.0810	0.9660	1.2531	1.5291	-0.3351	-0.3667	-1.7815	-1.8505
	Med	0.9129	0.8550	1.1291	1.3550	-0.2412	-0.2557	-1.7565	-1.8572
	SD	0.5129	0.4348	0.5130	0.6500	0.9144	0.9067	0.7686	0.8030
	MSE	0.9433	0.5463	1.8902	3.6342	1.7770	1.7320	4.3952	5.1070
Pr2	**Mean**	0.7858	0.6591	0.6691	0.7888	-0.5135	-0.5430	-0.2851	-0.4947
	Med	0.7568	0.6265	0.6010	0.7066	-0.4796	-0.5238	-0.2200	-0.4193
	SD	0.2451	0.2237	0.2632	0.3161	0.4617	0.4616	0.4467	0.4641
	MSE	0.1789	0.1098	0.1427	0.2605	0.7226	1.0658	0.5727	0.5604
Pr3	**Mean**	0.6322	0.5891	0.6979	0.7315	-0.1515	-0.1301	-0.1863	-0.2155
	Med	0.6163	0.5771	0.6025	0.6399	-0.1145	-0.1350	-0.0585	-0.1114
	SD	0.1744	0.1618	0.2740	0.2815	0.4085	0.4430	0.5423	0.5297
	MSE	0.0566	0.0610	0.2203	0.2243	0.4496	0.4609	0.9371	1.1870
Pr4	**Mean**	0.5412	0.6464	0.9467	1.0930	-0.3689	-0.1252	-1.1824	-1.2059
	Med	0.5178	0.6135	0.9289	1.0729	-0.3628	-0.1033	-1.1694	-1.2132
	SD	0.1841	0.2246	0.3315	0.3914	0.3306	0.3371	0.5489	0.5195
	MSE	0.0511	0.1021	0.4324	0.7461	0.2136	0.3673	1.4070	1.5174

Table 4.20: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_2 Sc_1.

Q2	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\beta_{23}}$	$\boldsymbol{\beta_{32}}$
Sc1		**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.3471	1.1287	0.3699	1.4920	-0.3781	-0.7063	-0.4401	-0.3737
	Med	0.3244	1.0626	0.3113	1.2801	-0.3713	-0.7009	-0.4194	-0.2159
	SD	0.1195	0.3734	0.2208	0.7474	0.5152	0.5236	0.8034	0.7961
	MSE	0.0323	0.4294	0.1254	1.9027	0.7217	0.6030	1.8020	2.3600
Pr2	**Mean**	0.3288	1.0888	0.2887	1.1894	-0.3735	-0.2253	-0.1840	-0.3247
	Med	0.3087	1.0216	0.2661	1.1041	-0.3642	-0.2093	-0.1402	-0.3336
	SD	0.1113	0.3540	0.1218	0.4539	0.5463	0.5453	0.4917	0.5675
	MSE	0.0303	0.3395	0.0172	0.7160	0.8164	1.0227	0.8716	1.0075
Pr3	**Mean**	0.3184	1.0581	0.2232	0.8765	-0.7145	-0.7089	-0.9941	-1.3692
	Med	0.2922	0.9853	0.2035	0.8443	-0.6976	-0.6574	-1.1006	-1.6219
	SD	0.1195	0.3835	0.1065	0.3424	0.5970	0.5919	0.8051	1.0287
	MSE	0.0285	0.3500	0.0264	0.2651	1.3274	1.3588	1.4918	2.9390
Pr4	**Mean**	0.3062	0.9316	0.2958	1.3194	-0.0039	-0.0672	-0.5034	-0.8659
	Med	0.2854	0.8781	0.2742	1.2356	-0.0365	-0.1327	-0.4112	-0.8616
	SD	0.1028	0.2948	0.1287	0.4981	0.4697	0.4582	0.8804	0.8852
	MSE	0.0321	0.2762	0.0293	0.7708	0.5567	0.4925	3.5469	5.0017

Table 4.21: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_2 Sc_2.

Q2	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\beta_{23}}$	$\boldsymbol{\beta_{32}}$
Sc2		**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.7521	2.1538	0.4181	1.6871	-0.4967	-0.7369	-0.8885	-0.8383
	Med	0.7020	2.0760	0.3760	1.6876	-0.4501	-0.6855	-0.8845	-0.8847
	SD	0.3396	0.8889	0.2231	0.6975	0.8453	0.8468	0.7431	0.7944
	MSE	0.5814	4.3847	0.0841	2.3349	2.6452	3.1382	2.3171	3.6099
Pr2	**Mean**	0.3647	1.3137	0.4228	1.2348	-1.1165	-0.6515	-0.4213	-0.1344
	Med	0.3162	1.1607	0.3849	1.1752	-1.0083	-0.5606	-0.6331	-0.2400
	SD	0.1710	0.5711	0.2090	0.5326	0.7594	0.7576	1.0543	1.0620
	MSE	0.0559	0.8943	0.0878	0.7529	1.6136	1.0635	4.4644	3.2623
Pr3	**Mean**	0.3138	1.2252	0.9066	1.8292	-0.2455	-0.1096	-0.5207	-0.2366
	Med	0.2745	1.0927	0.7962	1.8625	-0.1788	-0.0864	-0.4504	-0.4160
	SD	0.1516	0.5539	0.5696	0.9430	0.6184	0.5520	0.8235	0.8362
	MSE	0.0404	0.7043	1.3702	4.0935	1.3397	0.9767	2.2154	3.7844
Pr4	**Mean**	0.7298	2.0119	0.4598	1.5000	-1.4448	-1.3014	-0.6579	-1.0521
	Med	0.7416	2.1335	0.4201	1.4855	-1.3165	-1.1258	-0.7351	-1.2570
	SD	0.3339	0.8203	0.2524	0.6917	0.9058	0.8495	1.5437	1.6733
	MSE	0.9093	4.7244	0.1584	1.6142	3.2366	2.5379	6.2915	9.6339

Table 4.22: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_2 Sc_3.

Q2	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\beta_{23}}$	$\boldsymbol{\beta_{32}}$
Sc3		**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.2705	0.7944	0.3151	1.1552	-0.4792	-0.4047	-0.3273	-0.1870
	Med	0.2715	0.7986	0.2850	1.0757	-0.4793	-0.3846	-0.3503	-0.1786
	SD	0.0574	0.1577	0.1299	0.4560	0.2481	0.2443	0.3597	0.4526
	MSE	0.0045	0.0327	0.0347	0.4684	0.1441	0.0889	0.2792	0.6076
Pr2	**Mean**	0.2874	0.8029	0.2632	0.9016	-0.2016	-0.0780	-0.4358	-0.9394
	Med	0.2828	0.7984	0.2512	0.8833	-0.2267	-0.0715	-0.4728	-0.9011
	SD	0.0510	0.1394	0.0829	0.2515	0.2462	0.2689	0.3347	0.3627
	MSE	0.0072	0.0268	0.0163	0.1751	0.2120	0.3417	0.3174	0.8863
Pr3	**Mean**	0.3182	0.9857	0.2255	0.8436	-0.5106	-0.5799	-0.0261	-0.0133
	Med	0.3125	0.9714	0.2124	0.8273	-0.4658	-0.5476	-0.0025	-0.0215
	SD	0.0749	0.2148	0.0780	0.2407	0.3926	0.3821	0.4892	0.5252
	MSE	0.0120	0.1307	0.0136	0.0851	0.4743	0.4465	0.6000	0.7257
Pr4	**Mean**	0.2838	0.9006	0.2072	0.7157	-0.2241	-0.1481	-0.0097	-0.0234
	Med	0.2802	0.8995	0.2010	0.6978	-0.2131	-0.1329	-0.0044	-0.0168
	SD	0.0485	0.1423	0.0633	0.2010	0.2254	0.2173	0.3925	0.3402
	MSE	0.0057	0.0479	0.0087	0.0683	0.1531	0.1610	0.6250	0.5258

Table 4.23: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_2 Sc_4.

Q2	**Par**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Sc4		**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.4354	1.2806	0.5650	1.8784	-0.5172	-0.7881	-0.6000	-0.2648
	Med	0.4060	1.2192	0.5194	1.7587	-0.5351	-0.7786	-0.7117	-0.4003
	SD	0.1759	0.4962	0.2632	0.7996	0.4843	0.4773	0.6649	0.6079
	MSE	0.1125	0.9766	0.2582	3.5004	1.0888	1.1144	2.6563	1.8497
Pr2	**Mean**	0.3434	0.9931	0.3722	1.2446	-0.0469	-0.0074	-0.2452	-0.2474
	Med	0.3254	0.9577	0.3215	1.1074	-0.0368	-0.0200	-0.1174	-0.1496
	SD	0.1006	0.2660	0.1811	0.5685	0.4307	0.4390	0.6581	0.6635
	MSE	0.0232	0.1503	0.0678	0.8007	0.7209	0.6517	1.5157	1.4998
Pr3	**Mean**	0.3106	0.8945	0.3330	1.1041	-0.5733	-0.4369	-0.0022	-0.0389
	Med	0.2999	0.8745	0.3103	1.0544	-0.5466	-0.4081	-0.0806	-0.0003
	SD	0.0807	0.2117	0.1265	0.3407	0.3464	0.3431	0.5132	0.5573
	MSE	0.0122	0.0765	0.0350	0.3565	0.5548	0.3842	0.5473	0.6319
Pr4	**Mean**	0.3905	1.2095	0.1945	0.7726	-0.7957	-0.8672	-0.2401	-0.1831
	Med	0.3666	1.1546	0.1795	0.7382	-0.7876	-0.8346	-0.2620	-0.1946
	SD	0.1370	0.3947	0.0822	0.2867	0.4560	0.4574	0.4724	0.5300
	MSE	0.0589	0.5637	0.0101	0.0905	0.7930	0.7423	0.9139	0.9756

Table 4.24: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_3 Sc_1.

Q3	**Par**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Sc1		**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.8591	0.2718	0.9814	0.3157	-0.7191	-0.7459	-0.0725	-0.2510
	Med	0.8352	0.2563	0.9654	0.3054	-0.8202	-0.8292	-0.0496	-0.2373
	SD	0.2736	0.1082	0.2232	0.0814	0.5174	0.6896	0.3919	0.3923
	MSE	0.1445	0.0314	0.1263	0.0156	1.0830	2.1538	0.5016	0.2937
Pr2	**Mean**	0.9278	0.2860	0.7472	0.2547	-0.4843	-0.7860	-0.4278	-0.3194
	Med	0.8929	0.2681	0.7452	0.2485	-0.4687	-0.7410	-0.4220	-0.3402
	SD	0.2805	0.1088	0.1548	0.0627	0.3839	0.5127	0.3046	0.3463
	MSE	0.2429	0.0211	0.0264	0.0061	0.2338	0.6574	0.4334	0.2058
Pr3	**Mean**	0.9563	0.2533	0.7585	0.3121	-0.2893	-0.8702	-0.0423	-0.0789
	Med	0.8748	0.2258	0.7426	0.3020	-0.2996	-0.7883	-0.0344	-0.0926
	SD	0.3632	0.1238	0.1836	0.0844	0.4610	0.6080	0.3631	0.3748
	MSE	0.3070	0.0294	0.0796	0.0163	0.5402	1.0262	0.3324	0.6490
Pr4	**Mean**	1.4193	0.4026	0.8245	0.2754	-1.0990	-1.1128	-0.0392	-0.0403
	Med	1.3202	0.3658	0.8187	0.2717	-1.0503	-1.0844	-0.0238	-0.0398
	SD	0.5860	0.1845	0.1719	0.0656	0.6439	0.6744	0.3039	0.2931
	MSE	1.2039	0.0651	0.0450	0.0081	1.4820	1.4878	0.4553	0.3111

Table 4.25: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_3 Sc_2.

Q3	**Par**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Sc2		**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	1.6393	0.6811	1.1168	0.3833	-1.6821	-1.3923	-0.4959	-0.4686
	Med	1.6016	0.6246	1.0387	0.3468	-1.7264	-1.5105	-0.5588	-0.5870
	SD	0.6759	0.3519	0.4586	0.1780	0.8112	0.7434	0.5777	0.5984
	MSE	2.2869	0.5153	0.6921	0.0914	3.4428	2.1744	0.9048	0.7102
Pr2	**Mean**	1.6605	0.5311	1.1371	0.4062	-0.6153	-0.2425	-0.4981	-0.2368
	Med	1.5906	0.4973	1.0293	0.3670	-0.5191	-0.2196	-0.5070	-0.2451
	SD	0.6215	0.2269	0.4534	0.1716	0.6756	0.7676	0.4846	0.5129
	MSE	2.5518	0.2223	0.6926	0.0757	1.0131	1.0534	0.6684	0.7946
Pr3	**Mean**	1.9006	0.4743	1.4761	0.6120	-0.0353	-0.2598	-1.0996	-0.6528
	Med	1.6252	0.4121	1.1819	0.4823	-0.0309	-0.2838	-1.0067	-0.5303
	SD	0.8367	0.2621	0.7938	0.3664	0.5906	0.5900	0.6901	0.6931
	MSE	4.3810	0.2225	1.8658	0.4330	0.8370	0.7839	3.0015	2.1936
Pr4	**Mean**	1.1097	0.4999	1.0593	0.4067	-1.4552	-0.9818	-0.2522	-0.0219
	Med	1.0100	0.4356	0.9633	0.3694	-1.3095	-0.8401	-0.2491	-0.0207
	SD	0.5164	0.2674	0.4202	0.1704	0.8015	0.9762	0.4667	0.4727
	MSE	0.7161	0.2503	0.3408	0.0686	4.8556	3.8530	1.0509	0.6557

Table 4.26: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_3 Sc_3.

Q3	**Par**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Sc3		**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.7424	0.2586	0.9096	0.3028	-0.0144	-0.3332	-0.5703	-0.6247
	Med	0.7474	0.2527	0.9062	0.2979	-0.0085	-0.3472	-0.5532	-0.5850
	SD	0.1609	0.0696	0.1916	0.0722	0.3127	0.3671	0.2955	0.3154
	MSE	0.0269	0.0071	0.0821	0.0086	0.5190	1.3627	0.2265	0.2717
Pr2	**Mean**	0.8425	0.2428	0.8301	0.2617	-0.3715	-0.2665	-0.1720	-0.1261
	Med	0.8178	0.2339	0.8267	0.2586	-0.3876	-0.2723	-0.1700	-0.1234
	SD	0.2122	0.0745	0.1419	0.0500	0.4164	0.5333	0.2469	0.2744
	MSE	0.0703	0.0069	0.0793	0.0045	0.4445	0.9521	0.2040	0.1920
Pr3	**Mean**	0.7695	0.2016	0.9273	0.3065	-0.0549	-0.0969	-0.1550	-0.0664
	Med	0.7625	0.1964	0.9131	0.3027	-0.0618	-0.1408	-0.1608	-0.0737
	SD	0.1765	0.0622	0.1862	0.0679	0.2877	0.4493	0.2636	0.2648
	MSE	0.0368	0.0077	0.1004	0.0134	0.4063	0.7276	0.1980	0.3452
Pr4	**Mean**	0.9181	0.2904	0.8461	0.2708	-0.2230	-0.5609	-0.4252	-0.2348
	Med	0.9128	0.2837	0.8376	0.2650	-0.2517	-0.5647	-0.4232	-0.2361
	SD	0.2115	0.0775	0.1520	0.0541	0.4664	0.4272	0.2437	0.2858
	MSE	0.1583	0.0193	0.0544	0.0051	0.4436	0.4215	0.1300	0.1950

Table 4.27: Simulation results of modelling multi-state data using the transition probabilities with a covariate included in the model - Q_3 Sc_4.

Q3	**Par**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Sc4		**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	1.2263	0.3690	1.6916	0.5743	-0.7970	-0.9217	-1.4062	-1.2072
	Med	1.1040	0.3269	1.6655	0.5423	-0.8224	-0.9451	-1.4120	-1.2099
	SD	0.5617	0.1880	0.7504	0.2771	0.6291	0.6310	0.7040	0.7200
	MSE	0.8302	0.0667	2.6528	0.3049	1.3441	1.8096	3.7041	2.7517
Pr2	**Mean**	0.7918	0.2846	0.8433	0.3071	-0.1945	-0.4049	-0.1742	-0.1761
	Med	0.7590	0.2637	0.8345	0.2992	-0.2258	-0.4312	-0.1979	-0.2078
	SD	0.2252	0.1044	0.1789	0.0738	0.3752	0.4328	0.3085	0.3454
	MSE	0.0869	0.0178	0.0709	0.0126	0.2443	0.2895	0.3484	0.4133
Pr3	**Mean**	1.0398	0.3322	0.8451	0.2811	-0.7652	-0.6832	-0.2847	-0.2531
	Med	0.9788	0.3108	0.8319	0.2747	-0.7369	-0.7187	-0.2714	-0.2153
	SD	0.3616	0.1342	0.2191	0.0816	0.4879	0.5178	0.2733	0.2982
	MSE	0.3948	0.0299	0.0919	0.0104	0.9972	1.0311	0.3417	0.2510
Pr4	**Mean**	0.9693	0.3188	0.8398	0.2912	-0.7551	-0.7071	-0.2182	-0.0583
	Med	0.8984	0.2946	0.8332	0.2840	-0.7253	-0.6852	-0.2094	-0.0492
	SD	0.3479	0.1255	0.1983	0.0793	0.5122	0.7338	0.2628	0.2803
	MSE	0.3039	0.0504	0.0748	0.0106	0.7633	1.7195	0.2368	0.6468

Table 4.28: Bias for a select number of models using the transition probabilities with a covariate included in the model.

Model	**Scenario**	**Prior**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Q1	1	1	0.298	0.263	0.451	0.481	0.804	0.716	0.784	1.154
		2	0.176	0.188	0.131	0.175	0.296	0.443	1.040	0.696
		3	0.187	0.161	0.164	0.202	0.294	0.443	0.474	0.599
		4	0.187	0.159	0.664	0.838	0.533	0.660	1.220	1.278
	3	1	0.078	0.138	0.164	0.123	0.344	0.472	0.740	0.574
		2	0.079	0.141	0.159	0.166	0.290	0.311	0.167	0.231
		3	0.118	0.124	0.103	0.157	0.371	0.503	0.487	0.506
		4	0.141	0.116	0.146	0.153	0.455	0.433	0.455	0.435
Q3	1	1	0.264	0.140	0.277	0.094	0.903	1.296	0.590	0.374
		2	0.405	0.096	0.049	0.046	0.294	0.628	0.584	0.293
		3	0.418	0.119	0.214	0.096	0.572	0.810	0.448	0.713
		4	0.928	0.176	0.124	0.061	1.033	1.016	0.602	0.475
	3	1	0.031	0.047	0.213	0.058	0.649	1.108	0.373	0.415
		2	0.159	0.037	0.243	0.044	0.521	0.817	0.378	0.342
		3	0.075	0.062	0.256	0.094	0.569	0.725	0.358	0.524
		4	0.337	0.115	0.177	0.046	0.476	0.489	0.266	0.337

Table 4.29: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals of modelling multi-state data using the transition probabilities with a covariate included in the model - $Q_1\ Sc_1\ Pr_1$.

Interval	**Statistic**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Cred	Cov %	0.8743	0.9341	0.9162	0.8802	0.9123	0.9102	0.9581	0.9281
HPD	Cov %	0.9162	0.9521	0.9281	0.9401	0.9042	0.9341	0.9521	0.9281
Cred	Mean length	0.8870	0.8156	0.7879	0.8835	1.2853	1.2649	1.2514	1.2276
HPD	Mean length	0.8344	0.7636	0.7349	0.8275	1.2319	1.2219	1.1955	1.1759

Table 4.30: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals of modelling multi-state data using the transition probabilities with a covariate included in the model - $Q_3\ Sc_4\ Pr_1$.

Interval	**Statistic**	$\boldsymbol{\lambda}_{12}$	$\boldsymbol{\lambda}_{21}$	$\boldsymbol{\lambda}_{23}$	$\boldsymbol{\lambda}_{32}$	$\boldsymbol{\beta}_{12}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\beta}_{23}$	$\boldsymbol{\beta}_{32}$
Cred	Cov %	0.9498	0.9369	0.9374	0.9587	0.9258	0.9347	0.9325	0.9458
HPD	Cov %	0.9454	0.9376	0.9571	0.9269	0.9145	0.9258	0.9249	0.9365
Cred	Mean length	0.7675	0.7856	0.7958	0.6698	0.9856	0.9875	0.8958	0.9258
HPD	Mean length	0.6756	0.6654	0.7458	0.7697	0.7644	0.8258	0.9649	0.7258

4.4 Illustrative Examples

In this section, the B-MSM's developed in this chapter are used to model a real world data set. The data set, taken from Sharples *et al.* (2003) and Jackson (2011), is a study of the progression of coronary allograft vasculopathy (CAV), a post-transplant deterioration of the arterial walls.

The example will take the following form:

1) Background is given about the data being used and an extract of the data is provided.
2) The multi-state model that will be fitted to the data is presented.
3) The priors that will be used when modelling the data using the limiting probabilities and transition probabilities of the process are given.
4) Posterior model assessment statistics are presented and interpreted (based on 3000 posterior variates for each parameter in the models).
5) The posterior parameter estimates and summary statistics are presented and interpreted for the best model based on the posterior assessment statistics from 4.

4.4.1 Coronary Allograft Vasculopathy (CAV)

The CAV data set, from Sharples *et al.* (2003) and Jackson (2011), consists of 614 individuals that were monitored after undergoing a heart transplant. Sharples *et al.* (2003) studied the progression of coronary allograft vasculopathy (CAV), a post-transplant deterioration of the arterial walls, using these data. Approximately each year after transplant, each patient had an angiogram, at which CAV could be diagnosed. The result of the test is used to classify a patient into 4 different states:

- 1, representing no CAV (denoted by **W**ell here).
- 2, representing mild/moderate CAV (denoted by **M**ild here).
- 3, representing severe CAV (denoted by **S**evere here).
- 4, recorded at the date of death (denoted by **D**eath here).

Also included in the data set as possible covariates influencing the progression of CAV, is the age of the heart transplant donor (DAGE) and, if preoperative ischemic heart disease (IHD) was the primary diagnosis, the reason for the transplant. Table 4.31 contains an extract of the data set for two patients.

Table 4.31: Extract from the CAV data set.

Subject	Time (years)	State	IHD	DAGE
⋮	⋮	⋮	⋮	⋮
6	0	1	1	20
6	2	1	1	20
6	3	1	1	20
6	4	1	1	20
6	5	1	1	20
6	6	1	1	20
6	7	2	1	20
6	11	4	1	20
⋮	⋮	⋮	⋮	⋮
300	0	1	0	26
300	2	1	0	26
300	4	2	0	26
300	6	2	0	26
300	9	2	0	26
300	10	2	0	26
300	11	4	0	26
⋮	⋮	⋮	⋮	⋮

The underlying multi-state model assumed by these authors is (Jackson, 2011)

$$Q = \begin{bmatrix} -(\lambda_{WM} + \lambda_{WD}) & \lambda_{WM} & 0 & \lambda_{WD} \\ \lambda_{MW} & -(\lambda_{MW} + \lambda_{MS} + \lambda_{MD}) & \lambda_{MS} & \lambda_{MD} \\ 0 & \lambda_{SM} & -(\lambda_{SM} + \lambda_{SD}) & \lambda_{SD} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The models discussed in this chapter are only applicable to 3-state models (their extension to 4-state models will form part of future research). Due to this, the data set is altered by combining all severe and death state observations into a single state 3. This gives the states and the corresponding multi-state model that will be used here as:

- 1, representing no CAV (denoted by **W**ell here).
- 2, representing mild/moderate CAV (denoted by **M**ild here).
- 3, representing severe or death CAV (denoted by **S**evere here),

and

$$Q = \begin{bmatrix} -\lambda_{WM} & \lambda_{WM} & 0 \\ \lambda_{MW} & -(\lambda_{MW} + \lambda_{MS} + \lambda_{MD}) & \lambda_{MS} \\ 0 & \lambda_{SM} & -\lambda_{SM} \end{bmatrix}.$$

The same 6 priors (two for the LP and 4 for the TP models) introduced and used in Section 4.3.1.1 are used as prior distribution for the parameters in this example.

The DIC and goodness-of-fit (GOF) values for the two B-MSM's and different priors are given in Tables 4.32 and 4.33. The GOF values presented here can be compared to the frequentist model value of 1362.19 ($p < 0.0001$). All GOF-values in Table 4.33 have p-values that have an upper limit that are smaller than 0.0001, based on an upper limit of the degrees of freedom of 70 (see Section 2.2.2).

These results indicate that the best fitting B-MSM is the model using the transition probabilities with covariates included in the model and using prior 4. It must however be noted that, similar to the results found by Sharples *et al.* (2003) and Jackson (2011) for the 4-state frequentist model, the p-values for the GOF statistic are extremely small, indicating a lack of fit for all these models. Jackson (2011) attributed this to the fact that the underlying multi-state model may not be strictly medically realistic and that a more complex pattern of time-dependence, or allowing the transition intensities to depend on covariates, would be expected to yield a better fit.

Table 4.32: DIC values for the10 B-MSM fitted to the reduced CAV data.

Prior	**Jeff**	**MDI**		
LP	3122.073	2923.668		
Prior	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**
TP-NoCov	2721.171	2720.632	2720.660	2720.206
TP-Cov	2701.011	2701.931	2705.597	**2697.537**

Table 4.33: Goodness-of-fit values for the10 B-MSM fitted to the reduced CAV data.

Prior	**Jeff**	**MDI**		
LP	1907.115	1650.113		
Prior	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**
TP-NoCov	1362.625	1363.018	1362.441	1363.703
TP-Cov	1361.496	1362.779	1362.594	**1356.804**

The summary statistics for the TP-Cov B-MSM of the posterior distributions (based on 3000 posterior variates) for the 8 parameters in the model are presented in Table 4.34. The posterior and frequentist hazard ratios, posterior prediction matrices and posterior predictive distributions are presented in Tables 4.35 to 4.37 and Figures 4.1 to 4.6.

Similar conlcusions are drawn from the hazard rates for both the B-MSM and the frequentist model: IHD significantly increases the hazard of CAV onset (63% increase for the Bayesian

model vs. 56% increase for the frequentist model), while none of the other rates are significantly influenced by IHD.

The prediction matrices give the probability of being in the three different states of the disease after one year in the study, for patients with and without IHD being the primary diagnosis. These are calculated as the mean values of the posterior predictive distributions that are presented in Figures 4.1 to 4.6. Both matrices (and the corresponding distributions) are very similar, but they do show that patients with primary diagnosis of IHD have a higher probability of being in the severe state after one year in the study.

As the GOF-values for all models (Bayesian and frequentist) indicate a lack of fit for the models it is important not to over interpret the results. The example is presented to illustrate the modelling process, rather than finding and interpreting the best model for this data set.

Table 4.34: Posterior summary for CAV data with covariates - TP (Pr_3).

TP (Prior 4)	**Mean**	**Med**	**SD**	**Perc$_L$**	**Perc$_U$**	**Perc$_{\bar{x}}$**	**HPD$_L$**	**HPD$_U$**	**HPD$_{\bar{x}}$**
λ_{WM}	0.1684	0.1664	0.0107	0.1493	0.1878	0.0385	0.1493	0.1871	0.0378
λ_{MW}	0.2744	0.2824	0.0481	0.1936	0.3485	0.1550	0.1936	0.3482	0.1546
λ_{MS}	0.5278	0.5390	0.0432	0.4389	0.6240	0.1851	0.4588	0.6240	0.1652
λ_{SM}	0.0966	0.0922	0.0279	0.0532	0.1801	0.1269	0.0472	0.1611	0.1139
β_{WM}^{IHD}	0.4898	0.4817	0.1006	0.2903	0.7139	0.4236	0.2833	0.6566	0.3733
β_{MW}^{IHD}	0.1635	0.1857	0.2189	-0.1952	0.5634	0.7586	-0.1952	0.4956	0.6908
β_{MS}^{IHD}	0.1171	0.1289	0.1151	-0.1587	0.3316	0.4903	-0.1587	0.2689	0.4277
β_{SM}^{IHD}	-0.1071	-0.0798	0.2187	-0.5997	0.2604	0.8601	-0.5997	0.2071	0.8068

Table 4.35: Posterior and Frequentist models hazard ratios (95% HPD and 95% CI) for CAV data.

	HR$^{\text{TP}}$	**HPD$_{95\%}^{L}$**	**HPD$_{95\%}^{U}$**	**HR$_{95\%}^{\text{Freq}}$**	**LL$_{95\%}^{\text{Freq}}$**	**UL$_{95\%}^{\text{Freq}}$**
β_{WM}^{IHD}	**1.632**	**1.337**	**2.042**	**1.564**	**1.212**	**2.017**
β_{MW}^{IHD}	1.178	0.823	1.757	0.999	0.543	1.839
β_{MS}^{IHD}	1.124	0.853	1.393	0.933	0.666	1.307
β_{SM}^{IHD}	0.898	0.549	1.297	0.485	0.159	1.483

Table 4.36: One year prediction matrix. Primary diagnosis - IHD.

IHD	**Well**	**Mild**	**Severe**
Well	0.827	0.131	0.042
Mild	0180	0.457	0.363
Severe	0.010	0.060	0.931

Table 4.37: One year prediction matrix. Primary diagnosis - non-IHD.

Non-IHD	Well	Mild	Severe
Well	0.889	0.087	0.024
Mild	0.168	0.497	0.335
Severe	0.009	0.068	0.922

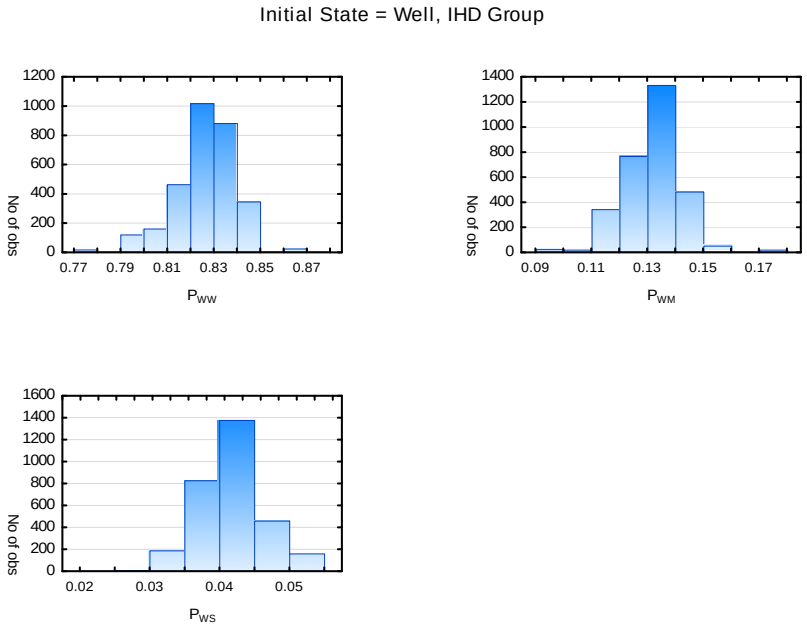

Figure 4.1: One year posterior predictive distributions for CAV data (Initial state = Well, IHD Group).

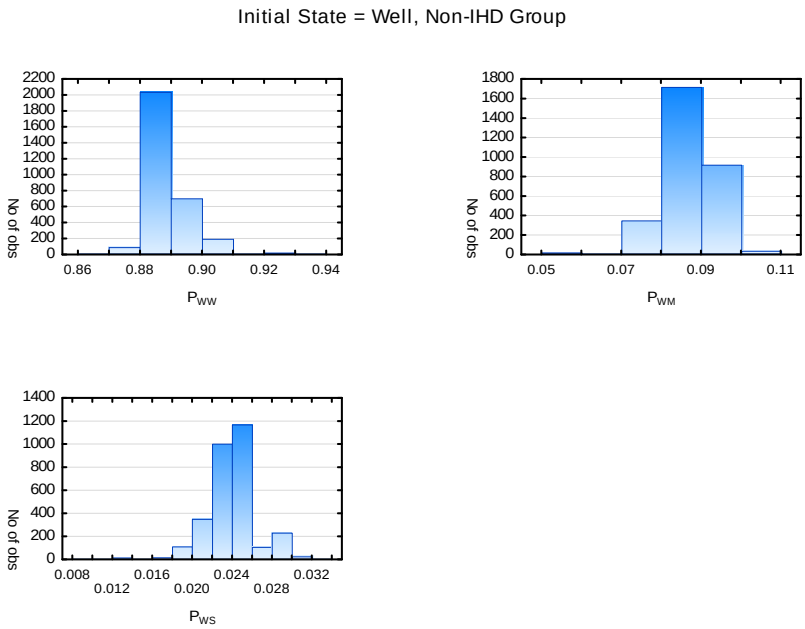

Figure 4.2: One year posterior predictive distributions for CAV data (Initial state = Well, Non-IHD Group).

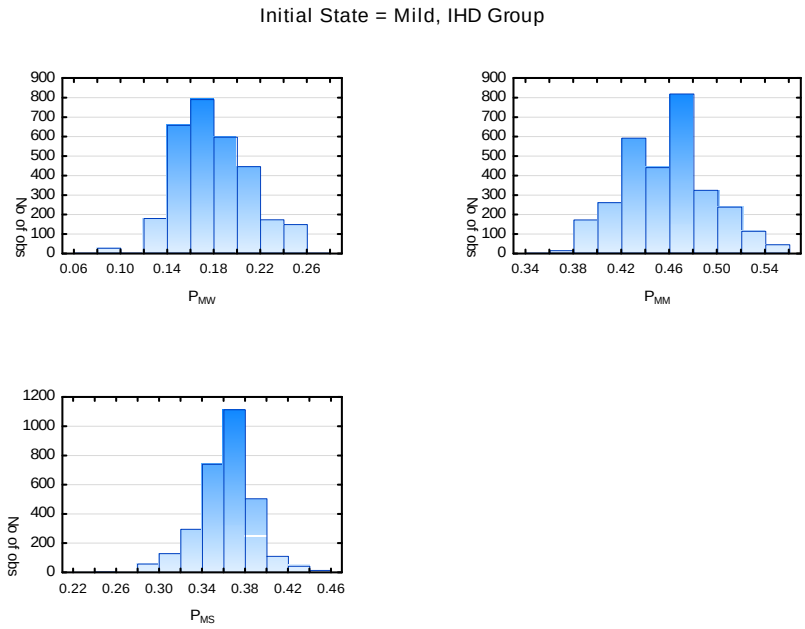

Figure 4.3: One year posterior predictive distributions for CAV data (Initial state = Mild, IHD Group).

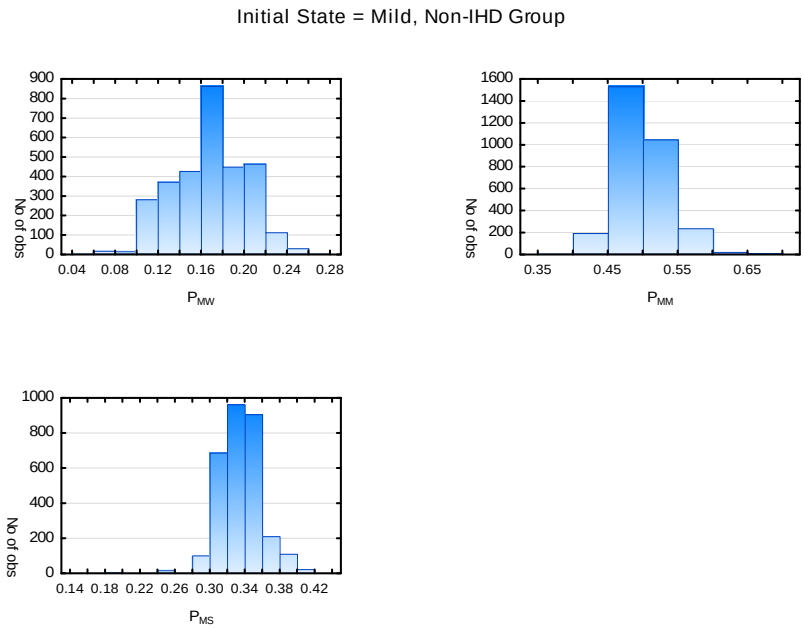

Figure 4.4: One year posterior predictive distributions for CAV data (Initial state = Mild, Non-IHD Group).

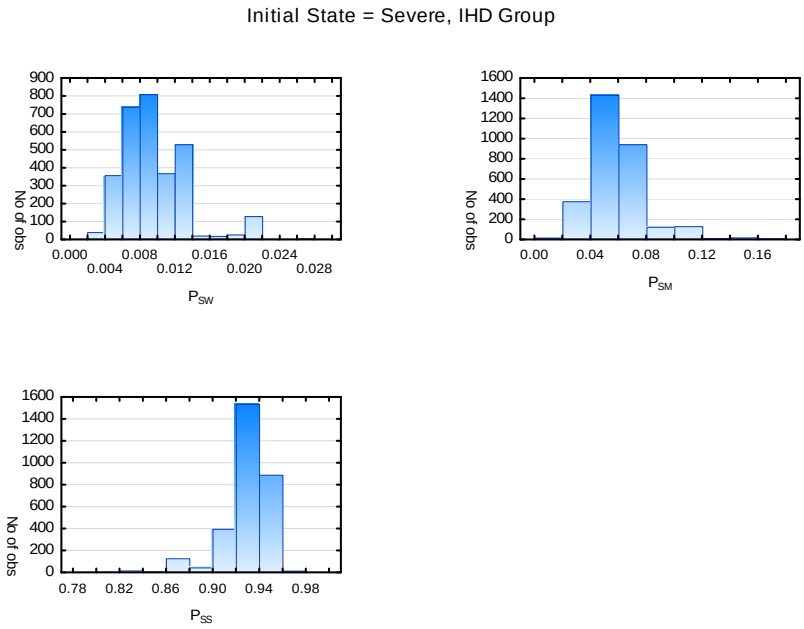

Figure 4.5: One year posterior predictive distributions for CAV data (Initial state = Severe, IHD Group).

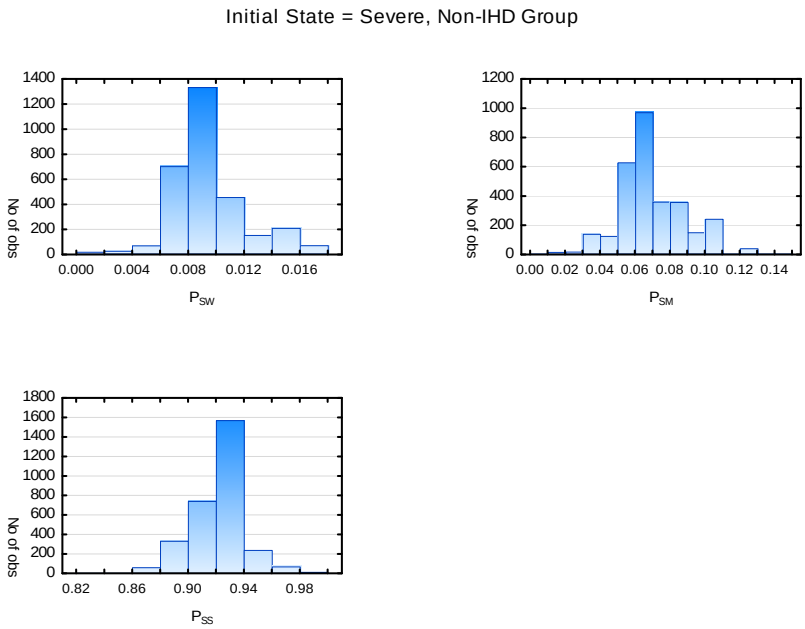

Figure 4.6: One year posterior predictive distributions for CAV data (Initial state = Severe, Non-IHD Group).

4.5 Conclusion

In this chapter, two different Bayesian approaches of modelling panel data using multi-state models were introduced.

In Section 4.1 a model was developed where the likelihood is written in terms of the long term probabilities of the multi-state process. The Jeffreys and the MDI priors were used as prior distributions for the long term probabilities and a Metropolis-Hastings algorithm was presented as an algorithm to sample values from the posterior distributions. Although this B-MSM can only been seen as a summary technique (the time between transitions is lost by only using the limiting probabilities in the model), it was shown in Section 4.3 that this method does yield results that are comparable (with regards to the MSE values) to those found in Section 2.3.3 where maximum likelihood estimates were calculated. One advantage of this B-MSM is that, irrespective of the sample size being used, no model presented itself with the instability of the parameters estimates, as was observed in Section 2.3.3.

In Section 4.2, a model was developed using the transition rates themselves in the likelihood function. The exponential distribution was used as prior distribution for the transition rates and a Metropolis-Hastings algorithm was utilised to sample variates from the posterior distribution. This model was also extended to incorporate covariates into the modelling process. This second B-MSM can be viewed as a more complete model in that the transition probabilities, which are functions of the transition rates, are modelled directly and the time between each transition is incorporated into the model. These models were generally found to have slightly larger MSE values than the models found in Section 2.3.3. It was however shown in Section 4.3 that if a prior distribution is selected that closely matches the underlying population rates, the MSE's were on par with the model based on the maximum likelihood estimates of the parameters. Unlike the maximum likelihood models, where extremely large MSE and standard deviation values were observed for smaller sample sizes, the B-MSM did not present itself with any of these large values. This clearly indicates that incorporating prior information into the modelling process allows one to fit multi-state models to smaller data sets than what is possible without incorporating prior information.

To illustrate the process of fitting B-MSM's to data, the models developed in this chapter were fitted to a published data set in Section 4.4. The results of the Bayesian model, selected based on DIC and GOF statistics, were compared to those of the corresponding frequentist model.

In Chapter 5 Bayesian imputation techniques are presented, where prior information is incorporated directly into a panel of data. A maximised likelihood multi-state model is then fitted to the imputed and enlarged panel data sets.

5

Bayesian Multi-State Imputing

Bayesian multi-state models where priors were placed on the parameters of the different models were introduced in the previous chapter. In this chapter priors will not be placed directly on the parameters of the model, but will rather be used to impute missing observations in the panel data itself. Once the data has been imputed using the proposed Bayesian multi-state imputation (B-MSI) techniques, the data will be analysed using the multi-state models introduced in Chapter 2.

Two Bayesian multi-state imputation techniques are discussed in this chapter.

Assume a 3-state progressive model with monthly follow-up visits. A typical patient's data consisting of the known observations, the observation times (in months) and the states, is shown in Table 5.1.

Table 5.1: Patient data with missing observations

Observation	1	2	3
Time	0	4	7
State	1	2	3

As the patient is not observed every month there are time-gaps between the observations where no states are recorded. In the 4 month period between observations 1 and 2, this patient made a transition from state 1 to state 2, and in the 3 month period between observations 2 and 3 there was a transition from state 2 to state 3. The exact times of these transitions are not known and it could have occurred at any time during the 4 and the 3 month periods. The aim of the techniques proposed in this chapter is to model the time of the transition or even the transition itself using prior knowledge about the disease progression.

The first imputation technique, discussed in Section 5.1, uses prior probability vectors to impute all missing observations. At each missing observation prior vectors obtained from clinical experts are used to generate the parameters of a Dirichlet distribution. From this

distribution, the parameters of a multinomial distribution are sampled and the multinomial distribution is then used to assign a state to the missing observation. Table 5.2 shows the final data matrix for this patient after all missing observations, I_1 to I_5, were imputed.

Table 5.2: Imputing missing observations

Observation	1	I_1	I_2	I_3	2	I_4	I_5	3
Time	0	**1**	**2**	**3**	4	**5**	**6**	7
State	1	**1**	**2**	**2**	2	**2**	**3**	3

The second imputation technique, discussed in Section 5.2, uses prior transition functions as the base functions for a Dirichlet process to impute the transition point between the known observations. In an interval where a known transition has taken place, the Dirichlet process is used to assign the point in the interval where the transition could have occurred. In Table 5.3, for example, the transition from state 1 to state 2 in the $(0; 4]$ time interval is imputed as having occurred at time 3 and the transition from state 2 to state 3 in the $(4; 7]$ interval, is imputed as having occurred at time 5.

Table 5.3: Imputting transition point.

Observation	1	I_1	2	I_2	3
Time	0	**3**	4	**5**	7
State	1	**2**	2	**3**	3

Both techniques lead to larger imputed data sets that are then analysed using the multi-state modelling techniques described in Chapter 2.

The theoretical underpinnings of each technique are firstly discussed in Sections 5.1 and 5.2.

In Section 5.3, an example is given to illustrate the different steps associated with both imputation techniques that were discussed in the theory sections.

In Section 5.4, an extensive simulation study is undertaken to assess the performance of these techniques under different models and data scenarios.

The chapter concludes with Section 5.5, where two published multi-state data sets are analysed using the two B-MSI techniques and the results are compared to those found by previous authors.

5.1 Imputing all unknown observations

The first approach is to incorporate clinicians' insight of disease progression into a multi-state model and to impute the state at each possible time point between two known visits. According to the natural order of the disease states (given a previous state), any neighbouring state can be visited by patients in the following time interval. In the setting of continuous-time Markov modelling, let the unique recorded visiting times for all patients be $t_0 < t_1 < ... < t_m$. A justification for considering this grid of time points where no disease state was observed is, that given the data, there remains a probability of transitioning at these time points. As the likelihood (2.17) is in essence a multinomial likelihood, a conjugate prior in this setting is the Dirichlet distribution which is at the base of imputing disease states and is described subsequently.

Assume that two clinicians have sufficient information regarding disease progression and that at each of the visiting times with an unknown state, they indicate the probability of a patient being in each possible state, given the previous state and/or other covariates or explanatory variables. Denote a $(k \times 1)$ probability vector

$$\mathbf{q}_i' = [q_{i1}, ..., q_{ik}], \tag{5.1}$$

where $q_{ij} = P(\text{in state } j \text{ at current time} \mid \text{in state } i \text{ at previous time}), i, j = 1, ..., k$ and $\sum_{j=1}^{k} q_{ij} = 1$. From (5.1), let $\mathbf{q}_i^1$ and $\mathbf{q}_i^2$ denote two clinicians' prior belief of the transitions in terms of probabilities, then two measures

$$a_{ij} = \frac{q_{ij}^1 + q_{ij}^2}{2}, \ \forall i, j \tag{5.2}$$

an average transition probability, and

$$d_{ij} = \left| q_{ij}^1 - q_{ij}^2 \right|, \ \forall i, j$$

a difference between the transition probability can be defined to describe their accuracy and variation (Congdon, 2002, pp. 37-41).

The underlying assumption is that the two $(k \times 1)$ given probability vectors, $\mathbf{q}_i^1$ and $\mathbf{q}_i^2$, each have a Dirichlet distribution with parameters $\alpha_{i1}, ..., \alpha_{ik}$, i.e.

$$\mathbf{q}_i^1 \sim Dirichlet(\alpha_{i1}, ..., \alpha_{ik}) \text{ and } \mathbf{q}_i^2 \sim Dirichlet(\alpha_{i1}, ..., \alpha_{ik}),$$

with $\sum_{j=1}^{k}\alpha_{ij} = c_i$, which means that if

$$R_i = \sum_{j=1}^{k}(q_{ij}^1 - q_{ij}^2)^2,$$

then it follows that

$$E[R_i] = 2\sum_{j=1}^{k} var(q_{ij}^1) = 2\sum_{j=1}^{k} var(q_{ij}^2),$$

since q_{ij}^1 and q_{ij}^2 are independent and identically distributed. It can easily be shown that

$$E[R_i] = 2\sum_{j=1}^{k}\frac{\alpha_{ij}(c_i - \alpha_{ij})}{c_i^2(c_i+1)}$$

and hence

$$E[R_i] = \frac{2}{c_i+1}(1 - \sum_{j=1}^{k}\eta_{ij}^2), \tag{5.3}$$

with $\eta_{ij} = \frac{\alpha_{ij}}{c_i}$ (Congdon, 2002, pp. 37-41). This (5.3) is now estimated by

$$\widehat{E[R_i]} = \frac{2}{c_i+1}(1 - \sum_{j=1}^{k}\widehat{\eta}_{ij}^2), \tag{5.4}$$

where $\widehat{\eta}_{ij} = a_{ij}$ and $\widehat{E[R_i]} = \sum_{j=1}^{k} d_{ij}^2$. Using (5.4), c_i can now be obtained and it can also be regarded as a prior weight in terms of the number of patients.

A Dirichlet prior can now be implemented sequentially to model disease states at all time points where no disease state information is available for a patient. These probabilities from the Dirichlet distribution enrich the likelihood (2.17) in a Bayesian setting by altering each patient's contribution to the likelihood according to the probabilities assigned by the clinicians.

5.1.1 Imputation methodology

The following scheme is used to impute all missing observations and to estimate the posterior distribution of the transition rates:

1) Let $t_0 < t_1 < ... < t_m$ be the unique observed times that form the grid of points where the disease states will be simulated in the sample of n patients. Some patients will therefore have an unobserved disease state for some t_l.
2) At each t_l for which no disease state was observed for a particular patient, generate a vector of transition probabilities $\boldsymbol{\theta}$, $\{\theta_j, j = 1, ..., k\}$, from a Dirichlet distribution with parameters,

$ca_{i1}, ..., ca_{ik}$, (5.2) depending on the disease state i at t_{l-1}.

3) Generate a disease state, $S(t_l)$ $(S(t_l) \in S = \{1, ..., k\})$, from a multinomial distribution with parameter $\boldsymbol{\theta}$.

4) Repeat steps 2 and 3 sequentially until all time points have a simulated disease state across all patients.

5) Fit a multi-state model to the updated data matrix and estimate the transition rates using the methods described in Section 2.1.2.1.

6) Repeat this process N times. This will provide Monte Carlo integration over the unobserved states and hence, posterior distributions of the transition rates are obtained.

The likelihood in (2.17) is enriched by a Bayesian imputation of disease states, which means that all n_{ijl} are non-zero for all observed $t_0 < t_1 < ... < t_m$. This scheme is illustrated by means of an example in Section 5.3.

5.2 Estimating the transition point

The second approach to incorporating clinicians' insight of disease progression into a multi-state model is to model the transition time between two visits. Let $t_0 < t_1 < ... < t_m$ be the unique visiting times in the data set and assume that some of these times have no recorded disease state for certain patients, i.e. a patient was not seen at that specific time point. Denote $S(t_l)$ $(S(t_l) \in S = \{1, ..., k\})$ as the observed disease state of a patient at visit l and time $t_l, l = 1, ..., m$. Assume a patient seen at time t_l is in state $S(t_l)$ and when next seen at time t_j the patient is in state $S(t_j)$, with $S(t_l) \neq S(t_j)$,

State		$S(t_l)$				$S(t_l)$	
Time	...	t_l	...	t_r	...	t_j	...

.

(It should be noted that transition times of disease states are only modelled if $S(t_l) \neq S(t_j)$ for two subsequent visits l and j.) If $S(t_j) = S(t_l) \pm 1$ the transition from $S(t_l)$ to $S(t_j)$ had to have happened in one of the $(j - l)$ time intervals between t_i and t_j. If $S(t_j) > S(t_l) + 1$ or $S(t_j) < S(t_l) - 1$, more than one transition had to have happened between time t_l and t_j. For the purposes of this study we assume that a patient can only move through the states in one direction in an interval between two known observations, i.e. patients can only move from state 1 to state 2 to state 3 etc. and not from state 1 to state 2 and then back to state 1 in an

interval between two observations. For example, if $S(1) = 1$ and $S(4) = 2$, the patient must have changed to state 2 during one of time intervals $(1;2], (2;3]$ or $(3;4]$, say $(2;3]$

State	1		**2***	2			1
Time	1	2	3	4	5	6	7

.

If $S(1) = 1$ and $S(4) = 3$, the patient must have changed to state 2 during one of time intervals $(1;2], (2;3]$ or $(3;4]$, say $(1;2]$, and then must have changed to state 3 during one of the time intervals $(2;3], (3;4]$, say $(2;3]$

State	1	**2***	**3***	3			1
Time	1	2	3	4	5	6	7

.

Assume now a random distribution G, distributed according to a Dirichlet process (DP), governs the transition between two known states $S(t_l)$ and $S(t_j)$. G is a Dirichlet process distributed with base distribution $F(t)$ and weight parameter γ, written as $G \sim DP(\gamma, F)$, if

$$(G(A_1), ..., G(A_{j-l})) \sim Dir(\gamma F(A_1), ..., \gamma F(A_{j-l})),$$

for every finite measurable partition $A_1, ..., A_{j-l}$ of the space $\Omega = A_1 \cup A_2 \cup \ldots \cup A_{j-l}$ (Hjort *et al.*, 2010, pp. 36-47). The base distribution can be thought of as the mean, since for any measurable set A we have $E[G(A)] = F(A)$, and the weight parameter, γ, can be understood as an inverse variance quantity (Hjort *et al.*, 2010, pp. 36-47)

$$Var[G(A)] = \frac{F(A)\{1 - F(A)\}}{(\gamma + 1)}.$$

The larger γ, the smaller the variance, and the DP will concentrate more of its mass around the mean (Hjort *et al.*, 2010, pp. 36-47). The weight parameter is also called the strength parameter, referring to the strength of the prior when using the DP as a nonparametric prior over distributions in a Bayesian nonparametric model, and the mass parameter, as the prior strength can be measured in units of sample size (or mass) of observations (Hjort *et al.*, 2010, pp. 36-47).

The $(j - l)$ intervals between the two known states $S(t_l)$ and $S(t_j)$ is regarded as a measurable partition $A_1, ..., A_{j-l}$ and the vector $(G(A_1), ..., G(A_{j-l}))$ is random since G is random. The Dirichlet process is utilised to model the transition time of a disease state by incorporating the h (h is the possible number of transitions in the multi-state model under study defined in 4.2) possible transition functions, $T_{rs}(t)$ $(r, s = 1, ..., k)$, monotonic functions over the interval

$(l, j]$, that capture the likelihood of a patient in state $S(0) = r$ being in state $S(t) = s$ after time t, as the base functions of the DP. Following a general trend often utilised in survival analysis, the transition distribution (from this point onwards the base function will be referred to as the transition function of the process) is assumed to be Weibull distributed

$$F(t) = T_{rs}(t) = 1 - e^{-\left(\frac{t}{\kappa}\right)^{\alpha}}$$

with given scale, κ, and shape, α, parameters; mainly due to its versatile hazard rate shape. (Ibrahim *et. al.*, 2001, pp. 102-103).

The general scheme of imputing disease states at any of these $(j-l)$ time intervals is discussed in Section 5.2.1, some characteristics of these transition functions, the base functions of the DP, and how to apply them is discussed in Section 5.2.2, an example explaining the different steps in the process is given in Section 5.3 and the simulation plan that will be used to assess the properties of the proposed imputing scheme is discussed in Section 5.4.

5.2.1 Imputation methodology

The following scheme is used to impute the transition time point between known observations and to estimate the posterior distribution of the subsequent transition rates:

1) For each patient, identify all the time intervals, $(t_l; t_j]$, in the data set where $S(t_l) \neq S(t_j)$.
2) Divide each of the intervals in step 1 into $(j-l)$ time partitions, $(t_l; t_{l+1}], (t_{l+1}; t_{l+2}], ..., (t_{j-1}; t_j]$ (a finite set of partitions).
3) Define the appropriate base/transition functions, $T_{rs}(t)$ for all $r \neq s$, and $r, s = 1, ..., k$, by generating the shape parameter from a normal distribution and setting the scale parameter equal to the length of the interval under study (see Section 5.2.2 for more on the transition functions).
4) Calculate for each of these $(j-l)$ intervals a β_w^m, $w = 1, ..., (j-l)$, where

$$\beta_w^m = \gamma F(A_w^m) = \begin{cases} \gamma |T^m_{S(t_l)[S(t_l)+1]}(t_{w-1}) - T^m_{S(t_l)[S(t_l)+1]}(t_w)|, & \text{if } S(t_l) < S(t_j) \text{ and} \\ \gamma |T^m_{S(t_l)[S(t_l)-1]}(t_{w-1}) - T^m_{S(t_l)[S(t_l)-1]}(t_w)|, & \text{if } S(t_l) > S(t_j) \end{cases} \tag{5.5}$$

where γ can be regarded as a prior weight expressed as the number of patients (see Figures 5.7 to 5.9 for a graphical representation of these values for $\gamma = 1$) and $m = 1, ..., g$.
5) Generate a vector of transition probabilities $\boldsymbol{\theta}^m$, $\{\theta_w^m, w = 1, ..., (j-l)\}$, from a Dirichlet

distribution with parameters, $\beta_1^m, ..., \beta_{j-l}^m$.

6) Repeat steps 3 to 5 g times.

7) Calculate the mean $((j-l)\times 1)$ transition probability vector $\bar{\boldsymbol{\theta}}$, with $\bar{\theta}_w = \frac{1}{g}\sum_{m=1}^{g}\theta_w^m, w = 1, ..., (j-l)$.

8) Generate z that indicates in which one of the $(j-l)$ intervals the appropriate transition occurred, from a multinomial distribution with parameters $\bar{\boldsymbol{\theta}}$, with $0 \leq z \leq (j-l)$.

9) At the interval determined in step 8, set

$$S(t_{l+z}) = \begin{cases} S(t_l)+1, & \text{if } S(t_l) < S(t_j) \text{ and} \\ S(t_l)-1, & \text{if } S(t_l) > S(t_j). \end{cases}$$

10) If $S(t_{l+z}) \neq S(t_j)$, repeat the process (go to step 2) for the remaining intervals, $(t_{l+z}; t_j]$.

11) Fit a multi-state model to the updated data matrix and estimate the transition rates using the methods described in Section 2.1.2.1.

12) Repeat steps 1 to 11 N times. This will provide Monte Carlo integration over the unobserved states and hence, posterior distributions of the transition rates are obtained.

This scheme is illustrated by means of an example in Section 5.3.

5.2.2 The transition function

The prior information about the multi-state model is incorporated into the modelling process via the transition function, $T_{rs}(t)$ $(r, s = 1, ..., k)$, of the Dirichlet process. This transition function is used to assign weights, the parameters β_w, $w = 1, ..., (j-l)$ in step 4 of section 5.2.1, to each one of the possible intervals between two observations where a transition occurred. Larger weights for an interval indicate that a patient is more likely to make a transition from one state to another within that specific interval.

In this study three types of transition functions are proposed. These three functions differ with regards to how the weights are assigned across the interval $(t_i; t_j]$. The three types of functions are plotted in Figures 5.1 to 5.3. T_B-functions (Figure 5.1) place more weight at the beginning of an interval and this type of function is used if a clinician suspects that a patient will move quickly from one state to the next. T_F-functions (Figure 5.2) are "flat" or non-informative functions that assign more or less equal weights to all intervals between t_i and t_j. This type of function is used if a clinician is uncertain as to where in the interval a patient

will move from one state to the next. T_E-functions (Figure 5.3) place most of the weight at the end of an interval and this type of function is used if a clinician suspects that a patient will stay in one state for a prolonged period of time before moving to the next state.
The transition functions are Weibull functions with differing scale, κ_{rs}, and shape, α_{rs}, parameters

$$T_{rs}(t) = 1 - e^{-\left(\frac{t}{\kappa}\right)^{\alpha}}, \ r, s = 1, ..., k. \tag{5.6}$$

The scale parameter is set to the length of the interval between the two known transition points. In Figures 5.1 to 5.3, a scale parameter of 5 is used and in Figures 5.4 to 5.6, a scale parameter of 3 is used. These figures show that the scale parameter ensures that the shape of the function stays the same across differing lengths of the intervals. It is thus possible to interpret the transition function in the exact same way, irrespective of the length of the time interval between two known observations.
The shape parameter controls how the distribution is spread across the interval. A shape parameter less than 1 is used if the probability of a transition is higher at the beginning of the interval (Figures 5.1, 5.4 and 5.7), a parameter between 1 and 1.5 is used if the probability is evenly distributed across the interval (Figures 5.2, 5.5 and 5.8) and a parameter larger than 1.5 is used if the probability of a transition is higher near the end of the interval (Figures 5.3, 5.6 and 5.9). In the simulation process the shape parameter is assumed to be a hyper-parameter and is sampled from a normal distribution

$$\alpha \sim nor(\mu_\lambda, \sigma_\lambda^2).$$

The parameters of the normal distribution, μ_λ and σ_λ^2, are chosen based on the type of transition function required (T_B, T_F or T_E), with $\mu \in \left\{\mu_{T_B}, \mu_{T_F}, \mu_{T_E}\right\}$ and $\sigma^2 \in \left\{\sigma^2_{T_B}, \sigma^2_{T_F}, \sigma^2_{T_E}\right\}$. As the transition function is defined by the hyper-parameter selected from a normal distribution, a sample of functions is used for each interval (steps 3 to 6 in Section 5.2.1). This is done so as to negate the effect of selecting just one value for the hyper-parameter. The standard deviation of the normal distribution, σ_λ, quantifies the uncertainty present in the choice of the shape parameter. The higher the level of uncertainty about the type of transition function to use, the bigger the standard deviation that will be used in the sampling process.
If it is believed that a patient would move quickly from one state to the next, g of these

T_B-functions are used. At each iteration of step 3 the shape parameter for the transition function is sampled from a normal distribution with $0 < \mu_{T_B} < 1$ (the hyper-distribution for T_B functions). In steps 4 and 5 the specific T_B-function is used to generate parameters $\beta_1^m, ..., \beta_{j-l}^m$, and in turn these are used in the Dirichlet distribution to generate a vector of probabilities, $\{\theta_w^m, w = 1, ..., (j-l)\}$, for $m = 1, ..., g$. These steps are repeated and at the end of the loop an average $((j-l) \times 1)$ vector of transition probabilities, $\bar{\boldsymbol{\theta}}$ with $\{\bar{\theta}_w = \frac{1}{g}\sum_{m=1}^{g}\theta_w^m, w = 1, ..., (j-l),\}$, is used in the multinomial distribution to generate an event indicating in which one of the $(j-l)$ intervals the transition occurred.

Table 5.4 summarises the effect of using shape parameters sampled from different normal distributions. In this table four possible transition intervals between the two known observations are assumed. Steps 3 to 7 of Section 5.2.1 are repeated $g = 200$ times for each interval and the results presented in Table 5.4. It is clear from Table 5.4 that as the shape parameter of the Weibull distribution is increased, i.e. moving from a T_B to a T_F to a T_E-function, the probability of an event happening within a specific interval increases from left to right. Using a small shape parameter ($\mu_\lambda = 0.1$) or T_B-function, the probability of the event occurring in the first interval is 90.3% compared to 5.4% in the second, 2.6% in the third and 1.7% in the fourth interval. As the shape parameter is increased and a T_F-function is used ($\mu_\lambda = 1.3$) the probability of the event occurring in the first interval is 24.4%, in the second it is 28.5% , in the third 26.4% and in the fourth interval it is 20.7%. Finally using a T_E-function ($\mu_\lambda = 7$) the probability has shifted towards the fourth interval and the probability of the event occurring in the first interval is 0.0%, in the second interval 1.7%, in the third interval 19.8% and in the fourth interval it is 78.5%.

In the remainder of this study T_B-functions are Weibull functions with the shape parameter sampled from a $nor(0.3, 0.05)$ distribution, T_F-functions are Weibull functions with the shape parameter sampled from a $nor(1.1, 0.05)$ distribution and T_E-functions are Weibull functions with the shape parameter sampled from a $nor(8, 0.75)$ distribution.

Table 5.4: The effect of differing transition functions.

Interval	$(0;1]$	$(1;2]$	$(2;3]$	$(3;4]$
$\mu_\lambda = 0.1$	0.903	0.054	0.026	0.017
$\mu_\lambda = 0.4$	0.685	0.153	0.091	0.071
$\mu_\lambda = 0.6$	0.536	0.201	0.166	0.097
$\mu_\lambda = 0.8$	0.415	0.271	0.181	0.133
$\mu_\lambda = 1.1$	0.314	0.277	0.221	0.188
$\mu_\lambda = 1.3$	0.244	0.285	0.264	0.207
$\mu_\lambda = 1.5$	0.206	0.276	0.287	0.238
$\mu_\lambda = 1.7$	0.152	0.256	0.312	0.280
$\mu_\lambda = 2$	0.093	0.271	0.313	0.323
$\mu_\lambda = 3$	0.033	0.146	0.375	0.446
$\mu_\lambda = 4$	0.006	0.079	0.331	0.584
$\mu_\lambda = 5$	0.000	0.046	0.282	0.672
$\mu_\lambda = 6$	0.000	0.032	0.248	0.720
$\mu_\lambda = 7$	0.000	0.017	0.198	0.785

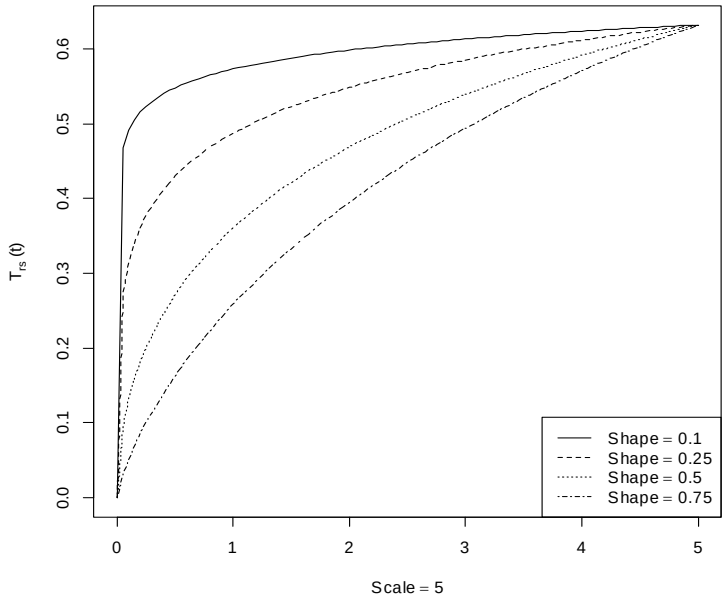

Figure 5.1: Transition functions with most weight at the beginning of an interval ($\kappa = 5$).

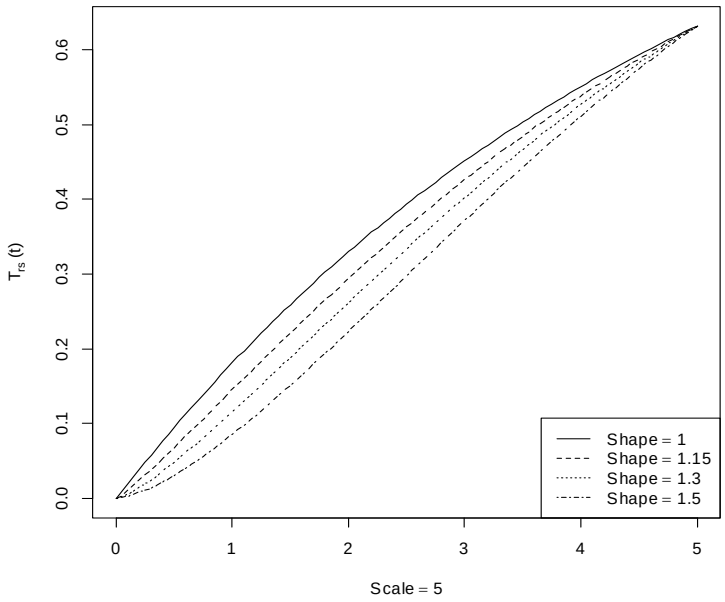

Figure 5.2: Transition functions with the weight spread evenly across an interval ($\kappa = 5$).

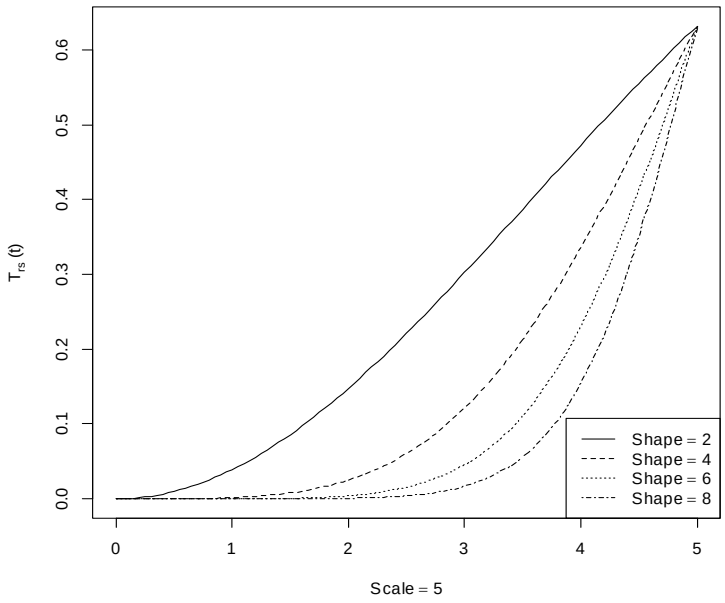

Figure 5.3: Transition functions with most weight at the end of an interval ($\kappa = 5$).

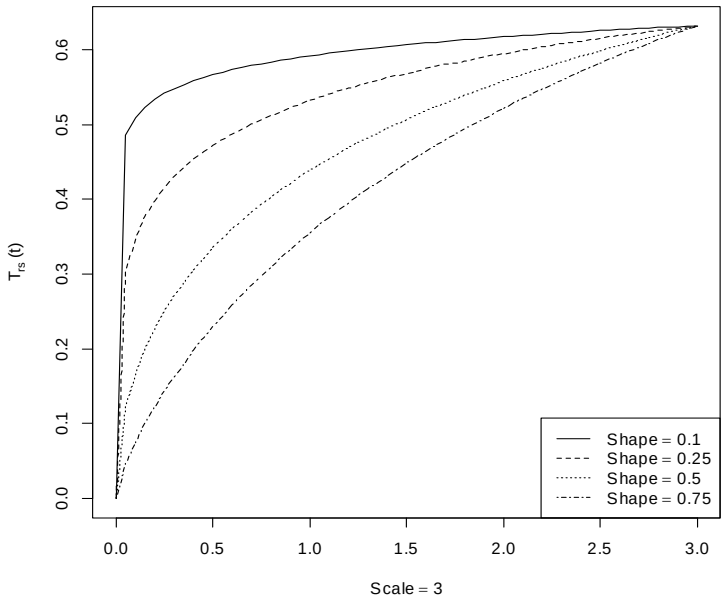

Figure 5.4: Transition functions with most weight at the beginning of an interval ($\kappa = 3$).

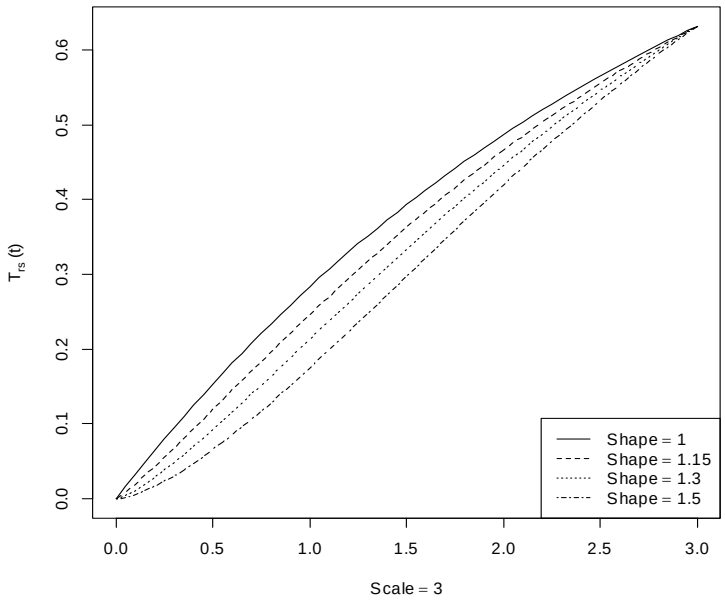

Figure 5.5: Transition functions with the weight spread evenly across an interval ($\kappa = 3$).

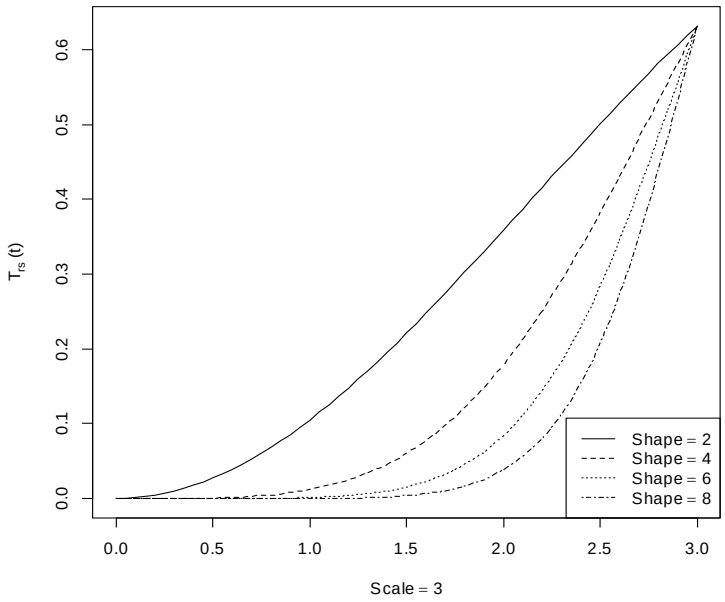

Figure 5.6: Transition functions with most weight at the end of an interval ($\kappa = 3$).

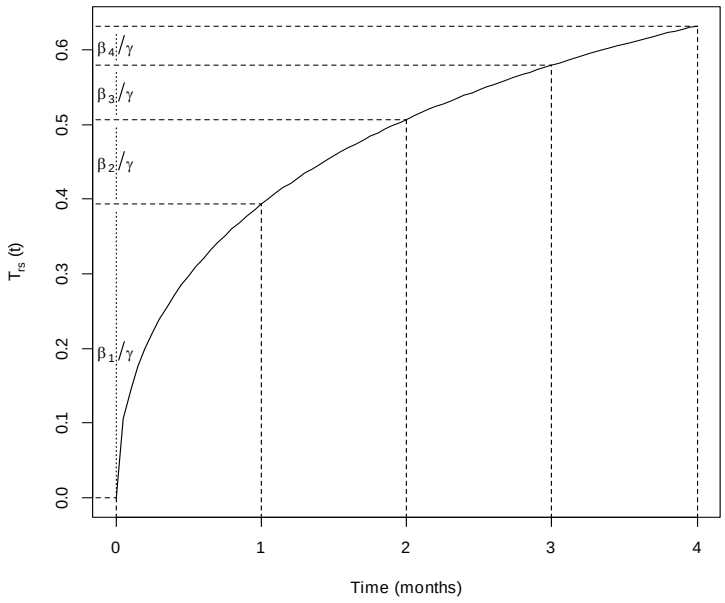

Figure 5.7: Transition function with $\kappa = 4$ and $\alpha = 0.5$.

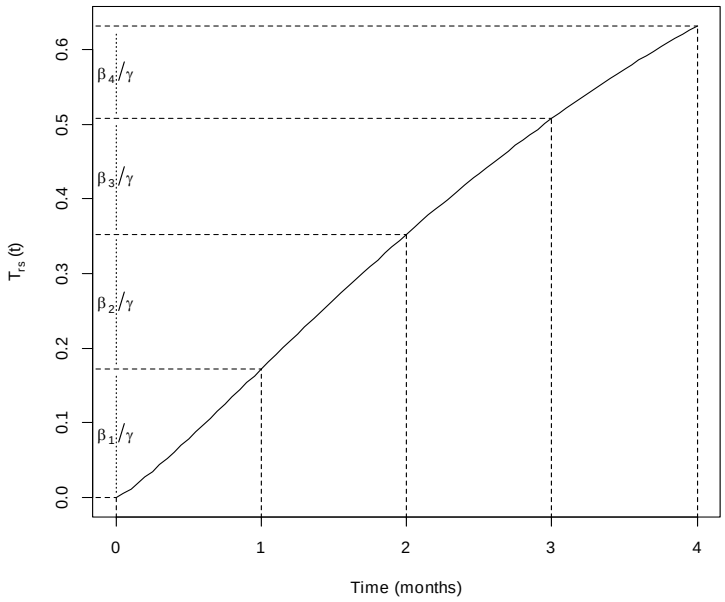

Figure 5.8: Transition function with $\kappa = 4$ and $\alpha = 1.2$.

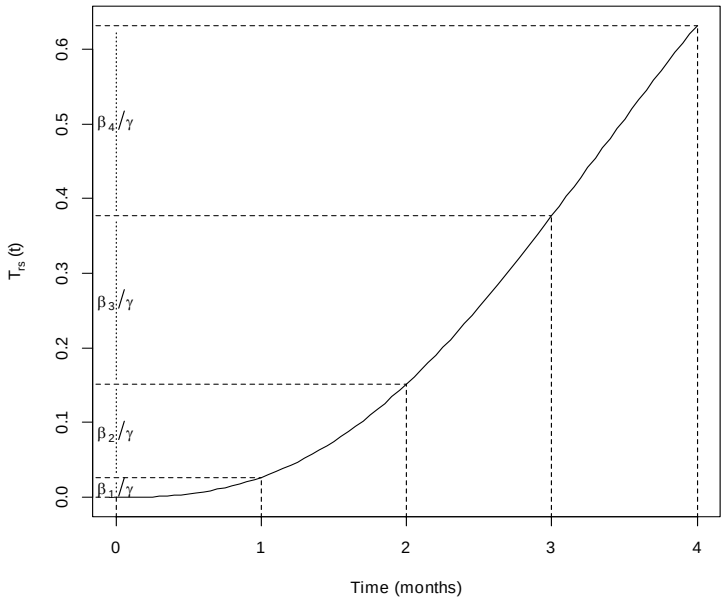

Figure 5.9: Transition function with $\kappa = 4$ and $\alpha = 2.6$.

5.3 Illustrating the imputation process: an example

In this section the two proposed methods are illustrated using the example presented in the introduction to this chapter. Assume a 3-state progressive model with monthly follow-up visits. A typical patient's data consisting of the known observations, the observation times (in months) and the states, is shown in Table 5.1.

5.3.1 Imputing all unknown observations

As the patients are assumed to be followed-up monthly, the time grid for this example is

$$t_0 = 0 < t_1 = 1 < ... < t_7 = 7.$$

Define Q^1, the first clinician's prior belief, as

$$Q^1 = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.15 & 0.7 & 0.15 \\ 0 & 0.3 & 0.7 \end{bmatrix},$$

and Q^2, the second clinician's prior belief, as

$$Q^2 = \begin{bmatrix} 0.5 & 0.45 & 0.05 \\ 0.3 & 0.5 & 0.2 \\ 0.05 & 0.45 & 0.5 \end{bmatrix}.$$

Q^1 shows that the first clinician's prior belief is that there is a 70% probability that a patient will stay in the current state at the next visit. If a patient is currently in state 1 there is a 30% probability that the patient would be in state 2 at the next visit. If a patient is currently in state 2 it is equally likely that the patient will be in either state 1 or state 3 at the next visit. If a patient is currently in state 3 there is a 30% probability that the patient will be in state 2 at the next visit.

Q^2 shows that the second clinician's prior belief is that there is a 50% probability that a patient will stay in the current state at the next visit. If a patient is currently in state 1 there is a 45% probability that the patient would be in state 2 and a 5% probability that the patient will be in state 3 at the next visit. If a patient is currently in state 2 there is a 30% probability that a patient will be in state 1 and a 20% probability that the patient will be in state 3 at the next visit. If a patient is currently in state 3 there is a 45% probability that the patient will be in state 2 and a 5% probability that the patient will be in state 1 at the next visit.

From Q^1 and Q^2, the average transition probability matrix A and the difference between the transition probabilities, matrix D, can be calculated as

$$A = \begin{bmatrix} 0.6 & 0.375 & 0.025 \\ 0.225 & 0.6 & 0.175 \\ 0.025 & 0.375 & 0.6 \end{bmatrix}, \tag{5.7}$$

$$D = \begin{bmatrix} 0.2 & 0.12 & 0.05 \\ 0.15 & 0.2 & 0.05 \\ 0.05 & 0.15 & 0.2 \end{bmatrix}. \tag{5.8}$$

Table 5.5 shows the time points that have missing observations and where the missing states will be imputed. At each time point a vector of transition probabilities,$[\theta_1, \theta_2, \theta_3]'$, is generated from a Dirichlet distribution with parameters (ca_1, ca_2, ca_3), with (a_1, a_2, a_3) selected from the corresponding row in (5.7) and c calculated using (5.4) with (5.8) and (5.7). If, for example, the missing observation is to be calculated for t_1, the following values are used:

$$\begin{aligned} (a_1, a_2, a_3) &= (0.6, 0.375, 0.025), \\ \widehat{E[R_1]} &= \sum_{j=1}^{3} d_{1j}^2 = 0.0569, \\ \sum_{j=1}^{3} \widehat{\eta}_{1j}{}^2 &= \sum_{j=1}^{3} a_{1j}^2 = 0.50125, \\ 0.0569 &= \frac{2}{c+1}(1 - 0.50125) \\ c &= 7.7654 \end{aligned}$$

and $[\theta_1, \theta_2, \theta_3]'$ is generated from a $Dirichlet(4.66, 2.91, 0.19)$ distribution. Using this distribution a possible probability vector at $t_1 = 1$ is given by $[0.508, 0.444, 0.048]'$, and this is now used as the parameters of the multinomial distribution to generate the unobserved state at time t_1, $S(1)$. For this example, $S(1)$ is generated as $S(1) = 1$ (Table 5.6).

This process is repeated at each time point with a missing observation until all missing values have been imputed for this specific patient (Table 5.6) and then for all patients under study.

Table 5.5: Missing observations that need to be imputed.

Observation	1	I_1	I_2	I_3	2	I_4	I_5	3
Time	0	**1**	**2**	**3**	4	**5**	**6**	7
State	1	?	?	?	2	?	?	3

Table 5.6: Imputing missing observations.

Observation	1	I_1	I_2	I_3	2	I_4	I_5	3
Time	0	**1**	**2**	**3**	4	**5**	**6**	7
State	1	**1**	**2**	**2**	2	**2**	**3**	3

The imputed data is now used to fit a multi-state model as described in Chapter 2 and to estimate the transition rates between the various states. Once the transition rates have been estimated, they are stored and the whole process is repeated 5000 times to generate a posterior distribution of the transition rates.

5.3.2 Estimating the transition point

As the patients are assumed to be followed-up monthly, the time grid for this example is

$$t_0 = 0 < t_1 = 1 < ... < t_7 = 7.$$

The time intervals during which transition occurred are $(0; 4]$ and $(4; 7]$ (Table 5.7). Each one of these intervals is divided into smaller one month intervals giving possible transition points between 0 and 4 as

$$(0; 1], (1; 2], (2; 3], (3; 4]$$

and between 4 and 7 as

$$(4; 5], (5; 6], (6; 7].$$

Over the 4 intervals between 0 and 4 the $T_{12}(t)$ transition function will be used and over the 3 intervals between 4 and 7 the $T_{23}(t)$ transition function will be used. Assume that from prior information it is known that patients currently in state 1 stay in state 1 for quite a while before transitioning to state 2 and that patients in state 2 makes a quick transition to state 3. This means that $T_{12}(t)$ will be defined by T_E-type transition functions and $T_{23}(t)$ by T_B-type transition functions. Figures 5.10 and 5.11 show one example of the $T_{12}(t)$ and $T_{23}(t)$ functions that are used over the two intervals ($T_{12}(t) = 1 - \exp(-(\frac{t}{4})^3)$ and $T_{23}(t) = 1 - \exp(-(\frac{t}{3})^{0.5})$). In (5.5) assume $\gamma = 5$, the prior weight, and a T_E-function with $\kappa = 4$ and $\alpha = 3$ for the transition function, then for each one of the intervals between 0 and 4, β_w, for $w = 1, 2, 3, 4$,

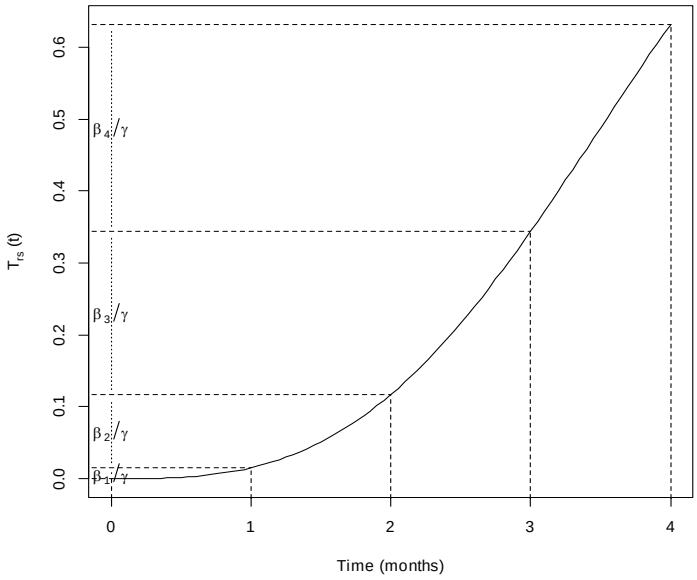

Figure 5.10: One of the 200 T_E-type T_{12} transition functions ($\kappa = 4, \alpha = 3$) used over the 4 time intervals between $[0 - 4]$.

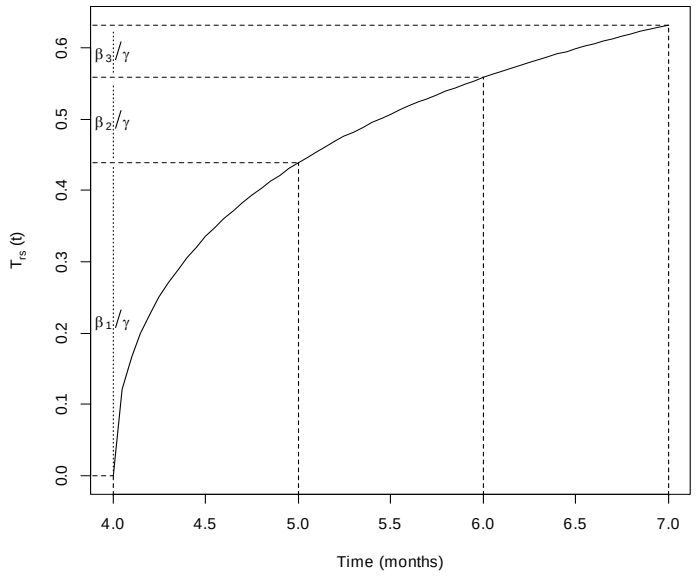

Figure 5.11: One of the 200 T_B-type T_{23} transition functions ($\kappa = 3, \alpha = 0.5$) used over the 3 time intervals between $[4 - 7]$.

and $[t_0 = 0, t_1 = 1, t_2 = 2, t_3 = 3, t_4 = 4]$ is calculated as

$$\beta_w = 5 \left| \left\{ 1 - \exp(-(\frac{t_{k-1}}{4})^3 \right\} - \left\{ 1 - \exp(-(\frac{t_k}{4})^3 \right\} \right|,$$

giving

$$\beta_1 = 1.97, \beta_2 = 0.57, \beta_3 = 0.36, \beta_4 = 0.26.$$

The probabilities that the transition took place in each one of the 4 intervals are now generated from a $Dirichlet(1.97, 0.57, 0.36, 0.26)$ distribution. A possible (4×1) transition probability vector from this distribution is

$$\boldsymbol{\theta}' = [0.056, 0.108, 0.202, 0.634],$$

indicating that there is a 5.6% probability that the transition occurred in the $(0; 1]$ interval, a 10.8% probability that it occurred in the $(1; 2]$ interval, a 20.2% probability that it occurred in the $(2; 3]$ interval and a 63.4% probability that it occurred in the $(3; 4]$ interval.

As $\boldsymbol{\theta}$ is highly dependent on the shape of the transition function, the process of generating the shape parameter using the transition function to calculate the $\beta-$values and generating a probability vector $\boldsymbol{\theta}$ for the possible transition interval, is repeated 100 times. On the next repetition the shape parameter may be 2.6 and $\boldsymbol{\theta}$ may be generated as

$$\boldsymbol{\theta}' = [0.00, 0.298, 0.141, 0.561].$$

A mean (4×1) transition probability vector $\bar{\boldsymbol{\theta}}$ is calculated after repeating the above process $g = 100$ times and, for this example, this is given by

$$\bar{\boldsymbol{\theta}}' = [0.021, 0.111, 0.356, 0.512]. \tag{5.9}$$

The mean transition probability vector is used as the parameter of a 4-state multinomial distribution and the interval in which the transition occurred is generated from this distribution. From Table 5.8 it can be seen that for this example, it is estimated that the transition occurred in the $(2; 3]$ interval, and thus it is known that the patient is in state 2 from time 3 onwards. Note that if the estimated interval was the $(3; 4]$ interval, the data would not have been changed and no new observation would have been imputed.

For the transition in the $(4; 7]$ interval the above mentioned process is repeated, with the only difference being that a T_B-type transition function is used (scale = 3, shape = 0.5) and there

are now only 3 possible transition intervals. Two possible (3×1) transition probability vectors for this interval are

$$\boldsymbol{\theta}' = [0.120, 0.835, 0.045]$$

and

$$\boldsymbol{\theta}' = [0.738, 0.239, 0.023]$$

leading to a mean (3×1) transition probability vector of

$$\overline{\boldsymbol{\theta}}' = [0.636, 0.218, 0.146]\,. \tag{5.10}$$

Using this mean vector the transition in the $(4; 7]$ interval is estimated to have taken place in the $(4; 5]$ interval, and thus it is known that the patient is in state 3 from time 5 onwards. The end result of this process can be seen in Table 5.8, were the original data has been augmented with the newly generated states at time 3 and time 5.

Table 5.7: Missing observations that need to be imputed.

Observation	1	I_1	I_2	I_3	2	I_4	I_5	3
Time	0	**1**	**2**	**3**	4	**5**	**6**	7
State	1	?	?	?	2	?	?	3

Table 5.8: Imputed transition points.

Observation	1	I_1	2	I_2	3
Time	0	**3**	4	**5**	7
State	1	**2**	2	**3**	3

The imputed data is now used to fit a multi-state model as described in Chapter 2 and to estimate the transition rates between the various states. Once the transition rates have been estimated, they are stored and the whole process is repeated 5000 time, to generate a posterior distribution of the transition rates.

Comparing (5.9) and (5.10) the effect of the different types of transition functions can clearly be seen. When using a T_B-function (5.10) the highest probability of a transition is placed near the beginning of the interval under study. For the T_E-function (5.9) this was reversed, with the highest probability of a transition placed near the end of the interval under study. Although not used in this example, when using a T_F transition function, typical (4×1) and

(3×1) mean transition probability vectors are given by

$$\bar{\boldsymbol{\theta}}' = [0.258, 0.262, 0.257, 0.223]\,,$$

or

$$\bar{\boldsymbol{\theta}}' = [0.341, 0.339, 0.320]\,,$$

showing that the probability of a transition is spread out evenly across the intervals.

5.4 Simulation study

Two different Bayesian methods of imputing multi-state data have been presented in this chapter:

- Firstly, all missing observations in a panel data set were imputed using prior probability vectors obtained from clinical experts. It is assumed that these vectors come from the same Dirichlet distribution and it is this distribution that is used to estimate the missing states at each unobserved time point. In the remainder of this simulation section these models will be referred to as **Fill** models.
- Secondly, the transition point between two known observations were imputed using a Dirichlet process prior with base function defined by known transition functions. The transition functions are Weibull functions with varying shape and scale parameters. By varying the shape parameter, it is possible to assign different probabilities to the intervals between the two known transition points. In the remainder of this simulation section these models will be referred to as **TP** models.

In this section, a simulation study is performed to investigate the properties of these two modelling techniques.

5.4.1 Simulation process and methodology

The starting point for the simulation study is to generate data from a known population. The sample generated from a known population will then be used to assess the effectiveness of the proposed techniques. Effectiveness here is defined as the Bayesian technique being able to accurately estimate the population values. To this end, the simulation process that will be

employed here will proceed as follows:

1) Generate a multi-state data set from a specified population using the process described in Section 2.3 (see also Section 5.4.1.1 for the various models that will be used).
2) Impute the missing observations using the two approaches described in this chapter and fit a multi-state model to the imputed data set (see Sections 5.1.1 and 5.2.1 for the methodology that is used for each technique). For each method vary the prior information to assess its effect on the modelling process.
   - When imputing all unknown observations, various prior mean vectors (5.1) are used in the modelling process to assess the effect of differing prior vectors on the parameter estimates.
   - When imputing the transition point between two observations, various shapes (T_B, T_F or T_E-functions) of the underlying transition function are used in the modelling process to assess the effect of the differing functions on the parameter estimates.
3) Steps 1 and 2 are repeated 5000 times to generate the posterior distribution of the parameters.
4) Assess the posterior distribution to see how accurately the estimates are given the known population values used to generate the specific data set. See Section 5.4.1.2 for the posterior statistics that will be used to assess each simulation run.

To calculate the coverage probability of the process, steps 1 to 3 are repeated 1000 times. The number of times that the credible interval and HPD region include the known population value is counted and expressed as a percentage to give the coverage probability (see Section 5.4.1.2). After running the simulation for models with no covariates, a select number of models will be run with covariates included in the modelling process. The purpose of these models is not to place priors on the covariates, but rather to investigate how the effect of the covariates is influenced when imputing the missing states.

5.4.1.1 Different models and scenarios

The investigation of the effect of the different model structures in the simulation process is done by using 6 different Q matrices. As the interest in this dissertation is on the disease progression of individuals, it will be assumed that individuals can only make a transition from one state to

an adjoining state and that no instantaneous transitions are possible to non-adjoining states. The models that will be investigated are the three 3-state (Q_1 (2.26), Q_2 (2.27), Q_3 (2.28)) and three 4-state (Q_4 (2.29), Q_5 (2.30) and Q_6 (2.31)) models introduced in Section 2.3.1.1.

As noted in Section 2.3.3, the sample size plays an important role when fitting a multi-state model. To this end, the effect of the sample size and the number of observations per patient on the proposed techniques are also investigated. Six different data scenarios (Sc1 to Sc6 from Table 2.2) are assumed and compared across all models.

Along with these six data scenarios, data with covariates included will also be investiagted. As noted in Section 4.2.1 the aim of this dissertation is not to create a model with priors placed on the covariates and as such, when introducing covariates into the modelling process, only models with one categorical covariate (covariate model A as defined in Section 2.3.1.1) or models with one catogorical and one continuous variable (covariate model B as defined in Section 2.3.1.1) are investigated.

The first four data scenarios will be used when simulating models where no covariates are included in the modelling process. As the complexity of the models increases with the inclusion of covariates, scenarios 5 and 6 are specifically included to be used when the effect of covariates is being investigated (see Section 5.4.3).

The prior assumptions regarding the transition made by individuals in the data play an important role in the imputing process. As such it is critical to understand and investigate the effect of different priors on the modelling process. The different prior assumptions that will be investigated for the two proposed Bayesian multi-state imputation techniques are:

– When imputing all missing observations (Section 5.1) the following 5 priors are assumed to be the prior belief obtained from two clinicians:

$$\text{Prior}_1^{Fill} \begin{bmatrix} 0.50 & 0.50 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0.50 & 0.50 \end{bmatrix} \text{ and } \begin{bmatrix} 0.55 & 0.45 & 0 \\ 0.30 & 0.40 & 0.30 \\ 0 & 0.45 & 0.55 \end{bmatrix} \text{ for a 3-state model, or}$$

$$\begin{bmatrix} 0.50 & 0.50 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0.50 & 0.50 \end{bmatrix} \text{ and } \begin{bmatrix} 0.55 & 0.45 & 0 & 0 \\ 0.30 & 0.40 & 0.30 & 0 \\ 0 & 0.30 & 0.40 & 0.30 \\ 0 & 0 & 0.45 & 0.55 \end{bmatrix} \text{ for a 4-state model.}$$

$$\text{Prior}_2^{Fill} \begin{bmatrix} 0.80 & 0.20 & 0 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0.20 & 0.80 \end{bmatrix} \text{ and } \begin{bmatrix} 0.90 & 0.10 & 0 \\ 0.05 & 0.90 & 0.05 \\ 0 & 0.10 & 0.90 \end{bmatrix} \text{ for a 3-state model, or}$$

$$\begin{bmatrix} 0.80 & 0.20 & 0 & 0 \\ 0.10 & 0.80 & 0.10 & 0 \\ 0 & 0.10 & 0.80 & 0.10 \\ 0 & 0 & 0.20 & 0.80 \end{bmatrix} \text{ and } \begin{bmatrix} 0.90 & 0.10 & 0 & 0 \\ 0.05 & 0.90 & 0.05 & 0 \\ 0 & 0.05 & 0.90 & 0.05 \\ 0 & 0 & 0.10 & 0.90 \end{bmatrix} \text{ for a 4-state model.}$$

Prior_3^{Fill} $\begin{bmatrix} 0.20 & 0.80 & 0 \\ 0.40 & 0.20 & 0.40 \\ 0 & 0.80 & 0.20 \end{bmatrix}$ and $\begin{bmatrix} 0.10 & 0.90 & 0 \\ 0.45 & 0.10 & 0.45 \\ 0 & 0.90 & 0.10 \end{bmatrix}$ for a 3-state model, or

$$\begin{bmatrix} 0.20 & 0.80 & 0 & 0 \\ 0.40 & 0.20 & 0.40 & 0 \\ 0 & 0.40 & 0.20 & 0.40 \\ 0 & 0 & 0.80 & 0.20 \end{bmatrix} \text{ and } \begin{bmatrix} 0.10 & 0.90 & 0 & 0 \\ 0.45 & 0.10 & 0.45 & 0 \\ 0 & 0.45 & 0.10 & 0.45 \\ 0 & 0 & 0.90 & 0.10 \end{bmatrix} \text{ for a 4-state model.}$$

Prior_4^{Fill} $\begin{bmatrix} 0.20 & 0.80 & 0 \\ 0.10 & 0.10 & 0.80 \\ 0 & 0.20 & 0.80 \end{bmatrix}$ and $\begin{bmatrix} 0.30 & 0.70 & 0 \\ 0.15 & 0.15 & 0.70 \\ 0 & 0.30 & 0.70 \end{bmatrix}$ for a 3-state model, or

$$\begin{bmatrix} 0.20 & 0.80 & 0 & 0 \\ 0.10 & 0.10 & 0.80 & 0 \\ 0 & 0.10 & 0.10 & 0.80 \\ 0 & 0 & 0.20 & 0.80 \end{bmatrix} \text{ and } \begin{bmatrix} 0.30 & 0.70 & 0 & 0 \\ 0.15 & 0.15 & 0.70 & 0 \\ 0 & 0.15 & 0.15 & 0.70 \\ 0 & 0 & 0.30 & 0.70 \end{bmatrix} \text{ for a 4-state model.}$$

Prior_5^{Fill} $\begin{bmatrix} 0.80 & 0.20 & 0 \\ 0.80 & 0.10 & 0.10 \\ 0 & 0.80 & 0.20 \end{bmatrix}$ and $\begin{bmatrix} 0.70 & 0.30 & 0 \\ 0.70 & 0.15 & 0.15 \\ 0 & 0.70 & 0.30 \end{bmatrix}$ for a 3-state model, or

$$\begin{bmatrix} 0.80 & 0.20 & 0 & 0 \\ 0.80 & 0.10 & 0.10 & 0 \\ 0 & 0.80 & 0.10 & 0.10 \\ 0 & 0 & 0.80 & 0.20 \end{bmatrix} \text{ and } \begin{bmatrix} 0.70 & 0.30 & 0 & 0 \\ 0.70 & 0.15 & 0.15 & 0 \\ 0 & 0.70 & 0.15 & 0.15 \\ 0 & 0 & 0.70 & 0.30 \end{bmatrix} \text{ for a 4-state model.}$$

– When imputing the transition point between two known observations (Section 5.2) the following 5 priors are assumed to be the priors of the underlying transition functions:

Prior_1^{TP} All transition functions are assumed to be T_B-functions. Under this assumption it is assumed that the transition to the following state occurred at the beginning of the current observation interval.

Prior_2^{TP} All transition functions are assumed to be T_F-functions. Under this assumption it is assumed that the transition to the following state occurred at any point during the current observation interval.

Prior_3^{TP} All transition functions are assumed to be T_E-functions. Under this assumption it is assumed that the transition to the following state occurred near the end of the current observation interval.

Prior_4^{TP} Transition functions to higher states (T_{12} or T_{23} or T_{34}) are assumed to be T_B-functions

and transitions to lower states (T_{21} or T_{32} or T_{43}) are assumed to be T_E-functions. Under these assumptions it is assumed that patients transition quickly from a lower to a higher state and once in a higher state takes longer to transition to a lower state.

Prior_5^{TP} Transition functions to higher states (T_{12} or T_{23} or T_{34}) are assumed to be T_E-functions and transitions to lower states (T_{21} or T_{32} or T_{43}) are assumed to be T_B-functions. Under these assumptions it is assumed that patients transition quickly from a higher to a lower state and once in a lower state takes longer to transition to a higher state.

Combining the different model structures, data scenarios and prior assumptions give $6\times4\times5 = 120$ simulations studies for each one of the two techniques introduced in this chapter.[4] *(See footnote 4)*

Table 5.9: Simulation results comparing the use of 100 vs 200 transition functions (Mean).

Mean	**100 Functions**				**200 Functions**			
Run	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$
1	0.6025	0.5580	0.2514	0.4184	0.4866	0.3469	0.4436	0.6549
2	0.4835	0.3913	0.6033	0.5887	0.6116	0.5478	0.4060	0.5190
3	0.5763	0.4640	0.5612	0.5870	0.4590	0.3153	0.5276	0.7390
4	0.5740	0.4290	0.5688	0.4631	0.4796	0.3336	0.4473	0.5188
5	0.4199	0.2131	0.5174	0.5115	0.5991	0.5117	0.3624	0.4103
6	0.5062	0.4637	0.4331	0.4144	0.4993	0.4315	0.3937	0.4319
7	0.4846	0.3735	0.4085	0.5155	0.5208	0.3397	0.3970	0.4311
8	0.4965	0.3244	0.3719	0.3238	0.3823	0.3913	0.3424	0.3928
9	0.4926	0.3725	0.4894	0.5732	0.4907	0.5071	0.4687	0.4326
10	0.6248	0.4499	0.4259	0.5252	0.6531	0.5340	0.3656	0.3576
Mean	**0.5261**	**0.4039**	**0.4631**	**0.4921**	**0.5182**	**0.4259**	**0.4154**	**0.4888**
SD	**0.0646**	**0.0934**	**0.1067**	**0.0866**	**0.0810**	**0.0920**	**0.0565**	**0.1222**

[4] A single realisation of imputing the data and fitting a 3-state model to data consisting of 25 patients observed over a 24 month period and assuming 50% missing observations, takes 165 minutes when using the mean of $g = 200$ transition functions between each pair of known observations (simulations run on an Intel Core i7-2600 3.40GHz computer). As this needs to be repeated 1000 times to assess the coverage probability of the process and hence this needs to be repeated for different model structures and data sets, it can become computationally time consuming. Before continuing with the simulation study, the effect of the number of transition functions used in the imputing process is first investigated. To this end, the results when using 100 repetitions (using 100 repetitions, the imputing and subsequent modelling of the data takes 95 minutes) is compared to the results when using 200 repetitions (tables 5.9 to 5.11). From these tables it is clear that there is not a big difference in the mean, standard deviation and MSE of the estimated transition rates when using 100 vs. 200 repetitions and that the values are fairly stable across the 10 different runs performed here. For the remainder of this simulation study $g = 100$ repetitions will be used to calculate the mean vector across the possible transition space (see section 5.2.1).

Table 5.10: Simulation results comparing the use of 100 vs 200 transition functions (SD).

SD	**100 Functions**				**200 Functions**			
Run	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$
1	0.0385	0.0326	0.0069	0.0159	0.0341	0.0231	0.0309	0.0504
2	0.0363	0.0247	0.0426	0.0443	0.0194	0.0166	0.0406	0.0693
3	0.0317	0.0237	0.0354	0.0443	0.0192	0.0154	0.0308	0.0380
4	0.0305	0.0205	0.0209	0.0482	0.0442	0.0345	0.0178	0.0242
5	0.0280	0.0135	0.0343	0.0186	0.0283	0.0238	0.0177	0.0240
6	0.0263	0.0232	0.0198	0.0395	0.0369	0.0244	0.0232	0.0291
7	0.0286	0.0224	0.0307	0.0235	0.0180	0.0188	0.0178	0.0275
8	0.0295	0.0194	0.0176	0.0480	0.0255	0.0268	0.0235	0.0275
9	0.0316	0.0220	0.0278	0.0183	0.0771	0.0421	0.0190	0.0227
10	0.0319	0.0218	0.0349	0.0355	0.0215	0.0248	0.0187	0.0267
Mean	**0.0313**	**0.0224**	**0.0271**	**0.0401**	**0.0324**	**0.0250**	**0.0240**	**0.0339**
SD	**0.0037**	**0.0048**	**0.0107**	**0.0129**	**0.0179**	**0.0081**	**0.0077**	**0.0150**

Table 5.11: Simulation results comparing the use of 100 vs 200 transition functions (MSE).

MSE	**100 Functions**				**200 Functions**			
Run	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$
1	0.0120	0.0044	0.0619	0.0069	0.0013	0.0240	0.0041	0.0265
2	0.0016	0.0124	0.0432	0.0376	0.0142	0.0034	0.0092	0.0016
3	0.0068	0.0019	0.0050	0.0099	0.0203	0.0344	0.0024	0.0619
4	0.0064	0.0055	0.0052	0.0017	0.0149	0.0279	0.0037	0.0018
5	0.0072	0.0825	0.0015	0.0017	0.0118	0.0013	0.0192	0.0086
6	0.0007	0.0019	0.0049	0.0079	0.0008	0.0053	0.0116	0.0052
7	0.0011	0.0165	0.0093	0.0025	0.0019	0.0263	0.0112	0.0056
8	0.0009	0.0312	0.0167	0.0314	0.0142	0.0122	0.0252	0.0122
9	0.0011	0.0167	0.0009	0.0066	0.0007	0.0008	0.0015	0.0053
10	0.0166	0.0030	0.0171	0.0176	0.0130	0.0029	0.0184	0.0208
Mean	**0.0054**	**0.0176**	**0.0166**	**0.0124**	**0.0093**	**0.0139**	**0.0107**	**0.0149**
SD	**0.0055**	**0.0246**	**0.0202**	**0.0126**	**0.0074**	**0.0130**	**0.0081**	**0.0184**

5.4.1.2 Assessing the simulation process

The posterior MSE will be the main statistic used to assess the performance of the proposed techniques. A smaller MSE will indicate that the estimates of the transition rates can be viewed as representative of the population under study, while a large MSE indicates a departure from the population values. As the MSE can be influenced by extreme values, the MedSE will be presented in cases where extreme MSE values are observed. The use of the MedSE will be clearly highlighted in the results.

The bias, another indicator as to the performance of the simulation process, and defined as

$$Bias(\hat{\theta}) = \sqrt{MSE - var(\hat{\theta})},$$

will be presented in a select number of simulation runs.

The mean, median and standard deviation of the posterior distribution will also be investigated for each simulation run.

The credible interval (Cred) and highest posterior density (HPD) region is calculated for each data run and will be used to calculate the equivalent frequentist coverage probabilities of the process. For this simulation study, the coverage probabilities will be based on 1000 repetitions of the process described in Section 5.4.1. The mean length of the posterior intervals over the 1000 repetitions are also calculated to give an indication of the width of the posterior intervals.[5] (*see footnote* 5)

5.4.2 Simulation results

In this section the simulation results of the two different approaches used

i) imputing all unknown observations (See Appendix A.4 for the R functions and programs used to generate these results.), and

ii) estimating the transition point between two known observations (See Appendix A.5 for the R functions and programs used to generate these results.),

are shown, discussed and interpreted.

[5] As the computational time required to perform 1000 repetitions of the modelling process takes close on 15 days (this time is based on using all four processors found in most computers today), the posterior coverage probabilities and interval lengths will only be calculated for a select number of simulation studies.

5.4.2.1 Imputing all unknown observations

In Tables 5.12 to 5.32, models fitted when imputing all unknown observations in the data are investigated. Tables 5.12 to 5.29 summarise the posterior distributions across the different scenarios, in Table 5.30 the bias associated with a select number of scenarios is presented and in Tables 5.31 to 5.32, the posterior coverage probabilities and mean lengths of the credible and HPD intervals are shown.

The results can be summarised as follows:

- In this section, unlike in Sections 2.3.3 and 4.3, where the different scenarios had a big effect on the MSE's of the models being fitted, no real differences between the MSE's are observed across the different scenarios. This is to be expected, as the technique employed here imputes all unknown missing values, negating the fact that the different scenarios have different percentages of missingness.
- For models Q_1 and Q_4 (see Tables 5.12 to 5.15, 5.24 and 5.25), prior 1 has the smallest MSE values. Under these models the probability of transitioning to a higher or a lower state is exactly the same and, as this is what prior 1 suggests, it is to be expected that prior 1 should lead to the smallest MSE values for these models.
- When modelling Q_2 and Q_5 (see Tables 5.16 to 5.19, 5.26 and 5.27) the probability is larger that an individual will transition to a lower state rather than a higher state. For these models prior 5, the prior with the largest probability placed on moving backwards to a lower state, has the smallest MSE's.
- When modelling Q_3 and Q_6 (see Tables 5.20 to 5.23, 5.28 and 5.29) the probability is larger that an individual will transition to a higher rather than a lower state. For these models prior 4, the prior with the largest probability placed on moving forwards to the next higher state, has the smallest MSE's.
- Prior 3 has in general much larger MSE values across the different models and scenarios. This prior can be interpreted as an individual is unlikely to stay in his or her current state, but keeps moving up or down through the states. If in the lowest state (state 1) the individual would move up to the next state and if in the highest state (state 3 or 4) the individual would move to a lower state. This continuous up and down movement may be the reason why the parameter estimates when using prior 3 across all 6 models are the most

unstable, i.e. they have the highest SD and MSE values.

- The bias associated with model Q_1, data scenarios 1 and 3, and model Q_6, data scenarios 3 and 4, show no extreme bias values (see Table 5.30).
- The frequentist coverage probabilities are of similar order for both models Q_1 and Q_5 (see Tables 5.31 and 5.32), with some of the intervals having lower values than the expected 95%.
- No extreme MSE's are observed for any of the models fitted here, the largest value found is 2.2876 for model Q_2, scenario 4 with prior 3 (see Table 5.19).

From these results it is clear that this method of modelling multi-state data gives good all-round parameter estimates of the population transition rates. When the correct prior is used, i.e. the prior that best matches the underlying population model, the MSE's of the models become smaller, indicating that correctly specifying the prior does lead to better parameter estimates. When an incorrect prior is used, as is the case with prior 3, the models become more unstable and the SD and MSE's increase. This highlights the importance of using an appropriate prior and what effect the priors play on the Bayesian multi-state imputation technique presented here.

Table 5.12: Simulation results of imputing all unknown states - Q_1Sc_1.

Q1	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc1	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.7985	0.5415	0.5019	0.7201	**Pr2**	0.3387	0.2190	0.2418	0.3564
	Med	0.7933	0.5384	0.4992	0.7153		0.3366	0.2178	0.2405	0.3538
	SD	0.0831	0.0570	0.0517	0.0791		0.0345	0.0217	0.0225	0.0404
	MSE	0.0975	0.0057	0.0061	0.0589		0.0283	0.0799	0.0683	0.0227
Pr3	**Mean**	1.3820	0.8275	0.7737	1.3036	**Pr4**	0.8620	0.5253	0.7796	0.6067
	Med	1.3799	0.8257	0.7710	1.3014		0.8598	0.5233	0.7769	0.6036
	SD	0.1119	0.0805	0.0748	0.1055		0.0746	0.0441	0.0665	0.0618
	MSE	0.8039	0.1167	0.0835	0.6580		0.1432	0.0059	0.0855	0.0158
Pr5	**Mean**	0.5550	0.7775	0.4886	0.6881					
	Med	0.5516	0.7754	0.4865	0.6867					
	SD	0.0548	0.0683	0.0385	0.0590					
	MSE	0.0086	0.0827	0.0043	0.0417					

Table 5.13: Simulation results of imputing all unknown states - Q_1Sc_2.

Q1	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc2	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.8469	0.5428	0.5729	0.7774	**Pr2**	0.2967	0.1983	0.1743	0.2956
	Med	0.8408	0.5396	0.5698	0.7721		0.2943	0.1967	0.1731	0.2917
	SD	0.0993	0.0650	0.0642	0.0939		0.0360	0.0222	0.0197	0.0466
	MSE	0.1310	0.0080	0.0106	0.0862		0.0427	0.0923	0.1069	0.0442
Pr3	**Mean**	1.6973	1.0907	0.8554	1.6293	**Pr4**	0.9574	0.5172	0.9856	0.5970
	Med	1.6966	1.0890	0.8540	1.6269		0.9536	0.5147	0.9817	0.5932
	SD	0.1397	0.1003	0.0847	0.1431		0.0982	0.0517	0.0944	0.0687
	MSE	1.4631	0.3617	0.1390	1.3052		0.2198	0.0042	0.2463	0.0143
Pr5	**Mean**	0.5950	1.0151	0.4922	0.8574					
	Med	0.5909	1.0108	0.4895	0.8542					
	SD	0.0678	0.1016	0.0491	0.0924					
	MSE	0.0140	0.2830	0.0049	0.1377					

Table 5.14: Simulation results of imputing all unknown states - Q_1Sc_3.

Q1	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc3	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.6727	0.5460	0.5128	0.6682	**Pr2**	0.3632	0.2461	0.2582	0.3464
	Med	0.6714	0.5449	0.5115	0.6664		0.3618	0.2454	0.2575	0.3453
	SD	0.0469	0.0388	0.0365	0.0518		0.0261	0.0160	0.0159	0.0269
	MSE	0.0328	0.0047	0.0028	0.0354		0.0197	0.0650	0.0590	0.0245
Pr3	**Mean**	1.3919	0.8647	0.7918	1.3713	**Pr4**	0.8678	0.5577	0.7487	0.5459
	Med	1.3898	0.8632	0.7908	1.3705		0.8666	0.5565	0.7478	0.5445
	SD	0.0792	0.0580	0.0534	0.0816		0.0509	0.0321	0.0450	0.0400
	MSE	0.8048	0.1376	0.0891	0.7888		0.1393	0.0050	0.0676	0.0044
Pr5	**Mean**	0.5869	0.8366	0.4953	0.7637					
	Med	0.5852	0.8351	0.4943	0.7622					
	SD	0.0410	0.0522	0.0294	0.0511					
	MSE	0.0105	0.1179	0.0011	0.0744					

Table 5.15: Simulation results of imputing all unknown states - Q_1Sc_4.

Q1	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc4	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.8057	0.5656	0.5500	0.7663	**Pr2**	0.3000	0.1783	0.1878	0.2994
	Med	0.8030	0.5642	0.5484	0.7633		0.2987	0.1775	0.1871	0.2978
	SD	0.0647	0.0460	0.0436	0.0655		0.0265	0.0148	0.0149	0.0301
	MSE	0.0982	0.0077	0.0047	0.0758		0.0410	0.1039	0.0978	0.0418
Pr3	**Mean**	1.7851	1.0019	0.9266	1.7590	**Pr4**	0.9516	0.5280	0.9890	0.6036
	Med	1.7840	1.0012	0.9263	1.7586		0.9501	0.5268	0.9868	0.6016
	SD	0.1073	0.0689	0.0647	0.1102		0.0675	0.0366	0.0665	0.0481
	MSE	1.6649	0.2581	0.1870	1.5985		0.2090	0.0027	0.2442	0.0132
Pr5	**Mean**	0.6242	1.0891	0.4771	0.8915					
	Med	0.6224	1.0871	0.4759	0.8893					
	SD	0.0497	0.0772	0.0356	0.0731					
	MSE	0.0183	0.3557	0.0028	0.1634					

Table 5.16: Simulation results of imputing all unknown states - Q_2Sc_1.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc1	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.6091	0.9218	0.2884	0.8658	**Pr2**	0.2195	0.4293	0.0871	0.4886
	Med	0.6051	0.9158	0.2857	0.8574		0.2183	0.4275	0.0857	0.4734
	SD	0.0651	0.0942	0.0439	0.1327		0.0221	0.0367	0.0152	0.1124
	MSE	0.1343	0.0406	0.0040	0.0331		0.0017	0.1070	0.0270	0.0824
Pr3	**Mean**	1.1603	1.3262	0.4869	1.5700	**Pr4**	0.8457	1.0753	0.4006	0.6678
	Med	1.1558	1.3210	0.4859	1.5674		0.8440	1.0719	0.3988	0.6620
	SD	0.0975	0.1180	0.0583	0.1532		0.0687	0.0769	0.0457	0.0872
	MSE	0.8514	0.3596	0.0633	0.7058		0.3675	0.1222	0.0259	0.0171
Pr5	**Mean**	0.5001	0.6214	0.2052	0.8738					
	Med	0.4957	0.6113	0.2029	0.8695					
	SD	0.0635	0.0521	0.0368	0.1440					
	MSE	0.0683	0.0835	0.0047	0.0365					

Table 5.17: Simulation results of imputing all unknown states - Q_2Sc_2.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc2	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.7323	0.9212	0.3204	0.8588	**Pr2**	0.2179	0.3011	0.0805	0.3919
	Med	0.7265	0.9132	0.3182	0.8493		0.2166	0.2993	0.0792	0.3749
	SD	0.0837	0.0991	0.0482	0.1285		0.0258	0.0310	0.0162	0.1168
	MSE	0.2404	0.0414	0.0080	0.0286		0.0018	0.2035	0.0291	0.1433
Pr3	**Mean**	1.6992	1.4761	0.6346	1.8218	**Pr4**	1.0065	0.9350	0.6057	0.6206
	Med	1.6962	1.4719	0.6342	1.8165		1.0032	0.9317	0.6040	0.6164
	SD	0.1411	0.1284	0.0715	0.1800		0.0937	0.0733	0.0614	0.0739
	MSE	2.1246	0.5465	0.1538	1.1823		0.5835	0.0444	0.1310	0.0228
Pr5	**Mean**	0.6034	0.8137	0.2158	0.7340					
	Med	0.5972	0.7964	0.2133	0.7259					
	SD	0.0874	0.0329	0.0426	0.0735					
	MSE	0.1329	0.2045	0.0037	0.0269					

Table 5.18: Simulation results of imputing all unknown states - Q_2Sc_3.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc3	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.5744	0.8945	0.2703	0.8022	**Pr2**	0.2280	0.4142	0.0963	0.4252
	Med	0.5727	0.8922	0.2694	0.7986		0.2274	0.4131	0.0956	0.4180
	SD	0.0422	0.0622	0.0294	0.0847		0.0159	0.0261	0.0112	0.0696
	MSE	0.1074	0.0260	0.0015	0.0102		0.0010	0.1137	0.0239	0.1135
Pr3	**Mean**	1.1372	1.3815	0.4735	1.5914	**Pr4**	0.7633	0.9551	0.4451	0.6340
	Med	1.1353	1.3778	0.4728	1.5901		0.7620	0.9535	0.4442	0.6316
	SD	0.0707	0.0901	0.0414	0.1104		0.0432	0.0480	0.0339	0.0560
	MSE	0.7965	0.4072	0.0536	0.7350		0.2682	0.0450	0.0403	0.0169
Pr5	**Mean**	0.4809	0.8612	0.2028	0.9153					
	Med	0.4788	0.8557	0.2015	0.9128					
	SD	0.0431	0.0102	0.0256	0.1022					
	MSE	0.0552	0.0252	0.0032	0.0387					

Table 5.19: Simulation results of imputing all unknown states - Q_2Sc_4.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc4	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.7266	0.8773	0.3377	0.8503	**Pr2**	0.2116	0.3313	0.0814	0.3792
	Med	0.7235	0.8746	0.3365	0.8456		0.2107	0.3303	0.0808	0.3720
	SD	0.0583	0.0667	0.0349	0.0870		0.0180	0.0241	0.0114	0.0675
	MSE	0.2308	0.0214	0.0091	0.0177		0.0018	0.1760	0.0286	0.1448
Pr3	**Mean**	1.7532	1.4557	0.6565	1.8980	**Pr4**	0.9892	0.9493	0.5864	0.6629
	Med	1.7527	1.4540	0.6560	1.8945		0.9879	0.9475	0.5858	0.6603
	SD	0.1048	0.0908	0.0525	0.1355		0.0634	0.0519	0.0434	0.0559
	MSE	2.2876	0.5119	0.1682	1.3524		0.5526	0.0450	0.1156	0.0111
Pr5	**Mean**	0.5876	0.7503	0.2173	0.9121					
	Med	0.5845	0.7415	0.2162	0.9095					
	SD	0.0581	0.0527	0.0299	0.1066					
	MSE	0.1180	0.0264	0.0028	0.0408					

Table 5.20: Simulation results of imputing all unknown states - Q_3Sc_1.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc1	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.7933	0.2547	0.8732	0.5767	**Pr2**	0.5790	0.1112	0.4147	0.2190
	Med	0.7879	0.2523	0.8684	0.5734		0.5704	0.1097	0.4128	0.2177
	SD	0.0991	0.0400	0.0814	0.0611		0.0847	0.0168	0.0344	0.0237
	MSE	0.0121	0.0018	0.0225	0.1106		0.0500	0.0200	0.1161	0.0019
Pr3	**Mean**	1.5416	0.5158	1.3157	1.2974	**Pr4**	0.9960	0.2463	0.7327	0.4776
	Med	1.5404	0.5139	1.3110	1.2941		0.9923	0.2445	0.7262	0.4743
	SD	0.1286	0.0619	0.1149	0.1088		0.1155	0.0367	0.0269	0.0572
	MSE	0.6494	0.0745	0.3357	1.1246		0.0937	0.0030	0.0769	0.0625
Pr5	**Mean**	0.6932	0.4735	0.9580	0.7721					
	Med	0.6886	0.4719	0.9553	0.7702					
	SD	0.0777	0.0521	0.0669	0.0645					
	MSE	0.0113	0.0531	0.0558	0.2932					

Table 5.21: Simulation results of imputing all unknown states - Q_3Sc_2.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc2	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.8777	0.3437	0.8817	0.7268	**Pr2**	0.4343	0.0887	0.3149	0.2134
	Med	0.8712	0.3410	0.8766	0.7215		0.4257	0.0873	0.3131	0.2119
	SD	0.1128	0.0498	0.0897	0.0834		0.0717	0.0169	0.0303	0.0274
	MSE	0.0312	0.0113	0.0262	0.2362		0.1069	0.0264	0.1915	0.0024
Pr3	**Mean**	1.8897	0.6894	1.3707	1.7034	**Pr4**	0.7488	0.2407	0.6144	0.5418
	Med	1.8849	0.6880	1.3686	1.7018		0.7432	0.2380	0.6043	0.5377
	SD	0.1751	0.0772	0.1154	0.1410		0.1300	0.0431	0.0568	0.0701
	MSE	1.3317	0.2005	0.4047	2.1395		0.1129	0.0020	0.0137	0.0903
Pr5	**Mean**	0.6485	0.6623	0.8632	0.9443					
	Med	0.6443	0.6600	0.8605	0.9408					
	SD	0.0736	0.0701	0.0686	0.0913					
	MSE	0.0180	0.1757	0.0183	0.4937					

Table 5.22: Simulation results of imputing all unknown states - Q_3Sc_3.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc3	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.8164	0.2869	0.8826	0.5937	**Pr2**	0.5415	0.1034	0.4086	0.2153
	Med	0.8136	0.2856	0.8798	0.5920		0.5377	0.1027	0.4076	0.2147
	SD	0.0734	0.0303	0.0590	0.0448		0.0571	0.0119	0.0238	0.0162
	MSE	0.0113	0.0023	0.0231	0.1214		0.0496	0.0217	0.1188	0.0016
Pr3	**Mean**	1.4746	0.4949	1.2301	1.0783	**Pr4**	0.9500	0.2300	0.6398	0.4684
	Med	1.4740	0.4941	1.2264	1.0764		0.9491	0.2289	0.6372	0.4670
	SD	0.0892	0.0426	0.0764	0.0639		0.0712	0.0252	0.0852	0.0393
	MSE	0.5376	0.0657	0.2394	0.6958		0.0465	0.0014	0.0241	0.0505
Pr5	**Mean**	0.6228	0.4667	0.9102	0.6580					
	Med	0.6209	0.4660	0.9086	0.6569					
	SD	0.0492	0.0352	0.0452	0.0387					
	MSE	0.0186	0.0525	0.0322	0.1691					

Table 5.23: Simulation results of imputing all unknown states - Q_3Sc_4.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc4	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.8651	0.3618	0.8261	0.7123	**Pr2**	0.3974	0.0928	0.3149	0.2174
	Med	0.8616	0.3608	0.8236	0.7097		0.3938	0.0922	0.3139	0.2165
	SD	0.0799	0.0366	0.0605	0.0584		0.0478	0.0115	0.0208	0.0199
	MSE	0.0201	0.0141	0.0108	0.2174		0.1279	0.0249	0.1900	0.0016
Pr3	**Mean**	1.7848	0.6362	1.4112	1.5981	**Pr4**	1.0580	0.2276	0.7001	0.5343
	Med	1.7832	0.6354	1.4095	1.5962		1.0541	0.2264	0.6958	0.5318
	SD	0.1113	0.0514	0.0851	0.0935		0.0999	0.0303	0.0115	0.0499
	MSE	1.0877	0.1522	0.4469	1.8356		0.1111	0.0019	0.0777	0.0835
Pr5	**Mean**	0.6352	0.6020	0.8739	0.9060					
	Med	0.6330	0.6007	0.8724	0.9047					
	SD	0.0518	0.0466	0.0483	0.0598					
	MSE	0.0160	0.1273	0.0188	0.4351					

Table 5.24: Simulation results of imputing all unknown states - Q_4Sc_3.

Q4	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Sc3	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.7392	0.5657	0.4525	0.4554	0.5133	0.6718
	Med	0.7366	0.5635	0.4515	0.4540	0.5113	0.6678
	SD	0.0629	0.0496	0.0326	0.0375	0.0481	0.0682
	MSE	0.0618	0.0083	0.0038	0.0038	0.0035	0.0365
Pr2	**Mean**	0.3576	0.2560	0.2331	0.2352	0.2613	0.3752
	Med	0.3560	0.2550	0.2322	0.2339	0.2595	0.3715
	SD	0.0283	0.0199	0.0162	0.0210	0.0240	0.0447
	MSE	0.0215	0.0602	0.0719	0.0722	0.0577	0.0177
Pr3	**Mean**	1.3400	0.9125	0.4992	0.5280	0.8042	1.3324
	Med	1.3384	0.9105	0.4980	0.5271	0.8027	1.3309
	SD	0.0962	0.0755	0.0355	0.0422	0.0736	0.1110
	MSE	0.7184	0.1787	0.0020	0.0026	0.0992	0.7121
Pr4	**Mean**	0.9423	0.6465	0.5511	0.4775	0.7053	0.5355
	Med	0.9400	0.6446	0.5504	0.4766	0.7036	0.5329
	SD	0.0690	0.0462	0.0322	0.0321	0.0531	0.0502
	MSE	0.2060	0.0380	0.0072	0.0023	0.0460	0.0054
Pr5	**Mean**	0.6204	0.9097	0.4483	0.5557	0.5057	0.8259
	Med	0.6182	0.9075	0.4476	0.5549	0.5035	0.8235
	SD	0.0499	0.0657	0.0278	0.0388	0.0436	0.0803
	MSE	0.0171	0.1766	0.0038	0.0049	0.0062	0.1340

Table 5.25: Simulation results of imputing all unknown states - Q_4Sc_4.

Q4	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Sc4	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.7932	0.5791	0.4466	0.4604	0.5467	0.7528
	Med	0.7896	0.5765	0.4454	0.4589	0.5448	0.7477
	SD	0.0727	0.0553	0.0353	0.0418	0.0559	0.0830
	MSE	0.0922	0.0095	0.0042	0.0043	0.0061	0.0720
Pr2	**Mean**	0.3184	0.2159	0.2132	0.1888	0.2145	0.3496
	Med	0.3166	0.2149	0.2122	0.1876	0.2125	0.3442
	SD	0.0306	0.0210	0.0175	0.0203	0.0243	0.0543
	MSE	0.0342	0.0817	0.0828	0.0976	0.0822	0.0275
Pr3	**Mean**	1.6186	0.9917	0.5732	0.5720	0.8958	1.5058
	Med	1.6168	0.9902	0.5727	0.5710	0.8949	1.5048
	SD	0.1223	0.0861	0.0428	0.0482	0.0832	0.1275
	MSE	1.2694	0.2509	0.0076	0.0085	0.1667	1.0281
Pr4	**Mean**	1.0405	0.6519	0.6201	0.5259	0.7971	0.5633
	Med	1.0365	0.6500	0.6190	0.5249	0.7952	0.5605
	SD	0.0934	0.0562	0.0414	0.0387	0.0655	0.0560
	MSE	0.3016	0.0296	0.0173	0.0055	0.0975	0.0076
Pr5	**Mean**	0.6096	1.0247	0.4543	0.5936	0.5946	0.9031
	Med	0.6077	1.0213	0.4530	0.5924	0.5923	0.8971
	SD	0.0543	0.0831	0.0335	0.0496	0.0619	0.1096
	MSE	0.0154	0.2931	0.0054	0.0121	0.0131	0.1766

Table 5.26: Simulation results of imputing all unknown states - Q_5Sc_3.

Q5	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Sc3	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.6177	0.9660	0.2253	0.7088	0.2453	0.8149
	Med	0.6154	0.9620	0.2244	0.7052	0.2414	0.7911
	SD	0.0500	0.0742	0.0259	0.0820	0.0603	0.2152
	MSE	0.1389	0.0538	0.0015	0.0088	0.0040	0.0665
Pr2	**Mean**	0.2203	0.4104	0.0994	0.4056	0.1224	0.5268
	Med	0.2197	0.4093	0.0987	0.4011	0.1169	0.4976
	SD	0.0164	0.0285	0.0110	0.0621	0.0323	0.1911
	MSE	0.0013	0.1162	0.0230	0.1263	0.0181	0.1119
Pr3	**Mean**	1.2881	1.5648	0.2875	0.8317	0.3967	1.3971
	Med	1.2832	1.5593	0.2877	0.8308	0.3994	1.4253
	SD	0.1209	0.1484	0.0344	0.0953	0.0966	0.2946
	MSE	1.0985	0.6887	0.0026	0.0175	0.0316	0.5424
Pr4	**Mean**	0.8057	1.1149	0.3256	0.7681	0.4257	0.5533
	Med	0.8038	1.1124	0.3256	0.7676	0.4239	0.5475
	SD	0.0593	0.0774	0.0294	0.0693	0.0649	0.0986
	MSE	0.3139	0.1399	0.0069	0.0063	0.0354	0.0497
Pr5	**Mean**	0.4591	0.7394	0.1986	0.8840	0.2268	0.7084
	Med	0.4570	0.7348	0.1976	0.8834	0.2200	0.7078
	SD	0.0403	0.0964	0.0236	0.0902	0.0587	0.1927
	MSE	0.0455	0.0699	0.0034	0.0408	0.0083	0.0357

Table 5.27: Simulation results of imputing all unknown states - Q_5Sc_4.

Q5	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Sc4	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.7500	0.9340	0.2711	0.6599	0.3077	0.7929
	Med	0.7463	0.9305	0.2701	0.6567	0.3046	0.7801
	SD	0.0670	0.0800	0.0301	0.0744	0.0619	0.1605
	MSE	0.2546	0.0407	0.0015	0.0140	0.0073	0.0301
Pr2	**Mean**	0.2122	0.3128	0.0699	0.2791	0.1139	0.4699
	Med	0.2115	0.3118	0.0692	0.2730	0.1068	0.3980
	SD	0.0190	0.0247	0.0106	0.0563	0.0410	0.3220
	MSE	0.0018	0.1922	0.0327	0.2269	0.0214	0.2572
Pr3	**Mean**	1.7383	1.6307	0.3541	0.9012	0.4941	1.5402
	Med	1.7384	1.6304	0.3545	0.9012	0.4976	1.5611
	SD	0.1545	0.1465	0.0395	0.1002	0.1026	0.2740
	MSE	2.2424	0.8107	0.0131	0.0383	0.0712	0.7092
Pr4	**Mean**	1.1714	1.2506	0.3714	0.7385	0.4906	0.5434
	Med	1.1670	1.2469	0.3713	0.7390	0.4877	0.5377
	SD	0.1059	0.0996	0.0345	0.0664	0.0665	0.0875
	MSE	0.8655	0.2637	0.0167	0.0047	0.0629	0.0505
Pr5	**Mean**	0.5712	0.6491	0.2325	0.8235	0.2345	0.7590
	Med	0.5680	0.6412	0.2309	0.8203	0.2260	0.7057
	SD	0.0571	0.1042	0.0306	0.1066	0.0758	0.0857
	MSE	0.1067	0.1426	0.0021	0.0453	0.0070	0.0419

Table 5.28: Simulation results of imputing all unknown states - Q_6Sc_3.

Q6	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	λ_{34}	λ_{43}
Sc3	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.8664	0.3132	0.6933	0.2031	0.9256	0.5660
	Med	0.8607	0.3106	0.6911	0.2023	0.9234	0.5643
	SD	0.1046	0.0546	0.0539	0.0245	0.0660	0.0464
	MSE	0.0248	0.0083	0.0121	0.0033	0.0356	0.1027
Pr2	**Mean**	0.5631	0.1354	0.4702	0.0846	0.4251	0.2157
	Med	0.5556	0.1327	0.4678	0.0838	0.4240	0.2149
	SD	0.0684	0.0246	0.0397	0.0118	0.0284	0.0178
	MSE	0.0448	0.0143	0.0843	0.0276	0.1082	0.0017
Pr3	**Mean**	1.5567	0.5295	0.7845	0.2952	1.3397	1.1750
	Med	1.5636	0.5312	0.7837	0.2950	1.3352	1.1715
	SD	0.1819	0.0896	0.0631	0.0328	0.1119	0.0983
	MSE	0.6959	0.0877	0.0060	0.0036	0.3625	0.8798
Pr4	**Mean**	0.7724	0.2647	0.8406	0.2258	0.8185	0.4602
	Med	0.7623	0.2619	0.8398	0.2247	0.8149	0.4581
	SD	0.0168	0.0465	0.0525	0.0257	0.0790	0.0396
	MSE	0.0362	0.0038	0.0140	0.0023	0.1029	0.0461
Pr5	**Mean**	0.6949	0.5063	0.7075	0.2864	1.1409	0.7906
	Med	0.6906	0.5043	0.7063	0.2857	1.1385	0.7885
	SD	0.0767	0.0630	0.0430	0.0259	0.0728	0.0584
	MSE	0.0104	0.0745	0.0101	0.0032	0.1777	0.3011

Table 5.29: Simulation results of imputing all unknown states - Q_6Sc_4.

Q6	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	λ_{34}	λ_{43}
Sc4	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.9017	0.3764	0.6419	0.2539	0.8717	0.6933
	Med	0.8934	0.3735	0.6400	0.2529	0.8676	0.6899
	SD	0.1179	0.0605	0.0537	0.0297	0.0721	0.0655
	MSE	0.0374	0.0198	0.0151	0.0010	0.0208	0.2022
Pr2	**Mean**	0.5365	0.0934	0.3982	0.0905	0.3429	0.2091
	Med	0.5262	0.0904	0.3954	0.0897	0.3416	0.2084
	SD	0.0834	0.0256	0.0380	0.0133	0.0268	0.0215
	MSE	0.0601	0.0255	0.1258	0.0258	0.1679	0.0022
Pr3	**Mean**	1.6039	0.6356	0.7927	0.3458	1.4435	1.5812
	Med	1.6136	0.6393	0.7925	0.3460	1.4436	1.5801
	SD	0.1896	0.1013	0.0711	0.0399	0.1321	0.1481
	MSE	0.7685	0.1602	0.0073	0.0110	0.4997	1.7953
Pr4	**Mean**	0.7287	0.2815	0.9130	0.2616	0.7277	0.5064
	Med	0.7292	0.2782	0.9114	0.2605	0.7251	0.5044
	SD	0.0657	0.0572	0.0637	0.0298	0.0604	0.0458
	MSE	0.0478	0.0051	0.0338	0.0023	0.0917	0.0682
Pr5	**Mean**	0.6452	0.6343	0.7138	0.3770	1.1753	1.0775
	Med	0.6412	0.6319	0.7137	0.3765	1.1748	1.0763
	SD	0.0764	0.0757	0.0526	0.0352	0.0949	0.1016
	MSE	0.0182	0.1553	0.0054	0.0179	0.1918	0.6956

Table 5.30: Bias for a select number of models when imputing all unknown states.

Model	**Scenario**	**Prior**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Q1	1	1	0.301	0.049	0.058	0.230		
		2	0.165	0.282	0.260	0.145		
		3	0.890	0.332	0.279	0.804		
		4	0.371	0.063	0.285	0.110		
		5	0.075	0.279	0.053	0.196		
	3	1	0.175	0.057	0.039	0.181		
		2	0.138	0.254	0.242	0.154		
		3	0.894	0.366	0.294	0.884		
		4	0.370	0.063	0.256	0.053		
		5	0.094	0.339	0.015	0.268		
Q6	3	1	0.118	0.073	0.096	0.051	0.177	0.317
		2	0.200	0.117	0.288	0.166	0.328	0.038
		3	0.814	0.282	0.045	0.050	0.592	0.933
		4	0.350	0.041	0.106	0.040	0.443	0.211
		5	0.067	0.266	0.091	0.051	0.415	0.546
	4	1	0.153	0.127	0.111	0.012	0.125	0.445
		2	0.231	0.158	0.353	0.160	0.409	0.042
		3	0.856	0.387	0.048	0.097	0.694	1.332
		4	0.469	0.043	0.172	0.038	0.532	0.257
		5	0.111	0.387	0.051	0.129	0.428	0.828

Table 5.31: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals (Fill model) - $Q_1Sc_1Pr_1$.

Fill Q1.Sc1.Pr1	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$
Cred Int (Coverage %)	0.8371	0.9707	0.9629	0.9332
HPD Int (Coverage %)	0.8527	0.9727	0.9746	0.8410
Cred Int (Mean length)	0.2684	0.1578	0.1874	0.1687
HPD Int (Mean length)	0.3014	0.2687	0.1987	0.3598

Table 5.32: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals (Fill model) - $Q_5Sc_3Pr_5$.

Fill Q5.Sc3.Pr5	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Cred Int (Coverage %)	0.8988	0.9431	0.9243	0.9177	0.9047	0.9087
HPD Int (Coverage %)	0.8477	0.9490	0.9245	0.9196	0.9287	0.9578
Cred Int (Mean length)	0.3581	0.2380	0.2061	0.1029	0.2850	0.2344
HPD Int (Mean length)	0.1846	0.2487	0.1717	0.2174	0.2551	0.2253

5.4.2.2 Estimating the transition point

In Tables 5.33 to 5.53, models fitted when estimating the transition point between two known observations in the data are investigated. Tables 5.33 to 5.50, summarise the posterior distributions across the different scenarios, in Table 5.51 the bias associated with a select number of scenarios is presented and in Tables 5.52 to 5.53, the posterior coverage probabilities and mean lengths of the credible and HPD intervals are shown.

The results can be summarised as follows:

- With the exception of Q_5 (see Tables 5.47 and 5.48), the results here are similar to the results found in the previous section in that no real differences between the MSE's are observed across the different scenarios. This shows that, even for the smallest data sets (scenario 2), when imputing the transition point between two known observations, models can be fitted to the imputed data set without the sample size being as critical as was found in Sections 2.3.3 and 4.3.
- For models Q_1 and Q_4 (see Tables 5.33 to 5.36, 5.45 and 5.46), the rates at which transitions are made to higher and lower states are exactly the same. Although no differences are observed with regards to the MSE across the 5 different priors[6] (*see footnote 6*), the effect of the priors is visible when looking at the mean values of the posterior parameters:
  - Prior 1, has in general, larger posterior transition rate variates than the other priors. This prior assumes that transitions occur at the beginning of an interval, indicating a prior belief in higher transition rates.
  - Prior 2, has in general, posterior transition rate variates between those of priors 1 and 3. This prior assumes that transitions occur anywhere within the specified interval, indicating no real prior belief about the transition rates.
  - Prior 3, has in general, smaller posterior transition rate variates than the other priors. This prior assumes that transitions occur nearer to the end of an interval, indicating a prior belief in lower transition rates.
  - Prior 4, has larger posterior transition rate variates for the forward rates and smaller posterior transition rate variates for the backwards rates. This prior assumes forward

[6] The fact that the MSE's are so similar is due to the population parameters being 0.5 for all transition rates. This means that if a value is estimated as 0.4 and another value as 0.6, the MSE value associated with both of these estimates will be exactly the same.

transitions occur nearer to the beginning of an interval, indicating higher transition rates, and backward transitions nearer to the end of an interval, indicating lower transition rates.

- Prior 5, has smaller posterior transition rate variates for the forward rates and larger posterior transition rate variates for the backwards rates. This prior assumes forward transitions occur nearer to the end of an interval, indicating lower transition rates, and backward transitions nearer to the beginning of an interval, indicating higher transition rates.

– When modelling Q_2 and Q_5 (see Tables 5.37 to 5.40, 5.47 and 5.48) prior 5 appears to be the prior with the smallest MSE values. This prior assumes forward transitions occur nearer to the end of an interval, indicating lower transition rates, and backward transitions nearer to the beginning of an interval, indicating higher transition rates.

– When modelling Q_3 and Q_6 (see Tables 5.41 to 5.44, 5.49 and 5.50) prior 4 appears to be the prior with the smallest MSE values. This prior assumes forward transitions occur nearer to the beginning of an interval, indicating higher transition rates, and backward transitions nearer to the end of an interval, indicating lower transition rates.

– The bias associated with model Q_1, data scenarios 1 and 3, and model Q_6, data scenarios 3 and 4, show no extreme bias values (5.51). The values are generally smaller than those observed when all the missing observations are imputed (see Table 5.30).

– Unlike in the previous section, where big differences were observed when the results of using the "correct" (priors that closely matches the parameters in the model) and "incorrect" (priors that do not correspond to the parameters in the model) priors were compared, here those differences are not as pronounced. This indicates that imputing the transition point is more robust with regards to, and hence less sensitive to, the choice of the prior distribution than when imputing all unknown states. This is due to the fact that here the prior is only used to impute a transition point between two observations, while with the previous technique, the prior was used to impute every single unknown value.

– The frequentist coverage probabilities are of similar order for both models Q_1 and Q_5 (see Tables 5.52 and 5.53), with some of the intervals having lower values than the expected 95%.

From these results it is clear that this method of modelling multi-state data gives good all-round parameter estimates of the population transition rates. When the correct prior is used, i.e. the prior that best matches the underlying population model, the MSE's of the models become smaller, indicating that correctly specifying the prior does lead to better parameter estimates. When an incorrect prior is used, as is the case with prior 3 and 4 with model Q_5 (see Tables 5.47 and 5.27), the models become more unstable and the SD and MSE's increase. This instability is not as pronounced as in the previous section, indicating that if the transition point between two known observations is estimated, an incorrectly specified prior does not have as big an effect on the parameter estimates as when all unknown observations are imputed. This shows that this technique is less sensitive to the choice of prior distribution.

Table 5.33: Simulation results of imputing the transition point - Q_1Sc_1.

Q1	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc1	**TP**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.6709	0.4782	0.4665	0.4286	**Pr2**	0.3923	0.3092	0.3734	0.4388
	Med	0.6705	0.4781	0.4666	0.4282		0.3916	0.3086	0.3726	0.4378
	SD	0.0318	0.0203	0.0122	0.0174		0.0143	0.0108	0.0127	0.0189
	MSE	0.0344	0.0021	0.0276	0.0174		0.0219	0.0397	0.0237	0.0064
Pr3	**Mean**	0.4181	0.3163	0.4000	0.4699	**Pr4**	0.5427	0.4031	0.4978	0.4437
	Med	0.4180	0.3164	0.3999	0.4695		0.5428	0.4032	0.4981	0.4436
	SD	0.0077	0.0061	0.0097	0.0140		0.0196	0.0106	0.0111	0.0121
	MSE	0.0165	0.0409	0.0287	0.0090		0.0103	0.0131	0.0114	0.0050
Pr5	**Mean**	0.3743	0.4826	0.4194	0.4778					
	Med	0.3744	0.4827	0.4195	0.4777					
	SD	0.0057	0.0087	0.0084	0.0152					
	MSE	0.0231	0.0174	0.0117	0.0076					

Table 5.34: Simulation results of imputing the transition point - Q_1Sc_2.

Q1	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc2	**TP**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.6934	0.4508	0.4656	0.4479	**Pr2**	0.4455	0.4179	0.2883	0.3635
	Med	0.6888	0.4477	0.4631	0.4445		0.4403	0.4132	0.2860	0.3598
	SD	0.0768	0.0508	0.0452	0.0504		0.0446	0.0401	0.0203	0.0355
	MSE	0.0481	0.0141	0.0259	0.0356		0.0121	0.0293	0.0467	0.0227
Pr3	**Mean**	0.3660	0.4854	0.3766	0.4341	**Pr4**	0.5592	0.4170	0.4723	0.3421
	Med	0.3649	0.4843	0.3759	0.4326		0.5564	0.4555	0.4712	0.3404
	SD	0.0164	0.0247	0.0180	0.0246		0.0528	0.0364	0.0247	0.0237
	MSE	0.0310	0.0107	0.0173	0.0298		0.0214	0.0222	0.0213	0.0282
Pr5	**Mean**	0.3082	0.5206	0.4499	0.5302					
	Med	0.3080	0.5203	0.4483	0.5280					
	SD	0.0116	0.0165	0.0316	0.0484					
	MSE	0.0410	0.0436	0.0150	0.0200					

Table 5.35: Simulation results of imputing the transition point - Q_1Sc_3.

Q1	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc3	**TP**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.4496	0.4370	0.4952	0.4992	**Pr2**	0.4617	0.3859	0.3925	0.4207
	Med	0.4497	0.4370	0.4952	0.4991		0.4613	0.3856	0.3921	0.4201
	SD	0.0103	0.0095	0.0080	0.0106		0.0138	0.0105	0.0100	0.0137
	MSE	0.0035	0.0094	0.0154	0.0200		0.0035	0.0146	0.0132	0.0085
Pr3	**Mean**	0.4395	0.4101	0.3862	0.4312	**Pr4**	0.6608	0.5217	0.5401	0.3322
	Med	0.4394	0.4101	0.3861	0.4308		0.6608	0.5218	0.5401	0.3321
	SD	0.0067	0.0067	0.0066	0.0104		0.0187	0.0121	0.0060	0.0058
	MSE	0.0054	0.0093	0.0177	0.0102		0.0452	0.0047	0.0351	0.0331
Pr5	**Mean**	0.4101	0.4328	0.3736	0.4283					
	Med	0.4101	0.4329	0.3737	0.4281					
	SD	0.0059	0.0091	0.0060	0.0115					
	MSE	0.0099	0.0060	0.0164	0.0091					

Table 5.36: Simulation results of imputing the transition point - Q_1Sc_4.

Q1	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc4	**TP**	**0.5**	**0.5**	**0.5**	**0.5**		**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.6829	0.5864	0.4329	0.4102	**Pr2**	0.4337	0.3428	0.3578	0.4132
	Med	0.6814	0.5859	0.4318	0.4087		0.4322	0.3415	0.3564	0.4113
	SD	0.0496	0.0391	0.0261	0.0303		0.0251	0.0187	0.0196	0.0285
	MSE	0.0437	0.0217	0.0064	0.0111		0.0077	0.0273	0.0228	0.0107
Pr3	**Mean**	0.4387	0.3723	0.3445	0.4484	**Pr4**	0.6796	0.5366	0.3604	0.3503
	Med	0.4382	0.3719	0.3441	0.4474		0.6771	0.5352	0.3600	0.3497
	SD	0.0144	0.0116	0.0112	0.0195		0.0465	0.0319	0.0157	0.0156
	MSE	0.0090	0.0262	0.0249	0.0080		0.0372	0.0064	0.0221	0.0241
Pr5	**Mean**	0.4030	0.3898	0.5034	0.6842					
	Med	0.4026	0.3896	0.5026	0.6822					
	SD	0.0143	0.0173	0.0292	0.0517					
	MSE	0.0143	0.0172	0.0286	0.1330					

Table 5.37: Simulation results of imputing the transition point - Q_2Sc_1.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc1	**TP**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.2879	0.7755	0.1617	0.5981	**Pr2**	0.2219	0.6250	0.1139	0.5301
	Med	0.2871	0.7729	0.1605	0.5815		0.2215	0.6235	0.1131	0.5225
	SD	0.0166	0.0456	0.0104	0.0808		0.0084	0.0265	0.0063	0.0611
	MSE	0.0072	0.0348	0.0121	0.0352		0.0028	0.0274	0.0196	0.0581
Pr3	**Mean**	0.2270	0.7043	0.2150	0.6904	**Pr4**	0.2758	0.6881	0.1268	0.5792
	Med	0.2270	0.7043	0.2147	0.6875		0.2759	0.6882	0.1261	0.5675
	SD	0.0050	0.0249	0.0080	0.0413		0.0083	0.0175	0.0045	0.0454
	MSE	0.0031	0.0256	0.0035	0.0332		0.0038	0.0106	0.0159	0.0384
Pr5	**Mean**	0.2413	0.7565	0.1565	0.5756					
	Med	0.2413	0.7566	0.1564	0.5741					
	SD	0.0068	0.0304	0.0075	0.0497					
	MSE	0.0053	0.0398	0.0115	0.0632					

Table 5.38: Simulation results of imputing the transition point - Q_2Sc_2.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc2	**TP**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.4589	1.1182	0.1142	0.4466	**Pr2**	0.2247	0.5366	0.1347	0.5504
	Med	0.4570	1.1123	0.1136	0.4395		0.2220	0.5302	0.1214	0.4656
	SD	0.0584	0.1364	0.0076	0.0543		0.0208	0.0519	0.0375	0.2432
	MSE	0.0664	0.1977	0.0188	0.1052		0.0017	0.0571	0.0162	0.1307
Pr3	**Mean**	0.1925	0.6088	0.1911	0.6023	**Pr4**	0.3735	0.7623	0.1568	0.4772
	Med	0.1921	0.6077	0.1907	0.5990		0.3725	0.7604	0.1543	0.4619
	SD	0.0073	0.0348	0.0126	0.0581		0.0312	0.0575	0.0180	0.0787
	MSE	0.0045	0.0607	0.0062	0.0890		0.0223	0.0210	0.0113	0.1085
Pr5	**Mean**	0.2473	0.8676	0.2300	0.7108					
	Med	0.2461	0.8630	0.2274	0.7081					
	SD	0.0195	0.0799	0.0241	0.0611					
	MSE	0.0059	0.1218	0.0052	0.0661					

Table 5.39: Simulation results of imputing the transition point - Q_2Sc_3.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc3	**TP**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.2524	0.6777	0.1489	0.5200	**Pr2**	0.2408	0.6500	0.1312	0.4654
	Med	0.2524	0.6776	0.1490	0.5193		0.2406	0.6493	0.1308	0.4628
	SD	0.0062	0.0182	0.0049	0.0334		0.0053	0.0180	0.0043	0.0325
	MSE	0.0008	0.0058	0.0124	0.0671		0.0023	0.0148	0.0148	0.0872
Pr3	**Mean**	0.1750	0.6191	0.1878	0.7715	**Pr4**	0.2752	0.6366	0.1572	0.4786
	Med	0.1750	0.6191	0.1875	0.7699		0.2753	0.6368	0.1572	0.4764
	SD	0.0019	0.0118	0.0060	0.0434		0.0065	0.0132	0.0047	0.0230
	MSE	0.0059	0.0196	0.0064	0.0486		0.0023	0.0200	0.0093	0.0758
Pr5	**Mean**	0.2364	0.8185	0.1756	0.7164					
	Med	0.2364	0.8186	0.1757	0.7174					
	SD	0.0045	0.0232	0.0057	0.0373					
	MSE	0.0006	0.0168	0.0077	0.0311					

Table 5.40: Simulation results of imputing the transition point - Q_2Sc_4.

Q2	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc4	**TP**	**0.25**	**0.75**	**0.25**	**0.75**		**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.4349	1.1049	0.1334	0.4035	**Pr2**	0.2259	0.6450	0.1292	0.4907
	Med	0.4342	1.1039	0.1325	0.3977		0.2245	0.6414	0.1279	0.4824
	SD	0.0344	0.0808	0.0082	0.0414		0.0147	0.0418	0.0099	0.0649
	MSE	0.0523	0.1964	0.0148	0.1388		0.0022	0.0238	0.0157	0.0895
Pr3	**Mean**	0.2314	0.5873	0.1758	0.6808	**Pr4**	0.3168	0.6979	0.1437	0.4933
	Med	0.2312	0.5867	0.1701	0.6446		0.3163	0.6972	0.1432	0.4881
	SD	0.0062	0.0207	0.0218	0.1241		0.0194	0.0375	0.0063	0.0351
	MSE	0.0012	0.0347	0.0137	0.0421		0.0114	0.0292	0.0124	0.0845
Pr5	**Mean**	0.2666	0.8868	0.1807	0.6537					
	Med	0.2656	0.8833	0.1804	0.6522					
	SD	0.0165	0.0636	0.0098	0.0536					
	MSE	0.0026	0.0237	0.0074	0.0417					

Table 5.41: Simulation results of imputing the transition point - Q_3Sc_1.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc1	**TP**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.6215	0.1193	0.7125	0.2466	**Pr2**	0.7723	0.1829	0.6458	0.2227
	Med	0.6204	0.1193	0.7121	0.2465		0.7689	0.1822	0.6444	0.2223
	SD	0.0333	0.0047	0.0249	0.0092		0.0532	0.0087	0.0262	0.0088
	MSE	0.0324	0.0176	0.0038	0.0007		0.0305	0.0070	0.0152	0.0027
Pr3	**Mean**	0.6802	0.1766	0.7101	0.2578	**Pr4**	0.9296	0.2471	0.7601	0.2507
	Med	0.6790	0.1766	0.7100	0.2575		0.9306	0.2472	0.7595	0.2505
	SD	0.0173	0.0053	0.0202	0.0065		0.0679	0.0119	0.0273	0.0078
	MSE	0.0216	0.0107	0.0297	0.0020		0.0859	0.0063	0.0189	0.0104
Pr5	**Mean**	0.6173	0.1351	0.6465	0.2516					
	Med	0.6121	0.1350	0.6469	0.2519					
	SD	0.0239	0.0040	0.0172	0.0088					
	MSE	0.0273	0.0137	0.0181	0.0040					

Table 5.42: Simulation results of imputing the transition point - Q_3Sc_2.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc2	**TP**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.6265	0.1646	1.1599	0.4615	**Pr2**	0.7150	0.1706	0.6386	0.2513
	Med	0.6121	0.1620	1.1582	0.4606		0.7041	0.1682	0.6327	0.2483
	SD	0.0835	0.0201	0.0969	0.0467		0.0932	0.0194	0.0552	0.0251
	MSE	0.0757	0.0093	0.2735	0.0776		0.0336	0.0081	0.0456	0.0074
Pr3	**Mean**	0.5578	0.1601	0.8449	0.3068	**Pr4**	0.8374	0.1808	0.7750	0.2516
	Med	0.5560	0.1598	0.8356	0.3031		0.8371	0.1807	0.7727	0.2506
	SD	0.0289	0.0089	0.0790	0.0313		0.0810	0.0158	0.0553	0.0176
	MSE	0.0508	0.0108	0.0695	0.0118		0.0402	0.0058	0.0133	0.0046
Pr5	**Mean**	0.5766	0.1574	1.0152	0.5363					
	Med	0.5683	0.1568	1.0163	0.5358					
	SD	0.0511	0.0148	0.0913	0.0591					
	MSE	0.1278	0.0130	0.4470	0.2980					

Table 5.43: Simulation results of imputing the transition point - Q_3Sc_3.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc3	**TP**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.6442	0.1601	0.7644	0.2618	**Pr2**	0.8072	0.1760	0.6098	0.2149
	Med	0.6433	0.1602	0.7641	0.2618		0.8054	0.1756	0.6092	0.2147
	SD	0.0240	0.0043	0.0200	0.0074		0.0411	0.0070	0.0164	0.0056
	MSE	0.0228	0.0088	0.0058	0.0015		0.0236	0.0059	0.0280	0.0020
Pr3	**Mean**	0.6025	0.1669	0.6385	0.2087	**Pr4**	0.9051	0.2169	0.7186	0.2296
	Med	0.6023	0.1669	0.6385	0.2087		0.9056	0.2170	0.7185	0.2296
	SD	0.0147	0.0035	0.0127	0.0037		0.0311	0.0056	0.0163	0.0042
	MSE	0.0260	0.0080	0.0281	0.0039		0.0439	0.0038	0.0141	0.0027
Pr5	**Mean**	0.5813	0.1936	0.5509	0.1998					
	Med	0.5804	0.1937	0.5510	0.1998					
	SD	0.0147	0.0047	0.0099	0.0044					
	MSE	0.0290	0.0043	0.0407	0.0044					

Table 5.44: Simulation results of imputing the transition point - Q_3Sc_4.

Q3	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}
Sc4	**TP**	**0.75**	**0.25**	**0.75**	**0.25**		**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.7232	0.1766	0.9677	0.3879	**Pr2**	0.6695	0.1801	0.6366	0.2505
	Med	0.7209	0.1762	0.9667	0.3873		0.6657	0.1791	0.6341	0.2492
	SD	0.0497	0.0122	0.0641	0.0307		0.0517	0.0115	0.0369	0.0170
	MSE	0.0522	0.0072	0.0855	0.0316		0.0242	0.0054	0.0255	0.0023
Pr3	**Mean**	0.6019	0.1422	0.6261	0.1826	**Pr4**	0.8854	0.1874	0.7136	0.2174
	Med	0.6016	0.1421	0.6250	0.1821		0.8860	0.1874	0.7129	0.2170
	SD	0.0191	0.0060	0.0257	0.0082		0.0528	0.0110	0.0315	0.0098
	MSE	0.0371	0.0127	0.0246	0.0063		0.0666	0.0094	0.0121	0.0016
Pr5	**Mean**	0.5212	0.1082	0.8050	0.3799					
	Med	0.5203	0.1088	0.8034	0.3789					
	SD	0.0220	0.0050	0.0450	0.0267					
	MSE	0.0722	0.0205	0.0332	0.0315					

Table 5.45: Simulation results of imputing the transition point - Q_4Sc_3.

Q4	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{34}	λ_{43}
Sc3	**TP**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.5339	0.4456	0.3894	0.3676	0.4588	0.4767
	Med	0.5337	0.4455	0.3893	0.3673	0.4587	0.4764
	SD	0.0186	0.0144	0.0086	0.0103	0.0108	0.0151
	MSE	0.0044	0.0081	0.0131	0.0178	0.0224	0.0173
Pr2	**Mean**	0.4791	0.3879	0.3141	0.3217	0.3940	0.5528
	Med	0.4784	0.3870	0.3135	0.3210	0.3926	0.5508
	SD	0.0161	0.0129	0.0081	0.0114	0.0172	0.0314
	MSE	0.0065	0.0139	0.0356	0.0383	0.0119	0.0149
Pr3	**Mean**	0.4537	0.4206	0.3370	0.3218	0.3958	0.4541
	Med	0.4535	0.4206	0.3370	0.3216	0.3955	0.4534
	SD	0.0073	0.0086	0.0051	0.0079	0.0098	0.0154
	MSE	0.0049	0.0104	0.0291	0.0323	0.0123	0.0219
Pr4	**Mean**	0.6278	0.4865	0.5273	0.3146	0.4673	0.3987
	Med	0.6275	0.4863	0.5273	0.3145	0.4673	0.3982
	SD	0.0204	0.0130	0.0060	0.0073	0.0101	0.0121
	MSE	0.0234	0.0145	0.0330	0.0354	0.0191	0.0240
Pr5	**Mean**	0.4871	0.4897	0.3406	0.4473	0.4099	0.6013
	Med	0.4869	0.4896	0.3406	0.4471	0.4097	0.6007
	SD	0.0105	0.0142	0.0054	0.0092	0.0122	0.0273
	MSE	0.0021	0.0043	0.0257	0.0263	0.0147	0.0396

Table 5.46: Simulation results of imputing the transition point - Q_4Sc_4.

Q4	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{34}	λ_{43}
Sc4	**TP**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
Pr1	**Mean**	0.6149	0.4983	0.3518	0.3219	0.4558	0.5042
	Med	0.6137	0.4977	0.3517	0.3215	0.4549	0.5024
	SD	0.0383	0.0296	0.0109	0.0126	0.0272	0.0392
	MSE	0.0240	0.0040	0.0256	0.0383	0.0090	0.0156
Pr2	**Mean**	0.4696	0.4085	0.3497	0.2943	0.3696	0.5387
	Med	0.4674	0.4066	0.3490	0.2933	0.3665	0.5326
	SD	0.0291	0.0242	0.0138	0.0159	0.0295	0.0564
	MSE	0.0089	0.0129	0.0264	0.0442	0.0186	0.0106
Pr3	**Mean**	0.4136	0.3734	0.4321	0.3831	0.5153	0.5546
	Med	0.4130	0.3729	0.4318	0.3823	0.5130	0.5515
	SD	0.0118	0.0136	0.0130	0.0174	0.0321	0.0407
	MSE	0.0121	0.0192	0.0096	0.0239	0.0090	0.0092
Pr4	**Mean**	0.6204	0.4764	0.5357	0.3583	0.4417	0.3593
	Med	0.6186	0.4753	0.5353	0.3577	0.4411	0.3578
	SD	0.0429	0.0293	0.0132	0.0152	0.0179	0.0230
	MSE	0.0200	0.0046	0.0282	0.0245	0.0279	0.0269
Pr5	**Mean**	0.4387	0.4472	0.3460	0.3711	0.5530	0.5086
	Med	0.4380	0.4465	0.3458	0.3705	0.5504	0.5047
	SD	0.0196	0.0249	0.0103	0.0182	0.0422	0.0406
	MSE	0.0124	0.0100	0.0247	0.0248	0.0049	0.0135

Table 5.47: Simulation results of imputing the transition point - Q_5Sc_3.

Q5	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{34}	λ_{43}
Sc3	**TP**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.2673	0.7261	0.1420	0.5098	0.1195	0.4971
	Med	0.2672	0.7262	0.1418	0.5076	0.1182	0.4877
	SD	0.0077	0.0211	0.0041	0.0276	0.0096	0.0726
	MSE	0.0020	0.0015	0.0117	0.0607	0.0189	0.1690
Pr2	**Mean**	0.2450	0.6987	0.1571	0.5139	0.1398	0.5362
	Med	0.2444	0.6968	0.1565	0.5107	0.1366	0.5167
	SD	0.0086	0.0281	0.0069	0.0396	0.0181	0.1254
	MSE	0.0011	0.0075	0.0091	0.0758	0.0153	0.0827
Pr3	**Mean**	0.2299	0.6492	0.1516	0.5196	0.1818	0.7355
	Med	0.2299	0.6494	0.1515	0.5193	0.1855	0.7578
	SD	0.0025	0.0130	0.0030	0.0195	0.0116	0.0743
	MSE	0.0005	0.0122	0.0111	0.0612	0.0122	0.2709
Pr4	**Mean**	0.2594	0.6053	0.1235	0.4241	104.7	3748.8
	Med	0.2595	0.6053	0.1231	0.4208	0.1231	0.8497
	SD	0.0063	0.0146	0.0047	0.0261	725.5	26015.0
	MedSE	0.001**	0.016**	0.010**	0.028**	0.007**	0.034**
Pr5	**Mean**	0.2340	0.7667	0.1745	0.7220	0.2353	0.9195
	Med	0.2340	0.7666	0.1744	0.7223	0.2337	0.9188
	SD	0.0033	0.0189	0.0042	0.0335	0.0175	0.0973
	MSE	0.0004	0.0091	0.0061	0.0307	0.0222	0.2771

Table 5.48: Simulation results of imputing the transition point - Q_5Sc_4.

Q5	**Par**	λ_{12}	λ_{21}	λ_{23}	λ_{32}	λ_{34}	λ_{43}
Sc4	**TP**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**
Pr1	**Mean**	0.4792	1.1705	0.1221	0.3736	0.1918	0.5362
	Med	0.4781	1.1682	0.1218	0.3706	0.1881	0.5158
	SD	0.0440	0.0994	0.0050	0.0306	0.0269	0.1130
	MSE	0.0609	0.2293	0.0167	0.1477	0.0046	0.0898
Pr2	**Mean**	0.2511	0.6351	0.0994	0.4185	0.1809	0.6689
	Med	0.2494	0.6303	0.0987	0.4124	0.1730	0.6287
	SD	0.0204	0.0517	0.0063	0.0510	0.0352	0.1836
	MSE	0.0008	0.0224	0.0230	0.1322	0.0160	0.1855
Pr3	**Mean**	0.2092	0.5827	0.1719	0.6004	816.5	2936.9
	Med	0.2089	0.5814	0.1710	0.5987	1.0808	4.0159
	SD	0.0064	0.0243	0.0098	0.0493	2793.4	9929.1
	MedSE	0.003**	0.014**	0.002**	0.036**	0.008**	0.023**
Pr4	**Mean**	0.2578	0.6156	0.1294	0.3992	0.3360	2.5998
	Med	0.2572	0.6138	0.1287	0.3944	0.1259	0.6219
	SD	0.0174	0.0374	0.0082	0.0357	13.2	125.3
	MedSE	0.005**	0.006**	0.010**	0.039**	0.009**	0.069**
Pr5	**Mean**	0.2457	0.8035	0.1890	0.6039	0.5027	0.7419
	Med	0.2452	0.8018	0.1887	0.6013	0.5040	0.7443
	SD	0.0132	0.0508	0.0108	0.0586	0.0196	0.1087
	MSE	0.0023	0.0220	0.0055	0.0260	0.0682	0.0883

Table 5.49: Simulation results of imputing the transition point - Q_6Sc_3.

Q6	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Sc3	**TP**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.7958	0.2008	0.6606	0.1200	0.7031	0.2383
	Med	0.7966	0.2009	0.6603	0.1199	0.7029	0.2382
	SD	0.0337	0.0088	0.0183	0.0034	0.0167	0.0061
	MSE	0.0042	0.0062	0.0145	0.0184	0.0039	0.0008
Pr2	**Mean**	0.7214	0.1739	0.5379	0.1079	0.5678	0.2257
	Med	0.7191	0.1730	0.5368	0.1075	0.5673	0.2254
	SD	0.0457	0.0101	0.0198	0.0050	0.0184	0.0079
	MSE	0.0222	0.0067	0.0480	0.0204	0.0387	0.0020
Pr3	**Mean**	0.8653	0.2283	0.6057	0.1549	0.6272	0.2273
	Med	0.8651	0.2280	0.6058	0.1551	0.6271	0.2272
	SD	0.0357	0.0120	0.0143	0.0060	0.0136	0.0053
	MSE	0.1222	0.0242	0.0241	0.0093	0.0200	0.0040
Pr4	**Mean**	0.8800	0.1746	0.6614	0.1547	0.6191	0.2229
	Med	0.8812	0.1746	0.6615	0.1545	0.6188	0.2228
	SD	0.0434	0.0075	0.0199	0.0049	0.0168	0.0050
	MSE	0.0294	0.0092	0.0207	0.0098	0.0278	0.0015
Pr5	**Mean**	0.6659	0.2099	0.5331	0.1285	0.6615	0.2464
	Med	0.6654	0.2096	0.5326	0.1284	0.6614	0.2464
	SD	0.0216	0.0113	0.0128	0.0041	0.0155	0.0072
	MSE	0.0236	0.0038	0.0541	0.0159	0.0217	0.0016

Table 5.50: Simulation results of imputing the transition point - Q_6Sc_4.

Q6	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Sc4	**TP**	**0.75**	**0.25**	**0.75**	**0.25**	**0.75**	**0.25**
Pr1	**Mean**	0.8350	0.2129	0.6476	0.1620	0.6818	0.2687
	Med	0.8292	0.2114	0.6468	0.1613	0.6805	0.2676
	SD	0.0909	0.0249	0.0338	0.0118	0.0406	0.0204
	MSE	0.0469	0.0029	0.0128	0.0112	0.0163	0.0098
Pr2	**Mean**	0.9120	0.1526	0.5853	0.1792	0.5004	0.2015
	Med	0.9013	0.1501	0.5832	0.1781	0.4986	0.2007
	SD	0.1128	0.0220	0.0349	0.0131	0.0255	0.0120
	MSE	0.1845	0.0172	0.0383	0.0081	0.0690	0.0035
Pr3	**Mean**	0.6643	0.2203	0.6695	0.2373	0.6257	0.2065
	Med	0.6628	0.2193	0.6681	0.2363	0.6242	0.2059
	SD	0.0357	0.0160	0.0334	0.0188	0.0303	0.0106
	MSE	0.0261	0.0035	0.0298	0.0099	0.0180	0.0024
Pr4	**Mean**	0.9671	0.1471	0.6671	0.2025	0.6624	0.2445
	Med	0.9646	0.1463	0.6662	0.2018	0.6619	0.2440
	SD	0.0856	0.0133	0.0327	0.0118	0.0289	0.0116
	MSE	0.0635	0.0112	0.0095	0.0044	0.0151	0.0008
Pr5	**Mean**	0.6390	0.1772	0.5856	0.2049	0.8352	0.4034
	Med	0.6348	0.1759	0.5848	0.2044	0.8316	0.4015
	SD	0.0387	0.0145	0.0229	0.0129	0.0688	0.0417
	MSE	0.0477	0.0070	0.0315	0.0045	0.2201	0.1222

Table 5.51: Bias for a select number of models when imputing the transition point.

Model	**Scenario**	**Prior**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Q1	1	1	0.183	0.041	0.166	0.131		
		2	0.147	0.199	0.153	0.078		
		3	0.128	0.202	0.169	0.094		
		4	0.100	0.114	0.106	0.070		
		5	0.152	0.132	0.108	0.086		
	3	1	0.058	0.097	0.124	0.141		
		2	0.057	0.121	0.115	0.091		
		3	0.073	0.096	0.133	0.100		
		4	0.212	0.068	0.187	0.182		
		5	0.099	0.077	0.128	0.095		
Q6	3	1	0.056	0.078	0.119	0.135	0.060	0.028
		2	0.142	0.081	0.218	0.143	0.196	0.044
		3	0.348	0.155	0.154	0.096	0.141	0.063
		4	0.166	0.096	0.143	0.099	0.166	0.038
		5	0.152	0.061	0.232	0.126	0.146	0.040
	4	1	0.196	0.048	0.108	0.105	0.121	0.097
		2	0.414	0.129	0.193	0.089	0.261	0.058
		3	0.158	0.057	0.169	0.098	0.131	0.048
		4	0.237	0.105	0.092	0.065	0.120	0.025
		5	0.215	0.082	0.176	0.066	0.464	0.347

Table 5.52: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals (TP model) - $Q_1Sc_1Pr_1$.

TP Q1.Sc1.Pr1	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$
Cred Int (Coverage %)	0.8734	0.8387	0.8621	0.8520
HPD Int (Coverage %)	0.8695	0.8406	0.8602	0.8520
Cred Int (Mean length)	0.2757	0.1687	0.1248	0.2687
HPD Int (Mean length)	0.3544	0.1570	0.2004	0.4015

Table 5.53: Coverage probabilities ($\alpha = 0.05$) and mean length of posterior Credible and HPD intervals (TP model) - $Q_5Sc_3Pr_5$.

TP Q5.Sc3.Pr5	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\lambda_{34}}$	$\boldsymbol{\lambda_{43}}$
Cred Int (Coverage %)	0.9208	0.9667	0.9008	0.9118	0.9039	0.8980
HPD Int (Coverage %)	0.9510	0.8490	0.9029	0.9118	0.9320	0.8961
Cred Int (Mean length)	0.3542	0.2366	0.2049	0.1024	0.2832	0.2328
HPD Int (Mean length)	0.1738	0.2445	0.1704	0.2168	0.1546	0.3252

5.4.3 The effect on covariates

The aim of this chapter is not to develop models where covariates influence the prior effect on the transition rates (this will be the aim of follow-up research), but as most practical multi-state data sets include covariates, it is important to investigate what effect the proposed imputation techniques has on models when covariates are included in the modelling process. To this end, a select number of models are refitted to data sets that now include covariates. The two covariate models, model A and model B, as defined in Section 2.3.1.1 are investigated here.

The 6 models that will be refitted using data sets with firstly all missing observations imputed and secondly the transition points imputed between two known states, are:

1) Models Q_1, Q_2, and Q_3 for data scenario 2 with the categorical variable included in the data set. This scenario is chosen, as in Section 2.3.3 (see Tables 2.11 to 2.13) this scenario had very high MSE values across the 3 different models.
2) Models Q_1, Q_2, and Q_3 for data scenario 4 with both the categorical and continuous variables included in the data. This scenario is chosen, as in Section 2.3.3 (see Tables 2.14 to 2.16) it was found to be a scenario with very high MSE values across the 3 different models.

The results can be summarised as follows:

- With the exception of prior 3, none of the extremely large MSE's associated with the transition rates found in Section 2.3.3 are observed for the models fitted here. This is to be expected, as the Bayesian imputation techniques enlarge the data sets and enrich the likelihood that is being modelled and thus increases the amount of information available when modelling the data.
- As was found in Section 5.4.2.1, large MSE values associated with the transition rates are observed across all 6 models when prior 3 is used to impute all unknown observations in the data set (see Tables 5.54, 5.56, 5.58, 5.60, 5.62 and 5.64). This is not the case with the MSE associated with the covariate effect parameters when using prior 3. This suggests that the choice of prior has a big effect on the estimated transition rates, but that the estimated covariate effects are influenced to a lesser extent by the choice of prior distribution.
- The MSE values of the covariate effects are significantly smaller than those found when a prior distribution was placed on the covariate effects in the B-MSM (see Section 4.3.2.3).

This may be due to the fact that the imputation techniques introduced here better handle the inclusion of covariates in the data or that more care should be taken when selecting the prior distributions for the B-MSM's of Chapter 4.

- As noted in the earlier sections (see Sections 5.4.2.1 and 5.4.2.2) when the prior is chosen that best matches the underlying population model (for model Q_2 this is prior 5 and for model Q_3 prior 4), the MSE values of the transition rates are generally smaller than for the other priors. There is also some evidence that if the prior that best matches the underlying population model is used, the estimates of the covariate effects have smaller MSE values.
- The MSE's, means and medians of the covariate effect parameters are on par with those found when no prior information was used in the modelling process (see Section 2.3.3). This indicates that the structure of the data sets with regards to how the covariates influence the transition rates were not significantly altered by imputing the unknown values in the data set.

One of the drawbacks of imputing unknown values is that the inherent structure of the data may be altered, especially as the influence of the covariates on the transition rates is not taken into account during the imputing process. The results found here show that although the effect of the covariates plays no role in the imputing process, the imputing of the unknown state does not have a big impact on the covariate effects on the transition rates - the values here are on par with those found when the data sets were modelled without imputing any values. This shows that even after imputing the unknown states, the underlying relationship between the transition rates and the covariates is similar to what it was before any values were imputed.

Table 5.54: Simulation results of imputing all unknown states with 1 covariate - Q_1Sc_2.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta}_{12}^{Cat}$	$\boldsymbol{\beta}_{21}^{Cat}$	$\boldsymbol{\beta}_{23}^{Cat}$	$\boldsymbol{\beta}_{32}^{Cat}$
Sc2	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.809	0.558	0.613	0.734	-0.064	0.053	-0.189	-0.012
	Med	0.794	0.548	0.602	0.722	-0.058	0.053	-0.188	-0.007
	SD	0.153	0.101	0.107	0.133	0.258	0.258	0.253	0.260
	MSE	0.120	0.019	0.028	0.074	0.476	0.659	0.337	0.562
Pr2	**Mean**	0.345	0.183	0.212	0.263	-0.259	-0.035	-0.162	0.034
	Med	0.334	0.180	0.209	0.257	-0.261	-0.048	-0.168	0.029
	SD	0.086	0.042	0.042	0.061	0.338	0.316	0.295	0.335
	MSE	0.032	0.103	0.086	0.061	0.349	0.573	0.540	0.655
Pr3	**Mean**	1.989	1.164	1.087	2.139	-0.380	-0.373	-0.171	-0.363
	Med	1.933	1.132	1.063	2.085	-0.374	-0.361	-0.168	-0.360
	SD	0.403	0.256	0.248	0.442	0.261	0.299	0.308	0.275
	MSE	2.381	0.522	0.413	2.912	0.190	0.203	0.375	0.200
Pr4	**Mean**	0.944	0.480	1.096	0.671	-0.198	-0.365	0.032	-0.220
	Med	0.928	0.468	1.065	0.657	-0.195	-0.356	0.048	-0.205
	SD	0.200	0.100	0.209	0.135	0.322	0.317	0.263	0.275
	MSE	0.243	0.022	0.419	0.048	0.360	0.215	0.605	0.309
Pr5	**Mean**	0.652	1.085	0.503	0.935	-0.127	-0.013	-0.262	-0.126
	Med	0.642	1.072	0.497	0.921	-0.133	-0.011	-0.266	-0.126
	SD	0.118	0.174	0.087	0.169	0.247	0.218	0.253	0.256
	MSE	0.038	0.375	0.011	0.220	0.393	0.520	0.256	0.394

Table 5.55: Simulation results of imputing the transition point with 1 covariate - Q_1Sc_2.

Q1	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta}_{12}^{Cat}$	$\boldsymbol{\beta}_{21}^{Cat}$	$\boldsymbol{\beta}_{23}^{Cat}$	$\boldsymbol{\beta}_{32}^{Cat}$
Sc2	**TP**	**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.602	0.567	0.448	0.370	-0.155	0.034	-0.138	0.092
	Med	0.593	0.559	0.445	0.367	-0.146	0.050	-0.149	0.081
	SD	0.101	0.089	0.050	0.050	0.260	0.241	0.215	0.237
	MSE	0.056	0.038	0.011	0.033	0.475	0.908	0.442	1.235
Pr2	**Mean**	0.430	0.320	0.303	0.305	-0.513	-0.243	-0.123	0.009
	Med	0.421	0.314	0.298	0.298	-0.520	-0.247	-0.144	-0.010
	SD	0.075	0.052	0.046	0.051	0.253	0.242	0.252	0.279
	MSE	0.013	0.041	0.046	0.049	0.357	0.610	0.576	0.626
Pr3	**Mean**	0.415	0.337	0.298	0.551	-0.439	-0.317	0.062	-0.576
	Med	0.411	0.334	0.292	0.536	-0.446	-0.327	0.056	-0.575
	SD	0.046	0.036	0.041	0.081	0.140	0.143	0.172	0.161
	MSE	0.010	0.035	0.046	0.020	0.111	0.180	0.652	0.049
Pr4	**Mean**	0.687	0.430	0.278	0.342	-0.911	-0.623	0.642	-0.168
	Med	0.674	0.424	0.277	0.340	-0.912	-0.627	0.636	-0.173
	SD	0.108	0.060	0.034	0.039	0.196	0.157	0.153	0.141
	MSE	0.085	0.056	0.071	0.048	0.087	0.054	1.826	0.360
Pr5	**Mean**	0.555	0.689	0.271	0.336	-0.962	-0.893	0.404	0.725
	Med	0.550	0.684	0.270	0.334	-0.959	-0.900	0.409	0.734
	SD	0.075	0.102	0.023	0.036	0.154	0.166	0.162	0.194
	MSE	0.009	0.052	0.053	0.045	0.107	0.086	1.249	2.112

Table 5.56: Simulation results of imputing all unknown states with 1 covariate - Q_2Sc_2.

Q2	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}
Sc2	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.755	1.029	0.289	0.984	-0.111	-0.359	0.261	-0.271
	Med	0.735	1.003	0.281	0.943	-0.109	-0.350	0.252	-0.259
	SD	0.146	0.187	0.079	0.254	0.266	0.255	0.367	0.328
	MSE	0.277	0.114	0.008	0.121	0.418	0.185	1.059	0.292
Pr2	**Mean**	0.205	0.287	0.079	0.402	-0.040	-0.180	0.212	-0.069
	Med	0.203	0.283	0.077	0.390	-0.037	-0.176	0.233	-0.061
	SD	0.038	0.041	0.023	0.098	0.293	0.237	0.473	0.378
	MSE	0.004	0.216	0.030	0.132	0.546	0.388	1.065	0.565
Pr3	**Mean**	1.826	1.631	0.727	1.995	-0.321	-0.284	-0.280	-0.441
	Med	1.783	1.601	0.709	1.927	-0.322	-0.284	-0.278	-0.424
	SD	0.355	0.316	0.175	0.431	0.281	0.290	0.384	0.325
	MSE	2.620	0.876	0.261	1.737	0.224	0.259	0.329	0.196
Pr4	**Mean**	0.987	1.051	0.538	0.714	-0.310	-0.515	0.195	-0.185
	Med	0.977	1.042	0.533	0.701	-0.306	-0.509	0.196	-0.191
	SD	0.137	0.129	0.087	0.126	0.212	0.206	0.236	0.239
	MSE	0.567	0.133	0.091	0.019	0.198	0.078	0.857	0.322
Pr5	**Mean**	0.609	0.668	0.245	0.797	-0.383	-0.366	-0.348	-0.232
	Med	0.591	0.625	0.238	0.760	-0.389	-0.365	-0.325	-0.208
	SD	0.124	0.109	0.066	0.086	0.195	0.262	0.232	0.368
	MSE	0.145	0.144	0.005	0.019	0.155	0.261	0.304	0.371

Table 5.57: Simulation results of imputing the transition point with 1 covariate - Q_2Sc_2.

Q2	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}
Sc2	**TP**	**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.343	0.981	0.104	0.586	-0.825	-0.686	0.723	-0.508
	Med	0.342	0.977	0.103	0.578	-0.841	-0.690	0.718	-0.505
	SD	0.054	0.127	0.021	0.111	0.216	0.168	0.237	0.202
	MSE	0.030	0.124	0.024	0.050	0.161	0.164	2.264	0.130
Pr2	**Mean**	0.237	0.589	0.124	0.566	-0.676	-0.839	-0.958	-0.028
	Med	0.232	0.578	0.123	0.561	-0.694	-0.858	-0.782	-0.052
	SD	0.041	0.085	0.017	0.072	0.239	0.217	1.197	0.264
	MSE	0.004	0.033	0.020	0.057	0.658	0.105	6.311	0.991
Pr3	**Mean**	0.162	0.526	0.347	1.037	-0.281	0.142	-0.353	-1.086
	Med	0.160	0.514	0.336	1.021	-0.305	0.129	-0.361	-1.115
	SD	0.021	0.084	0.087	0.195	0.184	0.239	0.335	0.264
	MSE	0.010	0.079	0.031	0.147	0.228	0.837	0.691	0.240
Pr4	**Mean**	0.270	0.659	0.164	0.685	-0.234	-0.564	-0.340	-0.333
	Med	0.267	0.647	0.141	0.584	-0.262	-0.593	-0.326	-0.298
	SD	0.042	0.102	0.081	0.346	0.263	0.248	0.598	0.545
	MSE	0.003	0.034	0.022	0.212	0.299	0.090	1.403	0.497
Pr5	**Mean**	0.326	0.886	0.201	0.699	-0.997	-1.098	-0.384	-0.433
	Med	0.321	0.871	0.192	0.655	-0.904	-1.103	-0.360	-0.421
	SD	0.049	0.141	0.050	0.114	0.173	0.183	0.115	0.131
	MSE	0.011	0.042	0.011	0.123	0.350	0.213	0.313	0.181

Table 5.58: Simulation results of imputing all unknown states with 1 covariate - Q_3Sc_2.

Q3	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$
Sc2	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.874	0.326	0.887	0.734	-0.067	0.161	-0.240	-0.160
	Med	0.855	0.320	0.873	0.721	-0.063	0.161	-0.236	-0.158
	SD	0.179	0.074	0.139	0.124	0.293	0.316	0.232	0.244
	MSE	0.050	0.011	0.042	0.251	0.488	0.842	0.283	0.353
Pr2	**Mean**	0.420	0.070	0.338	0.214	-0.038	0.073	-0.207	-0.067
	Med	0.401	0.068	0.334	0.211	-0.031	0.084	-0.200	-0.061
	SD	0.116	0.022	0.046	0.039	0.386	0.491	0.211	0.270
	MSE	0.123	0.033	0.172	0.003	0.626	0.839	0.297	0.476
Pr3	**Mean**	9.390	2.794	2.544	2.831	-0.308	-0.155	-0.584	-0.269
	Med	1.908	0.671	1.710	1.866	-0.258	-0.091	-0.535	-0.235
	SD	18.99	13.11	10.4	12.3	0.544	0.572	0.476	0.471
	MSE	659.4	523.1	117.5	159.4	0.452	0.659	0.263	0.420
Pr4	**Mean**	0.706	0.196	0.856	0.665	-0.336	-0.305	-0.400	-0.165
	Med	0.708	0.188	0.875	0.634	-0.337	-0.281	-0.399	-0.171
	SD	0.011	0.080	0.166	0.180	0.396	0.336	0.280	0.319
	MSE	0.041	0.009	0.175	0.207	0.376	0.335	0.336	0.389
Pr5	**Mean**	0.772	0.551	1.158	1.011	-0.309	0.058	-0.376	-0.236
	Med	0.747	0.538	1.122	0.981	-0.290	0.068	-0.340	-0.221
	SD	0.179	0.127	0.219	0.212	0.317	0.303	0.281	0.285
	MSE	0.033	0.108	0.219	0.631	0.260	0.671	0.190	0.304

Table 5.59: Simulation results of imputing the transition point with 1 covariate - Q_3Sc_2.

Q3	**Par**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta_{12}^{Cat}}$	$\boldsymbol{\beta_{21}^{Cat}}$	$\boldsymbol{\beta_{23}^{Cat}}$	$\boldsymbol{\beta_{32}^{Cat}}$
Sc2	**TP**	**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**
Pr1	**Mean**	0.593	0.125	1.055	0.422	-0.263	0.664	-0.982	-1.034
	Med	0.568	0.120	1.048	0.419	-0.260	0.667	-0.994	-1.047
	SD	0.134	0.031	0.135	0.063	0.346	0.344	0.189	0.207
	MSE	0.085	0.018	0.211	0.051	0.362	2.284	0.155	0.196
Pr2	**Mean**	0.599	0.126	0.677	0.235	-0.191	-0.357	-0.405	-0.518
	Med	0.575	0.123	0.661	0.227	-0.198	-0.397	-0.414	-0.526
	SD	0.136	0.024	0.106	0.042	0.250	0.297	0.209	0.240
	MSE	0.111	0.022	0.047	0.007	0.954	0.234	0.444	0.235
Pr3	**Mean**	0.565	0.277	0.926	0.133	-0.373	-0.845	-1.380	-0.019
	Med	0.551	0.273	0.870	0.127	-0.411	-0.930	-1.392	-0.042
	SD	0.114	0.062	0.251	0.035	0.258	0.340	0.244	0.277
	MSE	0.053	0.006	0.364	0.017	0.525	0.203	0.886	0.578
Pr4	**Mean**	0.471	0.198	0.807	0.399	-1.082	-0.668	-0.654	-0.881
	Med	0.497	0.204	0.827	0.359	-1.022	-0.674	-0.686	-0.917
	SD	0.097	0.252	0.337	0.493	0.509	0.167	0.312	0.233
	MSE	0.498	0.098	0.214	0.816	0.659	0.150	0.263	0.188
Pr5	**Mean**	0.536	0.178	0.442	0.123	-0.214	0.059	0.157	0.230
	Med	0.519	0.176	0.440	0.123	-0.223	0.042	0.156	0.233
	SD	0.094	0.030	0.034	0.010	0.219	0.226	0.120	0.141
	MSE	0.055	0.013	0.106	0.016	0.385	1.174	0.751	1.229

Table 5.60: Simulation results of imputing all unknown states with 2 covariates - Q_1Sc_4.

Q1	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
Sc4	**Fill**	**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Pr1	**Mean**	0.921	0.672	0.611	0.837	-0.312	-0.386	-0.224	-0.282	-0.003	-0.002	-0.008	0.000
	Med	0.880	0.646	0.585	0.798	-0.307	-0.381	-0.222	-0.281	-0.003	-0.003	-0.007	0.001
	SD	0.258	0.198	0.173	0.237	0.269	0.277	0.266	0.259	0.026	0.026	0.026	0.026
	MSE	0.254	0.087	0.043	0.172	0.230	0.176	0.297	0.242	0.001	0.001	0.001	0.001
Pr2	**Mean**	0.292	0.179	0.153	0.278	0.344	0.136	0.341	0.316	-0.001	0.003	0.006	0.005
	Med	0.272	0.172	0.144	0.264	0.354	0.132	0.352	0.318	0.000	0.003	0.006	0.005
	SD	0.118	0.057	0.056	0.095	0.344	0.315	0.313	0.335	0.039	0.033	0.034	0.035
	MSE	0.062	0.108	0.124	0.059	1.253	0.956	1.222	1.194	0.002	0.001	0.001	0.001
Pr3	**Mean**	2.059	1.057	1.292	2.140	-0.302	-0.144	-0.398	-0.375	-0.009	-0.005	-0.006	-0.006
	Med	1.884	0.970	1.196	1.975	-0.314	-0.142	-0.417	-0.388	-0.007	-0.004	-0.004	-0.005
	SD	1.250	0.487	0.596	1.016	0.267	0.284	0.260	0.240	0.028	0.029	0.032	0.030
	MSE	4.823	0.623	1.227	4.196	0.241	0.404	0.201	0.196	0.001	0.001	0.001	0.001
Pr4	**Mean**	1.498	0.819	0.919	0.682	-0.468	-0.503	-0.246	-0.238	-0.030	-0.023	0.021	0.003
	Med	1.435	0.794	0.893	0.660	-0.469	-0.513	-0.251	-0.247	-0.030	-0.023	0.021	0.003
	SD	0.441	0.223	0.201	0.163	0.259	0.257	0.249	0.261	0.024	0.025	0.021	0.025
	MSE	1.352	0.179	0.217	0.064	0.131	0.108	0.290	0.289	0.002	0.002	0.001	0.001
Pr5	**Mean**	0.966	1.256	1.026	1.796	-0.450	-0.334	-0.675	-0.593	-0.021	-0.005	-0.024	-0.030
	Med	0.834	1.154	0.878	1.541	-0.442	-0.324	-0.673	-0.590	-0.019	-0.004	-0.022	-0.028
	SD	0.537	0.546	0.613	0.980	0.355	0.344	0.351	0.362	0.034	0.029	0.035	0.032
	MSE	0.603	0.918	1.114	3.565	0.238	0.266	0.183	0.281	0.002	0.001	0.003	0.003

Table 5.61: Simulation results of imputing the transition point with 2 covariates - Q_1Sc_4.

Q1	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
Sc4	**TP**	**0.5**	**0.5**	**0.5**	**0.5**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Pr1	**Mean**	0.903	1.013	0.251	0.247	-0.633	-0.606	0.009	-0.164	-0.007	-0.021	0.030	0.055
	Med	0.887	0.989	0.245	0.243	-0.637	-0.611	0.018	-0.154	-0.007	-0.021	0.031	0.055
	SD	0.232	0.257	0.044	0.045	0.221	0.204	0.189	0.199	0.023	0.022	0.019	0.020
	MSE	0.453	0.834	0.073	0.073	0.150	0.095	0.562	0.330	0.004	0.009	0.006	0.006
Pr2	**Mean**	0.494	0.379	0.317	0.483	-0.149	-0.364	-0.117	-0.245	-0.003	0.009	0.024	0.030
	Med	0.454	0.353	0.287	0.434	-0.132	-0.341	-0.089	-0.211	-0.003	0.009	0.024	0.030
	SD	0.196	0.125	0.143	0.220	0.321	0.303	0.314	0.335	0.039	0.034	0.035	0.034
	MSE	0.046	0.036	0.084	0.143	0.441	0.440	0.460	0.376	0.002	0.001	0.003	0.005
Pr3	**Mean**	0.245	0.281	0.791	0.697	-0.169	0.324	-0.056	-0.231	0.056	0.025	-0.046	-0.019
	Med	0.231	0.266	0.741	0.651	-0.158	0.347	-0.042	-0.222	0.055	0.024	-0.049	-0.021
	SD	0.084	0.089	0.305	0.250	0.252	0.244	0.299	0.284	0.031	0.031	0.037	0.034
	MSE	0.079	0.067	0.189	0.211	0.347	1.152	0.805	0.435	0.003	0.001	0.005	0.007
Pr4	**Mean**	1.159	0.912	0.205	0.262	-0.101	-0.173	-0.343	-0.265	-0.072	-0.059	0.075	0.036
	Med	1.134	0.897	0.203	0.259	-0.097	-0.165	-0.342	-0.267	-0.073	-0.060	0.075	0.036
	SD	0.281	0.194	0.024	0.035	0.208	0.193	0.198	0.191	0.022	0.020	0.013	0.015
	MSE	0.861	0.341	0.097	0.059	0.509	0.361	0.271	0.319	0.011	0.007	0.008	0.002
Pr5	**Mean**	0.369	0.282	0.406	0.546	-0.095	0.182	-0.558	-0.479	0.004	0.027	0.008	0.001
	Med	0.351	0.273	0.395	0.535	-0.092	0.188	-0.551	-0.475	0.005	0.027	0.008	0.002
	SD	0.103	0.066	0.082	0.107	0.151	0.142	0.158	0.160	0.021	0.018	0.019	0.018
	MSE	0.054	0.067	0.018	0.018	0.405	0.801	0.048	0.108	0.002	0.001	0.001	0.000

Table 5.62: Simulation results of imputing all unknown states with 2 covariates - Q_2Sc_4.

Q2	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
Sc4	**Fill**	**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Pr1	**Mean**	0.864	0.920	0.382	0.957	-0.287	-0.185	-0.273	-0.340	-0.012	0.001	-0.007	-0.002
	Med	0.836	0.889	0.373	0.925	-0.286	-0.178	-0.289	-0.350	-0.012	0.001	-0.008	-0.001
	SD	0.214	0.222	0.115	0.258	0.210	0.206	0.278	0.258	0.024	0.024	0.032	0.030
	MSE	0.423	0.079	0.032	0.110	0.220	0.315	0.262	0.212	0.001	0.001	0.001	0.001
Pr2	**Mean**	0.221	0.278	0.112	0.732	0.123	0.102	0.400	-0.026	0.009	0.030	0.022	0.011
	Med	0.198	0.242	0.083	0.432	0.148	0.125	0.403	0.030	0.010	0.032	0.022	0.013
	SD	0.111	0.154	0.211	3.639	0.406	0.380	0.537	0.550	0.044	0.047	0.070	0.078
	MSE	0.017	0.255	0.095	24.94	0.909	0.903	1.621	0.921	0.002	0.003	0.007	0.010
Pr3	**Mean**	2.195	1.861	0.846	2.340	-0.461	-0.483	-0.614	-0.556	-0.001	-0.002	0.001	0.001
	Med	1.942	1.686	0.747	2.039	-0.470	-0.492	-0.620	-0.562	0.000	-0.004	0.002	0.003
	SD	2.760	2.293	0.603	1.831	0.340	0.340	0.407	0.380	0.035	0.035	0.043	0.041
	MSE	17.35	10.03	0.984	8.167	0.179	0.176	0.177	0.172	0.002	0.001	0.004	0.003
Pr4	**Mean**	1.291	1.596	0.613	0.904	-0.405	-0.538	-0.337	-0.396	-0.015	-0.044	0.014	-0.012
	Med	1.216	1.505	0.579	0.839	-0.402	-0.542	-0.344	-0.401	-0.014	-0.044	0.015	-0.012
	SD	0.466	0.593	0.215	0.335	0.303	0.297	0.330	0.308	0.029	0.030	0.029	0.031
	MSE	1.308	1.100	0.180	0.144	0.183	0.115	0.245	0.191	0.001	0.004	0.001	0.001
Pr5	**Mean**	0.604	1.758	0.243	1.167	-0.180	-0.207	-0.336	-0.295	0.000	0.000	0.020	-0.008
	Med	0.572	1.684	0.223	1.083	-0.177	-0.205	-0.350	-0.297	0.001	-0.001	0.018	-0.007
	SD	0.193	0.524	0.107	0.440	0.273	0.240	0.381	0.344	0.029	0.026	0.041	0.033
	MSE	0.163	1.319	0.014	0.435	0.347	0.306	0.279	0.286	0.001	0.001	0.002	0.001

Table 5.63: Simulation results of imputing the transition point with 2 covariates - Q_2Sc_4.

Q2	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
Sc4	**TP**	**0.25**	**0.75**	**0.25**	**0.75**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Pr1	**Mean**	0.894	1.899	0.131	0.733	-0.697	-0.341	0.080	-0.062	-0.086	-0.064	0.000	-0.031
	Med	0.776	1.687	0.127	0.717	-0.701	-0.346	0.106	-0.051	-0.086	-0.064	0.000	-0.031
	SD	0.993	1.792	0.033	0.155	0.239	0.211	0.332	0.284	0.031	0.027	0.025	0.026
	MSE	2.487	8.366	0.016	0.033	0.081	0.193	1.196	0.494	0.012	0.010	0.001	0.003
Pr2	**Mean**	0.329	0.529	0.591	4.903	-0.284	-0.185	-0.265	-0.858	0.021	0.060	-0.015	-0.077
	Med	0.198	0.363	0.231	1.438	-0.268	-0.177	-0.239	-0.847	0.019	0.060	-0.009	-0.065
	SD	0.531	0.598	2.1	24.2	0.459	0.401	0.530	0.454	0.068	0.065	0.080	0.088
	MSE	0.583	0.726	8.6	1182	0.468	0.555	0.535	0.283	0.008	0.009	0.010	0.020
Pr3	**Mean**	0.202	0.649	0.308	0.797	-0.378	-0.196	0.520	0.248	0.015	0.004	-0.052	-0.018
	Med	0.199	0.641	0.298	0.783	-0.393	-0.202	0.540	0.249	0.015	0.004	-0.053	-0.019
	SD	0.031	0.095	0.069	0.135	0.170	0.151	0.293	0.285	0.018	0.016	0.025	0.022
	MSE	0.003	0.079	0.031	0.021	0.213	0.343	1.649	1.541	0.001	0.004	0.013	0.005
Pr4	**Mean**	0.279	0.866	0.105	0.442	-0.392	-0.362	-0.037	0.156	0.043	-0.015	0.002	-0.012
	Med	0.267	0.834	0.099	0.432	-0.378	-0.357	-0.014	0.173	0.042	-0.015	0.001	-0.013
	SD	0.079	0.218	0.036	0.090	0.247	0.222	0.287	0.211	0.026	0.024	0.032	0.021
	MSE	0.021	0.147	0.024	0.103	0.232	0.174	0.530	0.788	0.003	0.003	0.005	0.001
Pr5	**Mean**	0.215	0.561	0.209	1.111	-0.244	-0.283	0.024	0.007	0.028	0.050	0.054	-0.014
	Med	0.214	0.555	0.199	1.062	-0.248	-0.286	0.024	-0.001	0.028	0.051	0.054	-0.014
	SD	0.030	0.087	0.053	0.257	0.119	0.112	0.179	0.165	0.018	0.018	0.019	0.020
	MSE	0.010	0.045	0.017	0.330	0.244	0.203	0.568	0.528	0.005	0.002	0.005	0.001

Table 5.64: Simulation results of imputing all unknown states with 2 covariates - Q_3Sc_4.

Q3	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
Sc4	**Fill**	**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Pr1	**Mean**	0.995	0.436	0.876	0.779	-0.240	-0.254	-0.346	-0.246	-0.009	-0.011	0.010	-0.003
	Med	0.943	0.412	0.849	0.751	-0.241	-0.265	-0.341	-0.242	-0.009	-0.009	0.010	-0.003
	SD	0.309	0.148	0.212	0.196	0.281	0.313	0.225	0.233	0.030	0.035	0.024	0.025
	MSE	0.157	0.057	0.071	0.320	0.291	0.306	0.189	0.265	0.001	0.002	0.001	0.001
Pr2	**Mean**	0.366	0.072	0.359	0.204	-0.216	0.153	-0.137	-0.044	0.021	0.012	-0.008	0.000
	Med	0.321	0.067	0.352	0.199	-0.206	0.151	-0.130	-0.037	0.021	0.011	-0.009	0.000
	SD	0.183	0.030	0.070	0.049	0.300	0.277	0.151	0.186	0.048	0.042	0.020	0.026
	MSE	0.190	0.033	0.158	0.005	0.360	0.813	0.360	0.465	0.003	0.002	0.001	0.001
Pr3	**Mean**	2.922	0.973	1.757	2.051	-0.513	-0.481	-0.369	-0.331	-0.011	-0.009	-0.010	-0.007
	Med	2.383	0.823	1.602	1.878	-0.524	-0.489	-0.376	-0.342	-0.009	-0.008	-0.009	-0.006
	SD	2.392	0.674	0.783	0.902	0.365	0.391	0.307	0.312	0.040	0.041	0.035	0.033
	MSE	29.79	2.05	1.91	4.51	0.214	0.274	0.217	0.245	0.003	0.003	0.002	0.002
Pr4	**Mean**	1.246	0.347	1.562	0.585	-0.299	-0.209	-0.255	-0.253	-0.005	-0.014	0.005	-0.001
	Med	1.198	0.331	1.519	0.570	-0.304	-0.221	-0.254	-0.253	-0.005	-0.015	0.005	-0.002
	SD	0.384	0.110	0.324	0.141	0.235	0.285	0.172	0.205	0.032	0.036	0.023	0.027
	MSE	0.440	0.023	0.775	0.132	0.230	0.336	0.241	0.244	0.002	0.003	0.001	0.001
Pr5	**Mean**	0.765	0.782	1.176	1.261	-0.297	0.000	-0.582	-0.396	0.005	-0.014	0.004	0.000
	Med	0.703	0.718	1.039	1.114	-0.280	0.007	-0.598	-0.407	0.003	-0.016	0.004	0.000
	SD	0.305	0.328	0.610	0.685	0.341	0.349	0.312	0.317	0.036	0.034	0.037	0.037
	MSE	0.124	0.446	0.860	1.854	0.320	0.807	0.272	0.332	0.002	0.002	0.002	0.002

Table 5.65: Simulation results of imputing the transition point with 2 covariates - Q_3Sc_4.

Q3	Par	λ_{12}	λ_{21}	λ_{23}	λ_{32}	β_{12}^{Cat}	β_{21}^{Cat}	β_{23}^{Cat}	β_{32}^{Cat}	β_{12}^{Con}	β_{21}^{Con}	β_{23}^{Con}	β_{32}^{Con}
Sc4	**TP**	**0.75**	**0.25**	**0.75**	**0.25**	**-0.7**	**-0.7**	**-0.7**	**-0.7**	**0.01**	**0.01**	**0.01**	**0.01**
Pr1	**Mean**	0.548	0.162	0.582	0.227	0.358	0.479	-0.675	-0.525	-0.032	-0.069	0.093	0.052
	Med	0.531	0.154	0.575	0.222	0.366	0.489	-0.677	-0.525	-0.033	-0.069	0.093	0.051
	SD	0.114	0.048	0.127	0.054	0.229	0.280	0.178	0.216	0.024	0.040	0.020	0.023
	MSE	0.151	0.015	0.114	0.008	1.182	1.522	0.050	0.092	0.004	0.008	0.008	0.004
Pr2	**Mean**	0.322	0.078	0.981	0.247	-0.363	-0.163	-0.429	-0.298	0.107	0.069	-0.029	-0.011
	Med	0.315	0.077	0.959	0.239	-0.365	-0.160	-0.428	-0.296	0.106	0.069	-0.029	-0.012
	SD	0.079	0.017	0.212	0.055	0.213	0.189	0.138	0.157	0.032	0.024	0.017	0.018
	MSE	0.253	0.031	0.499	0.028	0.394	0.383	0.094	0.212	0.018	0.005	0.006	0.003
Pr3	**Mean**	0.852	0.372	0.805	0.256	-0.284	0.410	-0.156	-0.295	-0.063	-0.070	-0.007	0.006
	Med	0.831	0.496	0.744	0.246	-0.263	0.437	-0.140	-0.290	-0.069	-0.073	-0.006	0.006
	SD	0.209	0.812	0.275	0.070	0.341	0.400	0.227	0.240	0.068	0.069	0.026	0.027
	MSE	0.304	0.713	0.784	0.012	0.459	1.877	0.548	0.350	0.036	0.023	0.004	0.002
Pr4	**Mean**	0.998	0.281	0.716	0.290	-0.536	-0.190	-0.408	-0.684	0.033	0.041	0.005	-0.028
	Med	0.917	0.255	0.692	0.279	-0.498	-0.153	-0.416	-0.680	0.032	0.041	0.005	-0.029
	SD	0.395	0.111	0.162	0.075	0.313	0.348	0.196	0.227	0.035	0.036	0.022	0.025
	MSE	0.374	0.022	0.028	0.012	0.174	0.455	0.218	0.057	0.002	0.003	0.002	0.006
Pr5	**Mean**	1.171	0.323	0.530	0.218	-0.342	0.233	-0.640	-0.481	-0.058	-0.063	0.044	0.060
	Med	1.009	0.282	0.502	0.203	-0.336	0.253	-0.612	-0.449	-0.059	-0.065	0.045	0.060
	SD	0.657	0.172	0.154	0.075	0.312	0.329	0.252	0.282	0.047	0.045	0.030	0.032
	MSE	0.716	0.048	0.074	0.007	0.314	1.187	0.219	0.345	0.007	0.009	0.002	0.004

Table 5.66: Bias for a select number of imputation models with covariates present.

Imputation	**Model**	**Scenario**	**Prior**	$\boldsymbol{\lambda_{12}}$	$\boldsymbol{\lambda_{21}}$	$\boldsymbol{\lambda_{23}}$	$\boldsymbol{\lambda_{32}}$	$\boldsymbol{\beta}_{12}^{Cat}$	$\boldsymbol{\beta}_{21}^{Cat}$	$\boldsymbol{\beta}_{23}^{Cat}$	β_{32}^{Cat}
Fill	Q1	2	1	0.311	0.093	0.131	0.238	0.640	0.770	0.522	0.703
			2	0.158	0.318	0.291	0.239	0.485	0.688	0.673	0.737
			3	1.489	0.675	0.593	1.648	0.349	0.337	0.529	0.353
			4	0.450	0.110	0.613	0.173	0.507	0.338	0.732	0.483
			5	0.154	0.587	0.055	0.438	0.576	0.687	0.439	0.574
TP	Q1	2	1	0.214	0.173	0.092	0.175	0.638	0.922	0.629	1.086
			2	0.088	0.195	0.211	0.216	0.541	0.742	0.715	0.740
			3	0.088	0.183	0.210	0.115	0.302	0.399	0.789	0.151
			4	0.271	0.229	0.264	0.215	0.220	0.170	1.343	0.583
			5	0.061	0.203	0.229	0.209	0.289	0.243	1.106	1.440
Fill	Q3	2	1	0.133	0.076	0.150	0.485	0.634	0.862	0.478	0.542
			2	0.331	0.180	0.412	0.038	0.691	0.773	0.503	0.635
			3	17.284	18.741	2.954	2.737	0.395	0.577	0.191	0.445
			4	0.761	0.054	1.207	0.418	0.468	0.805	0.508	0.536
			5	0.037	0.304	0.414	0.766	0.399	0.761	0.333	0.472
TP	Q3	2	1	0.258	0.130	0.439	0.217	0.491	1.472	0.346	0.391
			2	0.305	0.146	0.189	0.075	0.944	0.381	0.633	0.422
			3	0.200	0.047	0.548	0.127	0.677	0.295	0.909	0.708
			4	0.699	0.186	0.357	0.757	0.632	0.349	0.407	0.366
			5	0.215	0.110	0.324	0.127	0.580	1.060	0.858	1.100

5.5 Illustrative Examples

In this section, the two techniques introduced in this chapter are used to model two real world data sets. The first data set, from by Sharples *et al.* (2003) and Jackson (2011), is a study of the progression of coronary allograft vasculopathy (CAV), a post-transplant deterioration of the arterial walls. The second data set, from Andersen *et al.* (1993, pp. 32-34) and de Wreede *et al.* (2010), is a study describing patients with liver cirrhosis.

Each example will take the following form:

1) Some background on the data is given and an extract of the data is provided.
2) The multi-state model that will be fitted to the data is presented.
3) The priors that will be used when imputing all missing observations is given (the Fill models).
4) The priors that will be used when imputing the transition point is given (the TP models).
5) Posterior model assessment statistics for the models without any covariates included are presented and interpreted. These statistics are based on repeating the modelling processes of Sections 5.1.1 and 5.2.1 3000 times for each prior, resulting in posterior distributions consisting of 3000 values for each parameter in the model.
6) Based on the results from 5, the best model is refitted (again using 3000 repetitions) now with covariates included and the posterior parameter estimates and summary statistics are presented and interpreted.

The aim of this section is not to assess the appropriateness of the multi-state models for the given data sets, but rather to repeat the analyses performed in the published articles and compare it to the Bayesian models proposed here.

5.5.1 Coronary Allograft Vasculopathy (CAV)

See Section 4.4.1 for a description of this data set.

The following prior assumption will be used for the Bayesian multi-state imputation techniques:

– Imputing all missing observations:

$$\text{Prior}_1^{Fill} \begin{bmatrix} 0.20 & 0.40 & 0 & 0.40 \\ 0.2\dot{6} & 0.20 & 0.2\dot{6} & 0.2\dot{6} \\ 0 & 0.40 & 0.20 & 0.40 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0.10 & 0.45 & 0 & 0.45 \\ 0.30 & 0.10 & 0.30 & 0.30 \\ 0 & 0.45 & 0.10 & 0.45 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Prior_2^{Fill} $\begin{bmatrix} 0.50 & 0.25 & 0 & 0.25 \\ 0.16\dot{6} & 0.50 & 0.16\dot{6} & 0.16\dot{6} \\ 0 & 0.25 & 0.50 & 0.25 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.55 & 0.225 & 0 & 0.225 \\ 0.15 & 0.55 & 0.15 & 0.15 \\ 0 & 0.225 & 0.55 & 0.225 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Prior_3^{Fill} $\begin{bmatrix} 0.80 & 0.10 & 0 & 0.10 \\ 0.06\dot{6} & 0.80 & 0.06\dot{6} & 0.06\dot{6} \\ 0 & 0.10 & 0.80 & 0.10 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.90 & 0.05 & 0 & 0.05 \\ 0.03\dot{3} & 0.90 & 0.03\dot{3} & 0.03\dot{3} \\ 0 & 0.05 & 0.90 & 0.05 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Prior_4^{Fill} $\begin{bmatrix} 0.20 & 0.40 & 0 & 0.40 \\ 0.10 & 0.20 & 0.35 & 0.35 \\ 0 & 0.10 & 0.20 & 0.70 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.30 & 0.35 & 0 & 0.35 \\ 0.15 & 0.30 & 0.275 & 0.275 \\ 0 & 0.15 & 0.30 & 0.55 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Prior_5^{Fill} $\begin{bmatrix} 0.80 & 0.10 & 0 & 0.10 \\ 0.70 & 0.20 & 0.05 & 0.05 \\ 0 & 0.70 & 0.20 & 0.10 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.70 & 0.15 & 0 & 0.15 \\ 0.40 & 0.30 & 0.15 & 0.15 \\ 0 & 0.55 & 0.30 & 0.15 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

– Imputing the transition point:

Prior_1^{TP} All transition functions are assumed to be T_B-functions. Under this assumption it is assumed that the transition to the following state occured at the beginning of the current observation interval, which corresponds to Prior_1^{Fill}.

Prior_2^{TP} All transition functions are assumed to be T_F-functions. Under this assumption it is assumed that the transition to the following state occurred at any point during the current observation interval, which corresponds to Prior_2^{Fill}.

Prior_3^{TP} All transition functions are assumed to be T_E-functions. Under this assumption it is assumed that the transition to the following state occurred near the end of the current observation interval, which corresponds to Prior_3^{Fill}.

Prior_4^{TP} Transition functions to higher states (T_{12}, T_{14}, T_{23}, T_{24} and T_{34}) are assumed to be T_B-functions and transitions to lower states (T_{21} and T_{32}) are assumed to be T_E-functions. Under these assumptions it is assumed that patients make transitions quickly from a lower to a higher state and once in a higher state take longer to make a transition to a lower state, which corresponds to Prior_4^{Fill}.

Prior_5^{TP} Transition functions to higher states (T_{12}, T_{14}, T_{23}, T_{24} and T_{34}) are assumed to be T_E-functions and transitions to lower states (T_{21} and T_{32}) are assumed to be T_B-functions. Under these assumptions it is assumed that patients make transitions quickly from a higher to a lower state and once in a lower state take longer to make

a transition to a higher state, which corresponds to Prior_5^{Fill}.

The DIC and goodness-of-fit (GOF) values for each one of 5 priors across the two techniques without any covariates included in the model are given in Tables 5.67 and 5.68. The GOF values presented here can be compared to the frequentist model value of 165.05 ($p < 0.0001$) (Jackson, 2011). All GOF-values in Table 5.68 have p-values that have an upper limit that is smaller than 0.0001, based on an upper limit of the degrees of freedom of 81 (see Section 2.2.2).

These results indicate that the best fitting Bayesian model is the model where the transition point is imputed between two known observations and prior 3 is used as the prior for the transition functions. It must however be noted that, similar to the results found by Sharples *et al.* (2003) and Jackson (2011), the p-values for the GOF statistic are extremely small, indicating a lack of fit for all these models. Jackson (2011) attributed this to the fact that the underlying multi-state model may not be strictly medically realistic and that a more complex pattern of time-dependent, or allowing the transition intensities to depend on covariates, would be expected to yield a better fit.

Table 5.67: DIC values for the 5 priors and 2 imputing techniques for the CAV data.

DIC	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**
Fill	4267.18	3737.13	3498.84	4071.49	3644.45
TP	3392.47	3378.06	**3371.42**	3392.45	3383.93

Table 5.68: Goodness-of-fit values for the 5 priors and 2 imputing techniques for the CAV data.

GOF	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**
Fill	699.07	198.84	164.75	314.56	148.01
TP	167.32	147.66	**138.08**	167.99	139.18

To improve the fit of the model, the "TP model" with prior 3 is refitted to the data together with the two covariates in the multi-state model. The GOF for this model is 333.64 (the frequentist model value is 299.95), with a p-value < 0.0001 (based on an upper limit of the degrees of freedom of 241), and the DIC is found to be 3991.87. This result indicates that even with the two covariates included in the model, the model still does not fit the data well.

The summary statistics of the posterior distributions (based on 3000 posterior variates) for the 21 parameters in the model are presented in Table 5.69. The posterior and frequentist

hazard ratios, posterior prediction matrices and posterior survival curves for the fitted model are presented in Tables 5.70 to 5.72 and Figures 5.12 and 5.13.

The hazard rates indicate that a diagnosis of IHD is associated with a 45% increase in the hazard of CAV onset and a 68% decrease in the risk of moving back to the mild state once reaching the severe state. Every 1 year increase in the donor age is associated with a 2% increase in risk of CAV, a 2% decrease in the risk of death once reaching the mild state, a 4% increase in the risk of moving back to the mild state once reaching the severe state and a 4% decrease in the risk of death once reaching the severe state.

The prediction matrices give the probability of being in the four different stages after one year in the study for patients with and without IHD being the primary diagnosis (both assuming the donor age was 28, the mean value in the data set). These are calculated as the mean values of the posterior predictive distributions that are presented in Figures 5.14 to 5.19. The survival curves show the survival probabilities for a patient starting in each one of the three none death states at the beginning of the study. Both the prediction matrices and the survival curves show that if IHD is the primary diagnosis, that patient will have a lower survival time and will progress more quickly to the severe state of CAV.

As the GOF-values for all models (Bayesian and frequentist) indicate a lack of fit for the models, it is important not to over interpret the result. It is however useful to highlight the differences found by the two modelling techniques. These differences are most notable when looking at the hazard rates of the covariates in the model. Based on the Bayesian model, IHD increases the risk of CAV onset and decreases the risk of moving to moderate from severe CAV, while for the frequentist model IHD only increases the risk of CAV onset. Under the Bayesian model the donor age increases the risk of CAV onset and moving from severe to moderate CAV, and decreases the risk of moving from moderate to death and from severe to death. Under the frequentist model donor age increases the risk of CAV onset as well as the risk of death without CAV.

Table 5.69: Posterior summary for CAV data with covariates - TP (Pr_3).

TP (Prior 3)	**Mean**	**Med**	**SD**	**Cred$_L$**	**Cred$_U$**	**Cred$_{\bar{x}}$**	**HPD$_L$**	**HPD$_U$**	**HPD$_{\bar{x}}$**
λ_{WM}	0.1527	0.1531	0.0042	0.1437	0.1605	0.0167	0.1440	0.1605	0.0164
λ_{WD}	0.0058	0.0051	0.0028	0.0027	0.0128	0.0101	0.0023	0.0113	0.0090
λ_{MW}	0.2575	0.2566	0.0147	0.2305	0.2876	0.0570	0.2313	0.2879	0.0566
λ_{MS}	0.2524	0.2517	0.0124	0.2290	0.2802	0.0511	0.2281	0.2784	0.0500
λ_{MD}	0.2199	0.2224	0.0172	0.1818	0.2472	0.0654	0.1870	0.2503	0.0630
λ_{SM}	0.1284	0.1289	0.0154	0.0984	0.1588	0.0603	0.0973	0.1576	0.0603
λ_{SD}	0.2516	0.2527	0.0207	0.2105	0.2889	0.0783	0.2120	0.2899	0.0777
β_{WM}^{IHD}	0.3734	0.3740	0.0512	0.2693	0.4758	0.2065	0.2675	0.4735	0.2060
β_{WS}^{IHD}	-0.4430	-0.4220	0.4013	-1.3217	0.2942	1.6160	-1.2401	0.3224	1.5625
β_{MW}^{IHD}	-0.0500	-0.0512	0.0897	-0.2228	0.1349	0.3577	-0.2377	0.1181	0.3559
β_{MS}^{IHD}	0.0372	0.0385	0.1156	-0.1868	0.2663	0.4532	-0.2050	0.2453	0.450
β_{MD}^{IHD}	-0.0413	-0.0597	0.1288	-0.2505	0.2661	0.5166	-0.2771	0.2184	0.4955
β_{SM}^{IHD}	-1.1454	-1.1392	0.2466	-1.6538	-0.6729	0.9809	-1.7050	-0.7361	0.9697
β_{SD}^{IHD}	0.0642	0.0535	0.1503	-0.2009	0.3760	0.5772	-0.1969	0.3792	0.5761
β_{WM}^{DAGE}	0.0169	0.0169	0.0023	0.0125	0.0215	0.0090	0.0122	0.0210	0.0088
β_{WD}^{DAGE}	0.0186	0.0274	0.0432	-0.0750	0.0814	0.1564	-0.0650	0.0859	0.1510
β_{MW}^{DAGE}	-0.0082	-0.0081	0.0050	-0.018	0.0015	0.0202	-0.0184	0.0017	0.0201
β_{MS}^{DAGE}	-0.0099	-0.0101	0.0050	-0.0203	0.0004	0.0207	-0.0191	0.0008	0.0200
β_{MD}^{DAGE}	-0.0197	-0.0190	0.0073	-0.0345	-0.0065	0.0279	-0.0346	-0.0068	0.0278
β_{SM}^{DAGE}	0.0352	0.0340	0.0140	0.0105	0.0628	0.0523	0.0102	0.0619	0.0516
β_{SD}^{DAGE}	-0.0373	-0.0372	0.0074	-0.0530	-0.0233	0.0296	-0.0514	-0.0224	0.0289

Table 5.70: Posterior and Frequentist models hazard ratios (95% HPD and 95% CI) for CAV data.

	HR^{TP}	**HPD$^L_{95\%}$**	**HPD$^U_{95\%}$**	**HR$^{Freq}_{95\%}$**	**LL$^{Freq}_{95\%}$**	**UL$^{Freq}_{95\%}$**
β_{WM}^{IHD}	**1.453**	**1.307**	**1.606**	**1.565**	**1.179**	**2.076**
β_{WD}^{IHD}	0.642	0.289	1.381	1.304	0.821	2.073
β_{MW}^{IHD}	0.951	0.788	1.125	0.937	0.519	1.691
β_{MS}^{IHD}	1.038	0.815	1.278	0.958	0.613	1.498
β_{MD}^{IHD}	0.959	0.758	1.244	1.786	0.230	13.873
β_{SM}^{IHD}	**0.318**	**0.182**	**0.479**	0.767	0.271	2.173
β_{SD}^{IHD}	1.066	0.821	1.461	0.757	0.456	1.257
β_{WM}^{DAGE}	**1.017**	**1.012**	**1.021**	**1.019**	**1.007**	**1.032**
β_{WD}^{DAGE}	1.019	0.937	1.090	**1.038**	**1.018**	**1.059**
β_{MW}^{DAGE}	0.992	0.982	1.002	0.998	0.973	1.024
β_{MS}^{DAGE}	0.990	0.981	1.001	0.985	0.967	1.004
β_{MD}^{DAGE}	**0.980**	**0.966**	**0.993**	0.932	0.845	1.028
β_{SM}^{DAGE}	**1.036**	**1.010**	**1.064**	0.998	0.948	1.050
β_{SD}^{DAGE}	**0.963**	**0.950**	**0.978**	0.988	0.965	1.013

Table 5.71: One year prediction matrix. Primary diagnosis - IHD; Donor age = 28 years.

IHD (DAGE = 28)	Well	Mild	Severe	Death
Well	0.845	0.118	0.016	0.021
Mild	0.161	0.500	0.154	0.185
Severe	0.006	0.043	0.721	0.230

Table 5.72: One year prediction matrix. Primary diagnosis - non-IHD; Donor age = 28 years.

Non-IHD (DAGE = 28)	Well	Mild	Severe	Death
Well	0.887	0.084	0.010	0.019
Mild	0.178	0.503	0.138	0.181
Severe	0.019	0.125	0.639	0.217

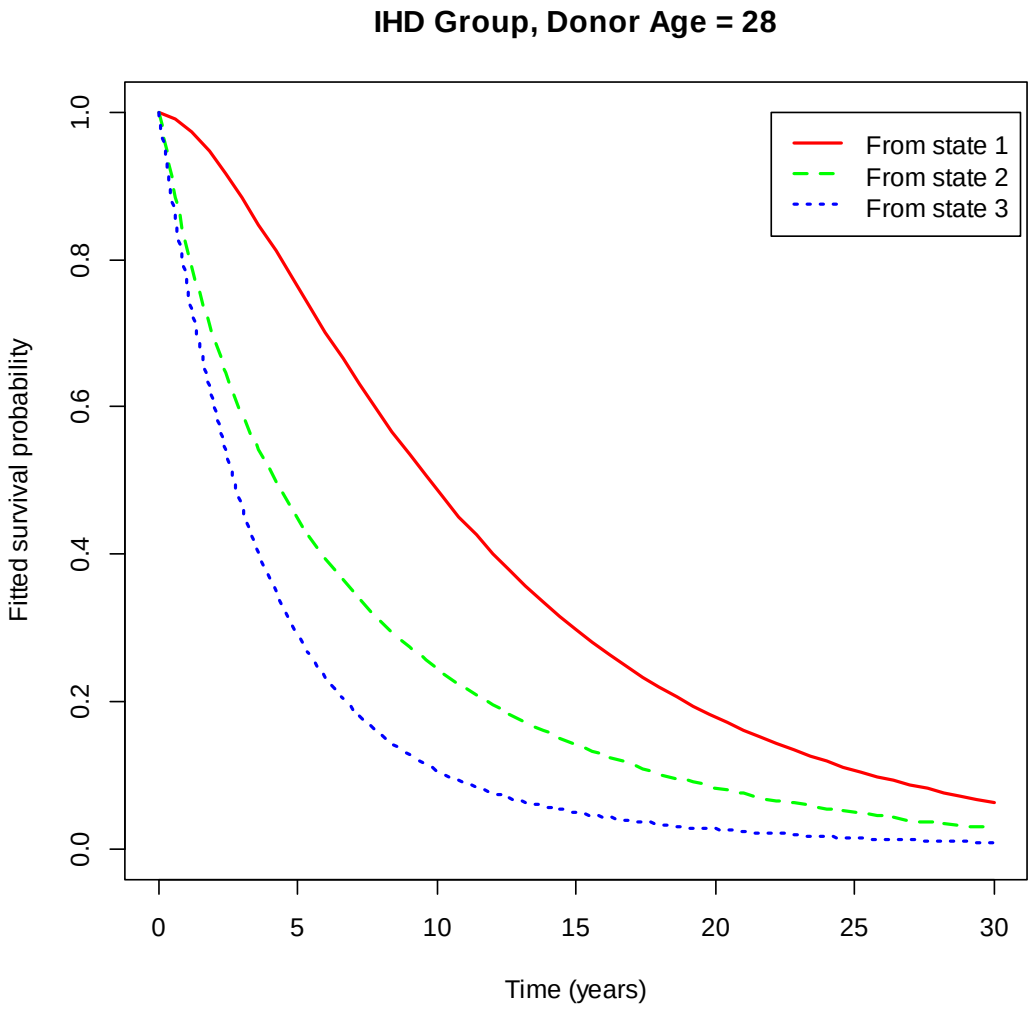

Figure 5.12: Posterior survival curves for CAV data (Primary diagnosis - IHD; Donor age = 28 years).

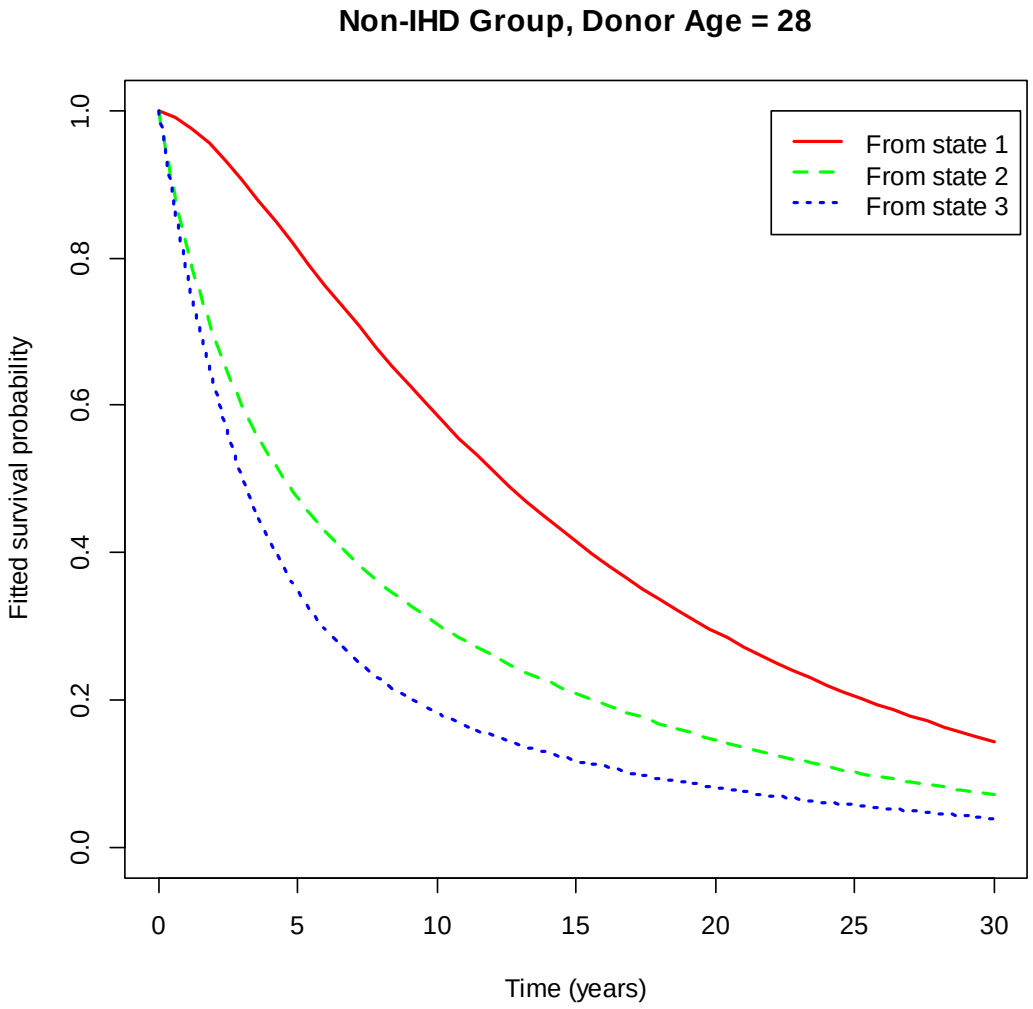

Figure 5.13: Posterior survival curves for CAV data (Primary diagnosis - Non-IHD; Donor age = 28 years).

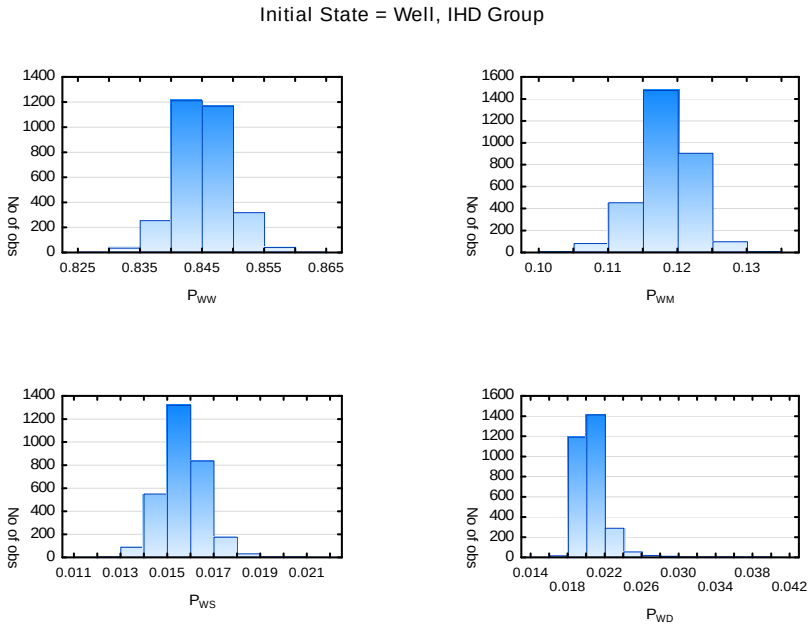

Figure 5.14: One year posterior predictive distributions for CAV data (Initial state = Well, IHD Group).

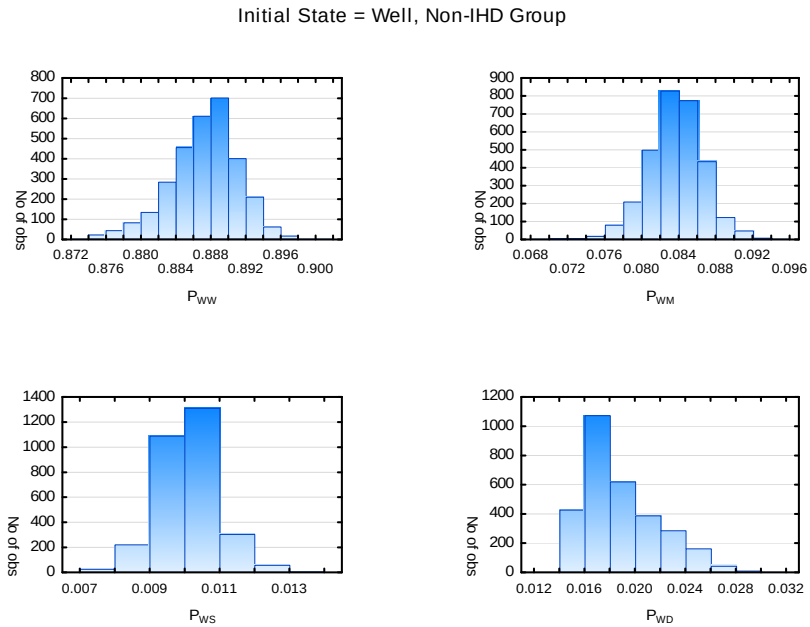

Figure 5.15: One year posterior predictive distributions for CAV data (Initial state = Well, Non-IHD Group).

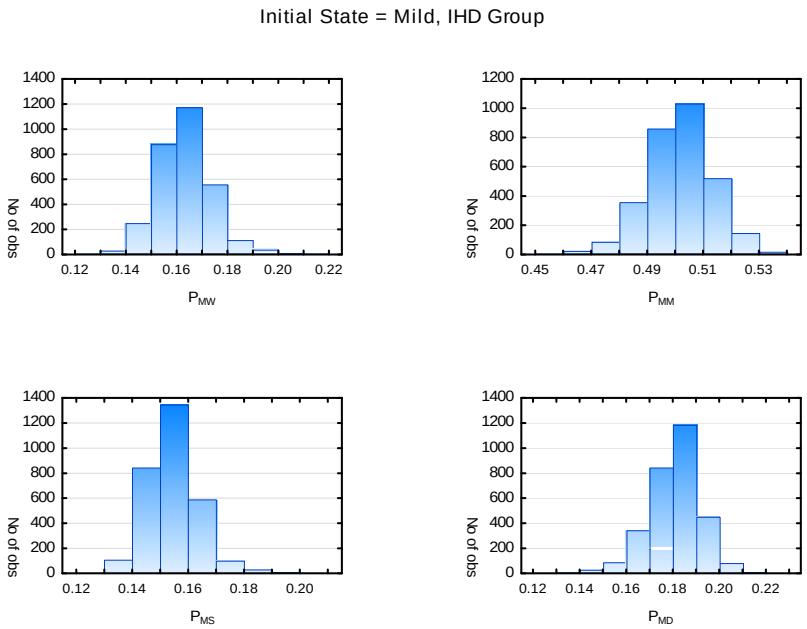

Figure 5.16: One year posterior predictive distributions for CAV data (Initial state = Mild, IHD Group).

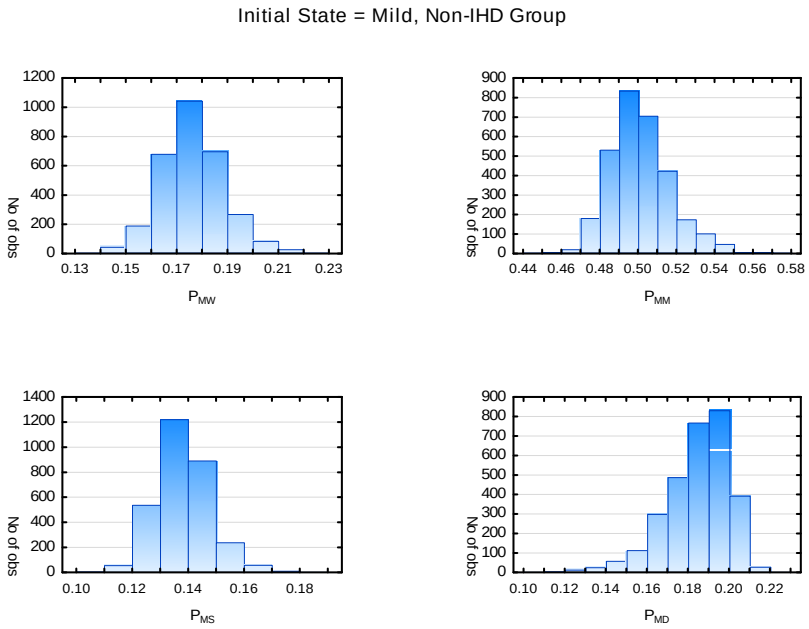

Figure 5.17: One year posterior predictive distributions for CAV data (Initial state = Mild, Non-IHD Group).

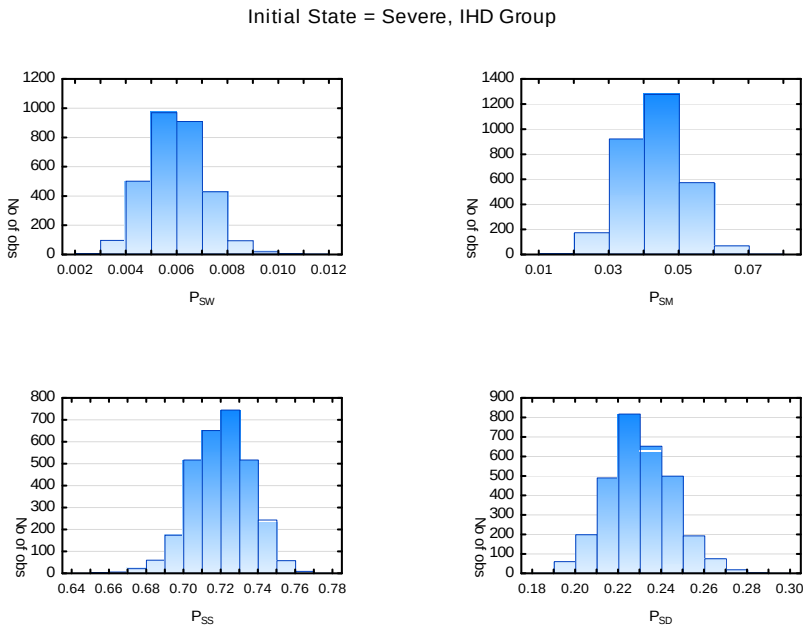

Figure 5.18: One year posterior predictive distributions for CAV data (Initial state = Severe, IHD Group).

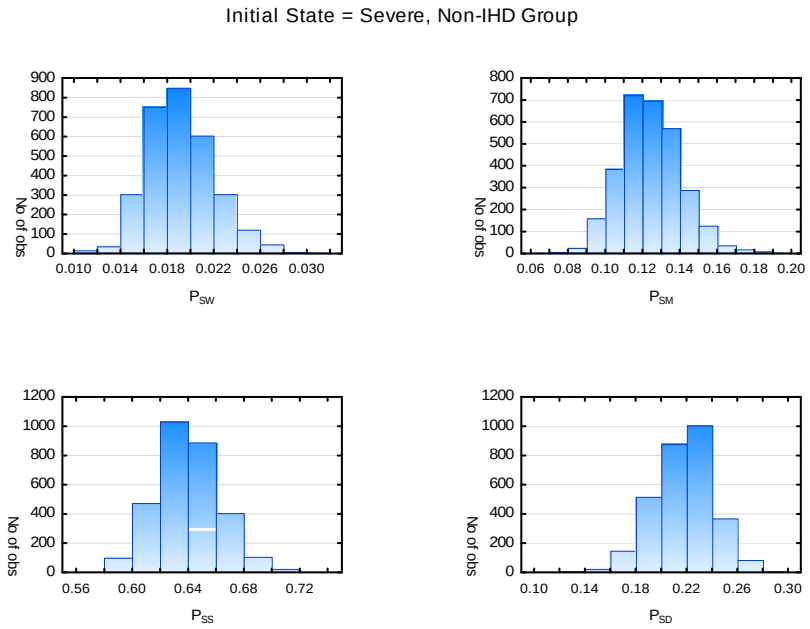

Figure 5.19: One year posterior predictive distributions for CAV data (Initial state = Severe, Non-IHD Group).

5.5.2 Liver cirrhosis (LC)

This data set is taken from Andersen *et al.* (1993, pp. 32-34) and de Wreede *et al.* (2010). The data describes patients with liver cirrhosis. They originate from a clinical trial performed in Copenhagen, in which patients were randomised to receive either the hormone prednisone or placebo. The primary goal was to investigate whether prednisone prolongs the survival of liver cirrhosis patients (de Wreede *et al.*, 2010). Here the prothrombin index, an indication of the functioning of the liver, is used to classify patients into two transient and one absorbing state(-s) for this study:

- 1, normal prothrombin index.
- 2, low prothrombin index.
- 3, death.

The data set consists of 251 patients who received prednisone and 237 who received the placebo. Of these 488 patients, only 265 had at least one follow-up visit before death; 141 receiving prednisone and 124 receiving the placebo. These 265 patients with at least one follow-up visit before death will be used in the analysis of this data set.

Table 5.73 contains an extract of the data for two patients.

Table 5.73: Extract from the LC data set.

Subject	**Time (days)**	**State**	**Prednisone**
⋮	⋮	⋮	⋮
2	0	2	0
2	251	1	0
2	434	2	0
2	729	1	0
⋮	⋮	⋮	⋮
560	0	1	1
560	78	2	1
560	431	1	1
560	1588	3	1
⋮	⋮	⋮	⋮

The underlying model assumed by the authors, and the same model that will be used here, is (de Wreede *et al.*, 2010):

$$Q = \begin{bmatrix} -(\lambda_{NL} + \lambda_{ND}) & \lambda_{NL} & \lambda_{ND} \\ \lambda_{LN} & -(\lambda_{LN} + \lambda_{LD}) & \lambda_{LD} \\ 0 & 0 & 0 \end{bmatrix}.$$

The following prior assumption will be used for the Bayesian multi-state imputation techniques:

– Imputing all missing observations:

Prior_1^{Fill} $\begin{bmatrix} 0.20 & 0.40 & 0.40 \\ 0.40 & 0.20 & 0.40 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.10 & 0.45 & 0.45 \\ 0.45 & 0.10 & 0.45 \\ 0 & 0 & 1 \end{bmatrix}$.

Prior_2^{Fill} $\begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0.25 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.55 & 0.225 & 0.225 \\ 0.225 & 0.55 & 0.225 \\ 0 & 0 & 1 \end{bmatrix}$.

Prior_3^{Fill} $\begin{bmatrix} 0.80 & 0.10 & 0.10 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0 & 0 & 1 \end{bmatrix}$.

Prior_4^{Fill} $\begin{bmatrix} 0.20 & 0.40 & 0.40 \\ 0.10 & 0.20 & 0.70 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.30 & 0.35 & 0.35 \\ 0.15 & 0.30 & 0.55 \\ 0 & 0 & 1 \end{bmatrix}$.

Prior_5^{Fill} $\begin{bmatrix} 0.50 & 0.33 & 0.17 \\ 0.36 & 0.50 & 0.14 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0.40 & 0.36 & 0.24 \\ 0.40 & 0.40 & 0.20 \\ 0 & 0 & 1 \end{bmatrix}$.

– Imputing the transition point:

Prior_1^{TP} All transition functions are assumed to be T_B-functions. Under this assumption it is assumed that the transition to the following state occurred at the beginning of the current observation interval, which corresponds to Prior_1^{Fill}.

Prior_2^{TP} All transition functions are assumed to be T_F-functions. Under this assumption it is assumed that the transition to the following state occurred at any point during the current observation interval, which corresponds to Prior_2^{Fill}.

Prior_3^{TP} All transition functions are assumed to be T_E-functions. Under this assumption it is assumed that the transition to the following state occurred near the end of the current observation interval, which corresponds to Prior_3^{Fill}.

Prior_4^{TP} Transition functions to higher states (T_{12}, T_{13} and T_{23}) are assumed to be T_B-functions and transitions to lower states (T_{21}) are assumed to be T_E-functions. Under these assumptions it is assumed that patients make ransitions quickly from a lower to a higher state and once in a higher state take longer to make a transition to a lower state, which corresponds to Prior_4^{Fill}.

Prior_5^{TP} Transition functions to higher states (T_{12}, T_{13} and T_{23}) are assumed to be T_E-functions and transitions to lower states (T_{21}) are assumed to be T_B-functions. Under these

assumptions it is assumed that patients make transitions quickly from a higher to a lower state and once in a lower state take longer to make a transition to a higher state, which corresponds to Prior_5^{Fill}.

The DIC and goodness-of-fit (GOF) values for each one of 5 priors across the two techniques without any covariates included in the model are given in Tables 5.74 and 5.75. The GOF values presented here can be compared to the frequentist model value of 665.52 ($p < 0.0001$). All GOF-values in Table 5.75 have p-values that have an upper limit that are smaller than 0.0001, based on an upper limit of the degrees of freedom of 36 (see Section 2.2.2).

These results indicate that the best fitting Bayesian model is the model where all missing observations are imputed using prior 5. The large GOF-values for the frequentist, as well as the Bayesian models show that none of these models really fit the data well.

Table 5.74: DIC values for the 5 priors and 2 imputing techniques for the LC data.

DIC	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**
Fill	5794.38	5455.41	5704.73	5995.57	**5361.67**
TP	5399.92	5382.39	5395.61	5430.54	5472.47

Table 5.75: Goodness-of-fit values for the 5 priors and 2 imputing techniques for the LC data.

GOF	**Prior 1**	**Prior 2**	**Prior 3**	**Prior 4**	**Prior 5**
Fill	891.22	897.11	755.74	757.69	**723.53**
TP	788.23	772.05	726.60	744.18	740.27

To improve the fit of the model, the "Fill model" with prior 5 is refitted to the data together with the covariate in the multi-state model. The GOF for this model is 753.83 (the frequentist model value is 694.05), with a p-value < 0.0001 (based on an upper limit of the degrees of freedom of 72), and the DIC is found to be 5469.92. This indicates that the goodness-of-fit has increased with the introduction of the covariate into the model, indicating a worse fitted model than the one without the covariate included.

The summary statistics for the posterior distributions based on the 3000 posterior variates are presented in Table 5.69. The posterior and frequentist hazard ratios, posterior prediction matrices and posterior survival curves for the fitted model are presented in Tables 5.77 to 5.79 and Figures 5.20 and 5.21.

The hazard rates indicate that receiving prednisone is associated with a 34% decrease in the hazard of death from the normal prothrombin group and a 29% increase in the risk of death

from the low prothrombin group.

The prediction matrices give the probability of being in the three different stages after two years in the study for the prednisone and placebo groups. These are calculated as the mean values of the posterior predictive distributions that are presented in Figures 5.22 to 5.25. The survival curves show the survival probabilities for a patient starting in each one of the two none death states at the beginning of the study. Both the prediction matrices and the survival curves show slight differences between the two treatment groups, the prednisone group is slightly less likely to progress to the death stage, but as was found by de Wreede *et al.* (2010) there is no clear difference between the two groups.

As the GOF-values for all models (Bayesian and frequentist) indicate a lack of fit for the models, it is important not to over interpret the results. It is however useful to highlight the differences found by the two modelling techniques. These differences are most notable when looking at the hazard rates of the covariate in the model. Based on the Bayesian model prednisone decreases the risk of death in the normal group and increases the risk of death in the low group. Under the frequentist model prednisone decreases the risk of moving from the normal to the low group and increases the risk of moving from the low to the normal group.

Table 5.76: Posterior summary for LC data with covariates - Fill (Pr_5).

Fill (Prior 5)	**Mean**	**Med**	**SD**	**Cred$_L$**	**Cred$_U$**	**Cred$_{\bar{x}}$**	**HPD$_L$**	**HPD$_U$**	**HPD$_{\bar{x}}$**
λ_{NL}	0.0373	0.0370	0.0026	0.0325	0.0433	0.0108	0.0318	0.0419	0.0101
λ_{ND}	0.0004	0.0004	0.0001	0.0003	0.0004	0.0001	0.0003	0.0004	0.0001
λ_{LN}	0.0591	0.0589	0.0039	0.0521	0.0685	0.0163	0.0510	0.0670	0.0159
λ_{LD}	0.0027	0.0027	0.0001	0.0026	0.0028	0.0002	0.0026	0.0028	0.0002
β_{NL}^{Treat}	-0.1065	-0.1083	0.1128	-0.3170	0.1334	0.4504	-0.3230	0.1144	0.4375
β_{ND}^{Treat}	-0.4133	-0.4246	0.1112	-0.6073	-0.1749	0.4324	-0.6150	-0.1892	0.4257
β_{LN}^{Treat}	0.0503	0.0525	0.1040	-0.1509	0.2641	0.4150	-0.1538	0.2641	0.4180
β_{LD}^{Treat}	0.2574	0.2596	0.0452	0.1563	0.3436	0.1873	0.1671	0.3490	0.1818

Table 5.77: Posterior and Frequentist models hazard ratios (95% HPD and 95% CI) for LC data.

	HR$^{\text{Fill}}$	**HPD$^{L}_{95\%}$**	**HPD$^{U}_{95\%}$**	**HR$^{\text{Freq}}_{95\%}$**	**LL$^{\text{Freq}}_{95\%}$**	**UL$^{\text{Freq}}_{95\%}$**
β_{NL}^{Treat}	0.899	0.724	1.121	**0.774**	**0.611**	**0.981**
β_{ND}^{Treat}	**0.661**	**0.541**	**0.828**	0.780	0.530	1.147
β_{LN}^{Treat}	1.052	0.857	1.302	**1.361**	**1.088**	**1.701**
β_{LD}^{Treat}	**1.294**	**1.182**	**1.418**	1.281	0.959	1.713

Table 5.78: Two year prediction matrix - Prednisone group.

Prednisone	Normal	Low	Death
Normal	0.248	0.165	0.587
Low	0.242	0.162	0.596

Table 5.79: Two year prediction matrix - Control group.

Control	Normal	Low	Death
Normal	0.248	0.141	0.611
Low	0.241	0.137	0.622

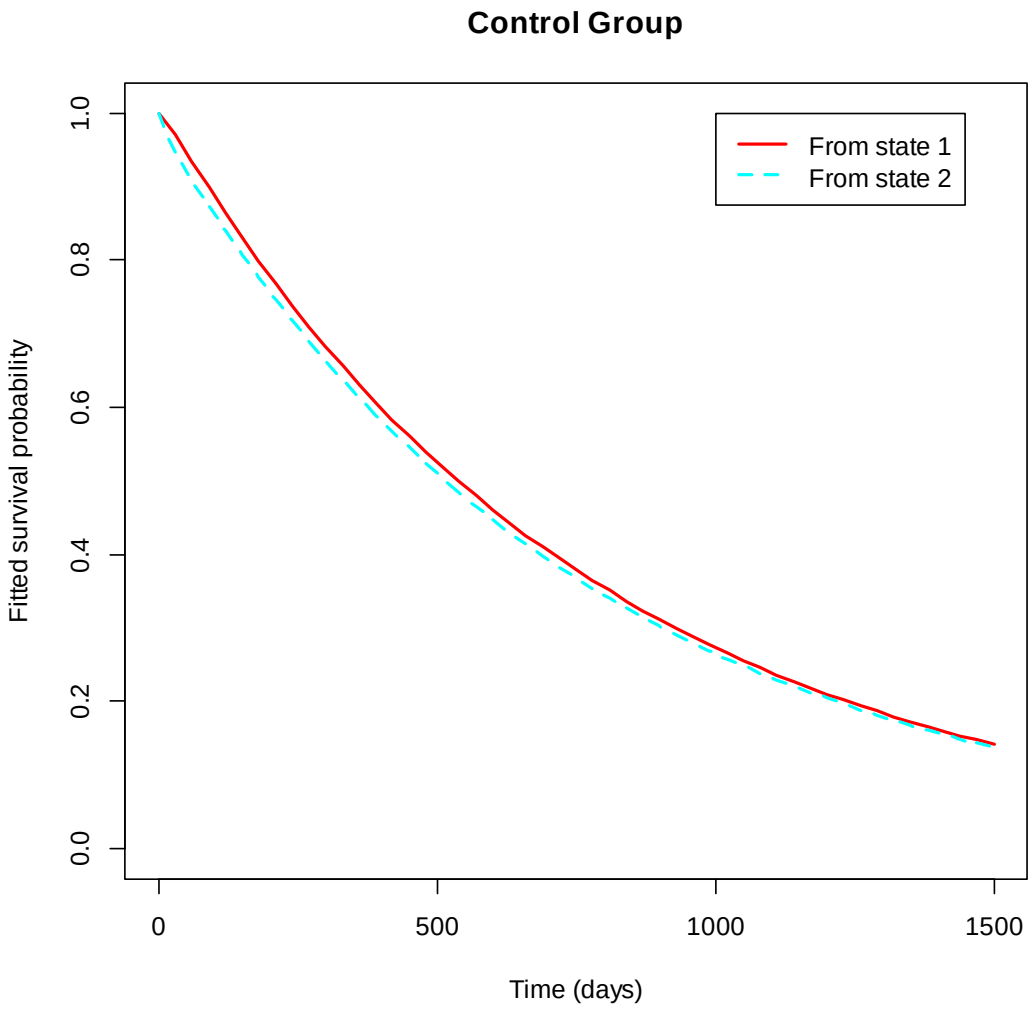

Figure 5.20: Posterior survival curves for LC data (Control group).

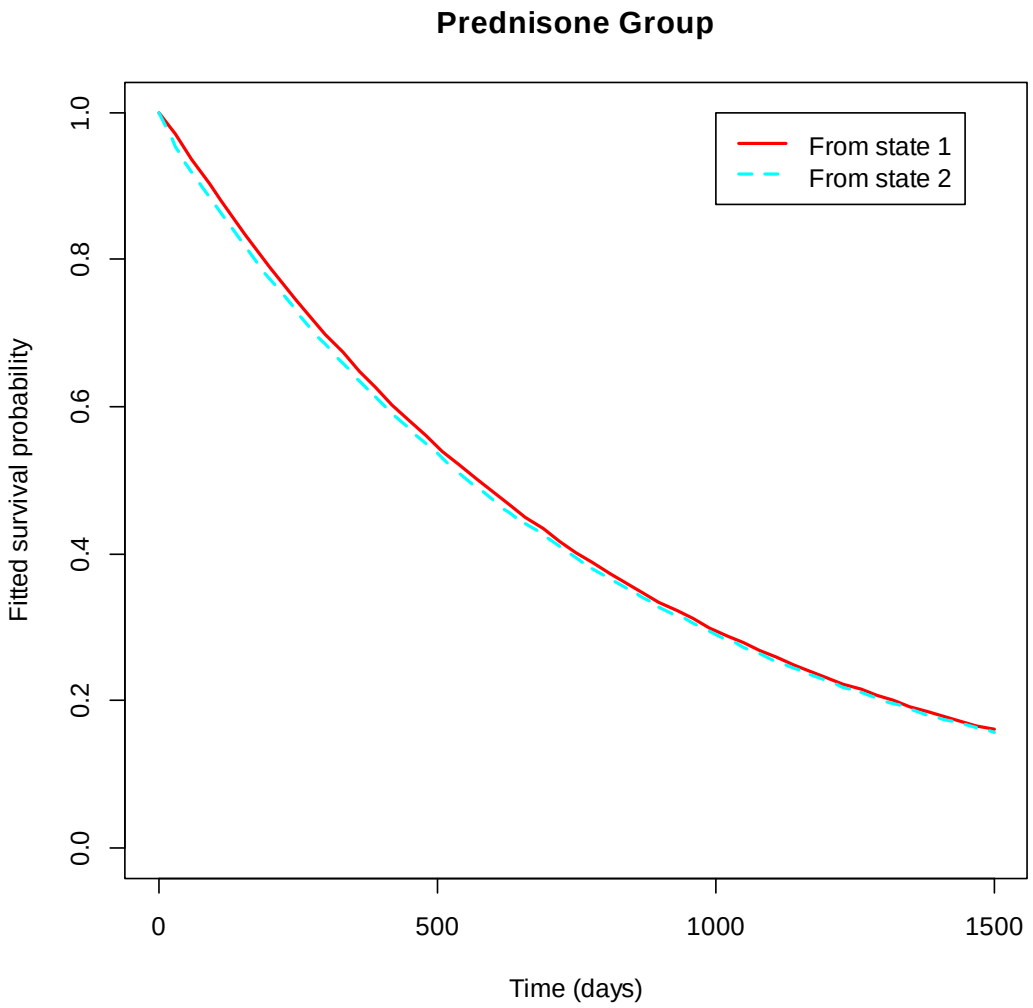

Figure 5.21: Posterior survival curves for LC data (Prednisone group).

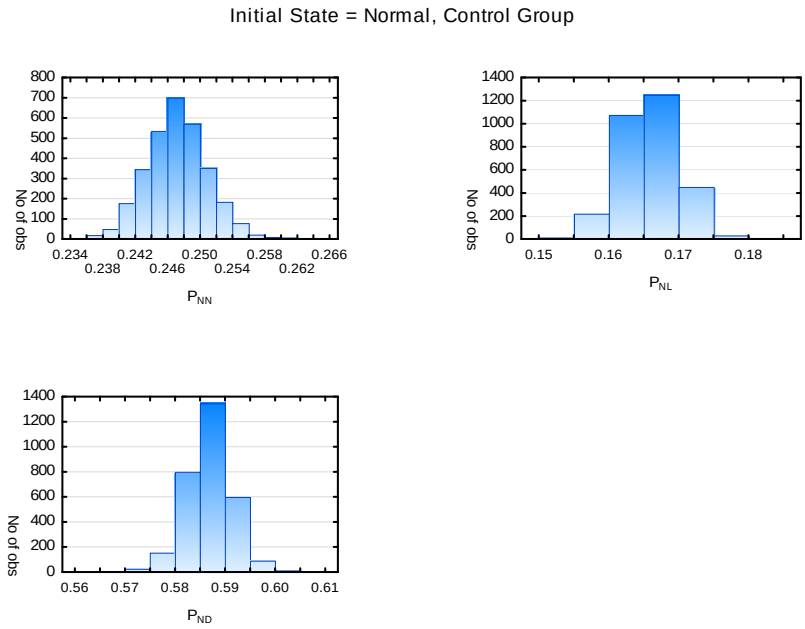

Figure 5.22: Two year posterior predictive distributions for LC data (Initial state = Normal, Control).

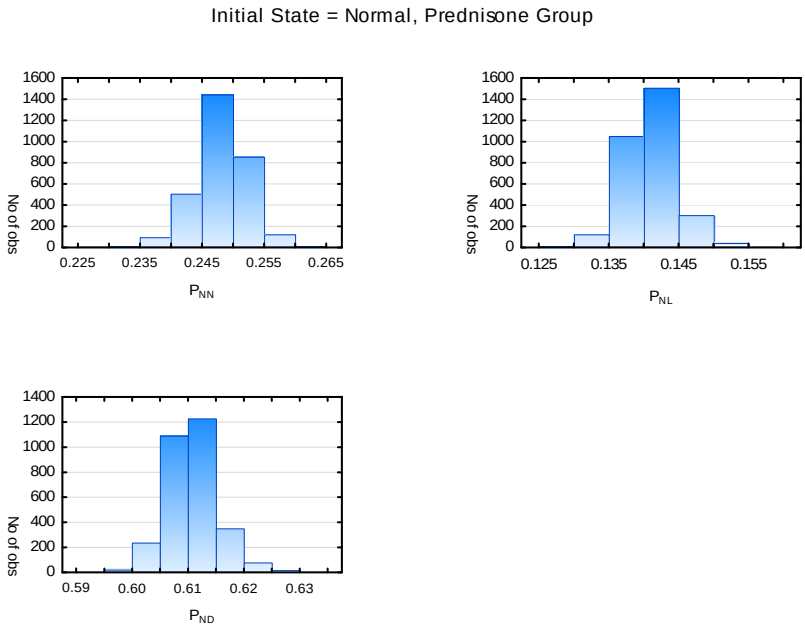

Figure 5.23: Two year posterior predictive distributions for LC data (Initial state = Normal, Prednisone).

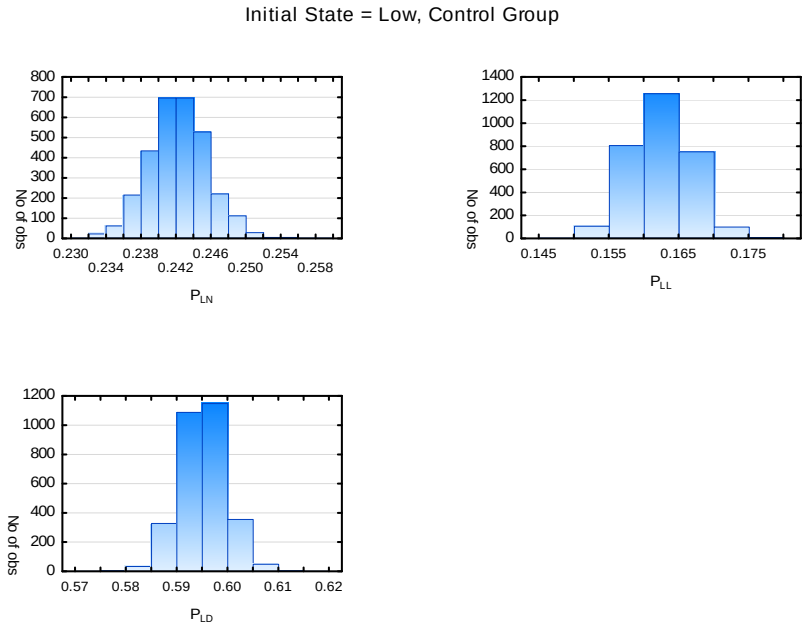

Figure 5.24: Two year posterior predictive distributions for LC data (Initial state = Low, Control).

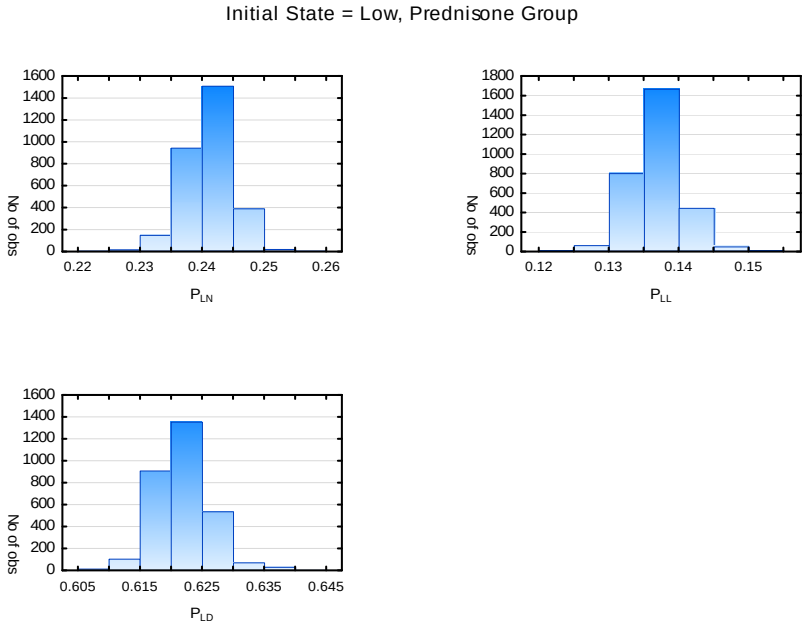

Figure 5.25: Two year posterior predictive distributions for LC data (Initial state = Low, Prednisone).

5.6 Conclusion

In this chapter, two Bayesian multi-state imputation techniques were developed to incorporate prior information into the multi-state modelling process.

In Section 5.1, an imputation technique that uses prior probability vectors to impute all missing observations was discussed. The unobserved state at each missing observation is sampled from a multinomial distribution with parameters sampled from a Dirichlet distribution. The parameters of the Dirichlet distribution were estimated based on prior probability vectors obtained from clinical experts.

In Section 5.2, an imputation technique that uses prior transition functions as the base functions for a Dirichlet process was discussed. A Dirichlet process is assumed to govern the distribution of the transition point between two known observations and it is used to generate the point at which a transition occurred between two known observations. Three different transition functions that can be used as the base function of the Dirichlet process were discussed. These base functions were used to calculate the parameters of a Dirichlet distribution and the Dirichlet distribution was then used to generate the parameters of a multinomial distribution. This multinomial distribution was then used to calculate in which interval between the two known observations the transition took place.

As it can be difficult to understand a process by just looking at the underlying theory, a short example was used in Section 5.3 to illustrate the two different imputation techniques.

In Section 5.4, an extensive simulation study was undertaken to assess the performance of the two imputation techniques under different models and data scenarios. In total, 6 different multi-state models, some with and without covariates, each under 6 different data scenarios were used to generate multi-state data sets. For each data set, 5 different prior distributions were used to impute the missing observations using the two proposed imputation techniques. Once the missing values in the data sets were imputed, a multi-state model was fitted to the data and the transition probabilities calculated. The posterior distribution for each parameter in the model was generated by repeating the imputation process 5000 times for simulated data sets. The MSE was the main statistic used to assess the posterior distributions and it was found that if the correct prior was used, both techniques give results that are comparable, and at times better, to those found in Sections 2.3.3 and 4.3.1.2. In this section it was shown that the

main advantage the two techniques presented here has over the frequentist method of modelling multi-state data discussed in Section 2.3.3, is that they gave consistent parameter estimates, i.e. small MSE's, even for small data sets with a large percentage of missing observations. This indicates that by incorporating prior information into the multi-state data sets we are able to fit multi-state models to data sets that would previously not yield conclusive results. The Bayesian imputation techniques were also shown to be sensitive to correctly specifying the prior distribution. If the prior distribution that best matches the underlying parameters was chosen, the models performed better, i.e. had smaller MSE's and bias, than models with prior distributions that do not match the parameters. It was found that imputing the transition point between two known observations is a more robust technique, i.e. less sensitive to incorrectly specified prior distributions, than when imputing all unknown states.

Finally in Section 5.5 two published multi-state data sets were modelled using the two imputation techniques. The results of the Bayesian models, selected based on DIC and GOF statistics, were compared to those of the corresponding frequentist models.

6

Conclusions and further research areas

The primary aim of this research was to develop Bayesian multi-state models that allow the incorporation of prior clinical expertise into the multi-state modelling process. This aim was achieved by developing the following four Bayesian methods of modelling multi-state data:

- A Bayesian multi-state model was developed where the likelihood is expressed in terms of the limiting probabilities of a Markov process. Prior distributions - namely the MDI and the Jeffreys priors - are placed on the limiting probabilities, and a Metropolis-Hastings algorithm is used to sample variates from the posterior distributions.
- A Bayesian multi-state model was developed where the transition rates are directly modelled in the likelihood, and priors are placed on the transition rates.[7] (*See footnote* 7) This model was extended to allow for the incorporation of covariates into the model.
- A Bayesian multi-state imputation technique was presented that uses prior probability vectors obtained from clinical experts to impute all missing observations in the data set. A multinomial distribution with parameters from a Dirichlet distribution is used to sample the unknown observations.
- A Bayesian multi-state imputation technique was presented where a Dirichlet process is used to estimate the unknown transition point between two known observations. Prior information about the transition process is incorporated in the Dirichlet process by means of prior transition functions that govern the imputation process. Three different transition functions were discussed.

The development of these four B-MSM's necessitated the development of a procedure that could generate panel data sets from populations with known parameters. To this end, a data generating procedure was developed. This procedure allows panel data sets to be generated for a myriad of different underlying multi-state models. This procedure was used extensively

[7] As the transition rates are the rates at which transitions are made to and from different states, the exponential distribution is used as a prior distribution.

to generate data sets so as to assess the performance of the proposed B-MSM's.

Extensive simulation studies were performed to investigate the properties of the four proposed techniques and to assess how they perform under different models and data scenarios. Through this simulation process it was shown that the estimates obtained from these methods are comparable to, and at times better than, those of the frequentist approach to modelling multi-state data. For smaller data sets or more complex models, they where able to provide more stable estimates than the frequentist approach and when fitted to published multi-state data sets they outperformed the frequentist models.

The secondary aim of this research was to investigate the properties of the frequentist Markov model when fitted to multi-state models under varying model and data size scenarios and to ascertain under what situations this process yields unstable results. From this investigation it was found that - when fitting simple 3-state models - stable parameter estimates could be obtained with as few as 25 individuals in the data set. However, as soon as more complex 4-state models were fitted or covariates were included in the models, not even data sets with as many as 75 individuals will give stable results.

Further research possibilities include the following:

- The simulation studies can be extended by considering more complex multi-state models and different data scenarios. This can be done by specifying multi-state models with more complex transition rate patterns, for example, 5-state recurring models, and increasing the sample size, the time period under study and the percentage of missing observations used to generate data sets from these more complex models.
- The B-MSM's presented in Chapter 4 can be extended to models other than recurring 3-state models. Currently the theory and computer programs used to fit B-MSM's are limited to 3-state models. By developing the necessary theory and altering the computer programs used to fit the models, the B-MSM's can be extended to allow it to be fitted to data sets with more than 3 states.
- Covariate information can be incorporated into the prior probability vectors used when imputing all missing observations. This would entail combining information from the data with the clinicians' prior beliefs so as to generate the prior probability vectors. One possible

way this can be accomplished is to use a continuation-ratio logits model to model the probability of being in each state (Agresti, 2002, pp. 289-290).

- The prior distributions placed on the transition function used to estimate the transition point between two known observations, can be formulated to include covariate information. One possible way of achieving this is by regressing the shape and scale parameters of the Weibull function on the covariates in the data, or by replacing the Weibull function with an appropriately specified regression function of the covariates in the data set.
- Coverage probabilities can be calculated for a larger number of multi-state models and data scenarios, to better assess the frequentist properties of the B-MSM's.
- Kay (1986) proposed using interpolation to estimate exact transition times and then using these times to create a complete data set (see Section 2.2). Tests can then be performed on the complete data set to assess the Markov assumption. Instead of using interpolation to complete the data set, the imputation methods developed here could be used to create the complete data set.

Multi-state models are actively being used to model and understand the behaviour of complex systems. By incorporating prior information into the modelling process of these complex systems it is the author's belief that the underlying process in these complex systems can be better modelled and thus better explained and understood.

Appendecis

In these appendecis the R-programs used in the dissertation are presented. The programs used in each section are given first and then the code that was used to generate the results, by making use of the programs, is given.

A.1 Simulating a panel data set

The following functions and program were used in Section 2.3.3.

```
simulate.data <- function(num.pat,pos.times,num.obs.pat,vis.dif,sim.qmatrix,cov.eff.x,cov.eff.y){
    data <- NULL
    reg <- 0
    num.states <- dim(sim.qmatrix)[[1]]
    while (reg < num.pat) {
        reg <- reg + 1
        num.obs <- round(runif(1,2,num.obs.pat))
        pat.times <- sort(sample(pos.times, size=num.obs))
        if (0 != min(pat.times)){
            pat.times <- c(0,pat.times)
            num.obs <- num.obs + 1
        }
        visits <- 1:num.obs
        ran.start <- round(runif(1,1,3))
        x.val <- round(runif(1,0,1))
        y.val <- round(runif(1,1,15))
        simul.ind.data <- simmulti.msm(data=data.frame(subject=reg, time=pat.times, x=x.val, y=y.val),
            qmatrix=sim.qmatrix, start=ran.start, covariates=list(x=cov.eff.x,y=cov.eff.y))
        act.num.obs <- dim(simul.ind.data)[[1]]
        stage.diff <- simul.ind.data[2:act.num.obs,"state"]-simul.ind.data[1:(act.num.obs-1),"state"]
        time.diff <- simul.ind.data[2:act.num.obs,"time"]-simul.ind.data[1:(act.num.obs-1),"time"]
        problem.int <- (abs(stage.diff) > 1) & (time.diff == vis.dif)
```

```
        if (sum(problem.int) == 0){
            data <- rbind(data,cbind(simul.ind.data,"Visit"=1:act.num.obs,"Indicator"=1))
        }
        else{
            reg <- reg - 1
        }
    }
    return(data)
}

Create.Sim.Data <- function(Pat.data,Pat.ID,Visit,Time,Stage,Indicator,Absorbing){
    IDs <- unique(Pat.data[,Pat.ID])
    num.pat <- length(IDs)
    uni.times <- sort(unique(Pat.data[,Time]))
    num.times <- length(uni.times)
    temp.visit <- matrix(NA,nrow=1,ncol=dim(Pat.data)[[2]],dimnames=list(c(),dimnames(Pat.data)[[2]]))
    temp.visit[,Indicator] <- 0
    new.data <- NULL
    for (i in IDs){
        num.obs <- dim(Pat.data[Pat.data[,Pat.ID]==i,])[[1]]
        x.val <- Pat.data[Pat.data[,Pat.ID]==i,'x'][1]
        y.val <- Pat.data[Pat.data[,Pat.ID]==i,'y'][1]
        act.visit <- Pat.data[Pat.data[,Pat.ID]==i,Time]
        mis.times <- uni.times[match(uni.times,act.visit,nomatch=0) == 0]
        num.new.obs <- length(mis.times)
        temp.new.data <- NULL
        temp.visit[,Pat.ID] <- i
        temp.visit[,'x'] <- x.val
        temp.visit[,'y'] <- y.val
        for (j in mis.times) {
            temp.visit[,Time] <- j
```

```
            temp.new.data <- rbind(temp.new.data,temp.visit)
        }
        new.pat.data <- rbind(Pat.data[Pat.data[,Pat.ID]==i,],temp.new.data)
        new.pat.data <- new.pat.data[match(sort(new.pat.data[,Time]),new.pat.data[,Time]),]
        new.pat.data[,Visit] <- c(1:dim(new.pat.data)[[1]])
        new.data <- rbind(new.data,new.pat.data)
    }
    return(new.data)
}
```

The following program code uses the above functions to generate multi-state data sets for different transition matrixes and data scenarios, fits a Markov model to the generate data and calculates and collates the results.

```
sim.q.mat <- c(1,2,3,4,5,6)
sim.scn.doen <- c(1,2,3,4,5,6)
num.rep <- 5500
for (q.mat in sim.q.mat) {
for (sim.scn in sim.scn.doen){
    start.time <-proc.time()
    top.dir.name <- paste('d:\\CJBMuller\\My Documents\\Navorsing\\PhD\\Multi-State Models\\
        Sagteware en Rekenaar Werk\\Finale Simulasie\\',sep='')
    next.dir.name <- paste('3 State\\Hfstk 6\\Covariates\\NO.COV.Q',q.mat,'.Sc',sim.scn,'.',sep='')
    dir.name <- paste(top.dir.name,next.dir.name,sep='')
    # Set type of model by the transition matrix
    if (q.mat == 1) {
        sim.qmatrix <- rbind(c(0, 0.5, 0),c(0.5, 0, 0.5),c(0, 0.5, 0))
        pop.par <- c(0.5,0.5,0.5,0.5)
    }
    if (q.mat == 2) {
        sim.qmatrix <- rbind(c(0, 0.25, 0),c(0.75, 0, 0.25),c(0, 0.75, 0))
        pop.par <- c(0.25,0.75,0.25,0.75)
```

```
}
if (q.mat == 3) {
    sim.qmatrix <- rbind(c(0, 0.75, 0),c(0.25, 0, 0.75),c(0, 0.25, 0))
    pop.par <- c(0.75,0.25,0.75,0.25)
}
if (q.mat == 4) {
    sim.qmatrix <- rbind(c(0, 0.5, 0, 0),c(0.5, 0, 0.5, 0),c(0, 0.5, 0, 0.5),c(0, 0, 0.5, 0))
    pop.par <- c(0.5,0.5,0.5,0.5,0.5,0.5,cov.effect.x,cov.effect.y)
}
if (q.mat == 5) {
    Sim.qmatrix <- rbind(c(0, 0.25, 0, 0),c(0.75, 0, 0.25, 0),c(0, 0.75, 0, 0.25),c(0, 0, 0.75, 0))
    pop.par <- c(0.25,0.75,0.25,0.75,0.25,0.75,cov.effect.x,cov.effect.y)
}
if (q.mat == 6) {
    sim.qmatrix <- rbind(c(0, 0.75, 0, 0),c(0.25, 0, 0.75, 0),c(0, 0.25, 0, 0.75),c(0, 0, 0.25, 0))
    pop.par <- c(0.75,0.25,0.75,0.25,0.75,0.25,cov.effect.x,cov.effect.y)
}
# Set Data scenario
# Info on time frame 0 to 24 months, number of patients, max number of observations per patient
times <- seq(0,24,1)
if (sim.scn == 1) {
    num.pat <- 25
    missing.perc <- 0.1
}
if (sim.scn == 2) {
    num.pat <- 25
    missing.perc <- 0.5
}
if (sim.scn == 3) {
    num.pat <- 50
```

```
        missing.perc <- 0.1
}
if (sim.scn == 4) {
        num.pat <- 50
        missing.perc <- 0.5
}
if (sim.scn == 5) {
        num.pat <- 75
        missing.perc <- 0.1
}
if (sim.scn == 6) {
        num.pat <- 75
        missing.perc <- 0.5
}
num.obs.pat <- round(length(times)*(1-missing.perc),0)
trans.names.one <- list("1->2","2->1","2->3","3->2")
# Create matrixes for final answers
comb.ml.model <- NULL
# Number of repitions for the type of data
# One Covariate
for (big.rep in 1:num.rep) {
        sim.data <- simulate.data.nocov(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,vis.dif=1)
        sim.model <- msm(state ~time,subject=subject,data=sim.data,qmatrix=sim.qmatrix)
        sim.model.est <- qmatrix.msm(sim.model)[[1]]
        if (q.mat < 4) sim.model.est.uni <- c(sim.model.est[1,c(2)],sim.model.est[2,c(1,3)],sim.model.est[3,c(2)])
        if (q.mat > 3) sim.model.est.uni <- c(sim.model.est[1,c(2)],sim.model.est[2,c(1,3)],sim.model.est[3,c(2,4)]
                ,sim.model.est[4,c(3)])
        ml.model.est <- t(sim.model.est.uni)
        dimnames(ml.model.est)[[2]] <- trans.names.one
        comb.ml.model <- rbind(comb.ml.model,ml.model.est)
```

```
        }
        comb.ml.model <- comb.ml.model[501:5500,]
        ml.mean <- apply(comb.ml.model,2,mean)
        ml.med <- apply(comb.ml.model,2,median)
        ml.sd <- apply(comb.ml.model,2,sd)
        ml.MSE <- cal.MSE(comb.ml.model,pop.par)
        comb.summary.ml <- rbind(ml.mean,ml.med,ml.sd,ml.MSE)
        file.name <- paste(dir.name,'ml.summary.csv',sep='')
        write.csv(round(comb.summary.ml,5),file=file.name)
}
}
```

A.2 Limiting probabilities in the likelihood

The following functions and program were used in Section 4.3.2.1.

```
posterior.jeff <- function(m1,m2,m3,init.est,iter){
        posterior <- NULL
        posterior <- rbind(posterior,init.est)
        n <- m1 + m2 + m3
        uniq.est <- 1
        i <- 1
        reject <- 0
        lamb <- 1
        while (uniq.est < iter){
                curr.est <- posterior[i,]
                cand.est <- curr.est
                if (lamb > 4) {
                        lamb <- 1
                }
                if (lamb == 1) {
                        pos.est.1 <- rexp(1,curr.est[1])
```

```
        while (pos.est.1 >= 1) {
                pos.est.1 <- rexp(1,curr.est[1])
        }
cand.est[1] <- pos.est.1
}
if (lamb == 2) {
        pos.est.2 <- rexp(1,curr.est[2])
        while (pos.est.2 >= 1) {
                pos.est.2 <- rexp(1,curr.est[2])
}
cand.est[2] <- pos.est.2
}
if (lamb == 3) {
        pos.est.3 <- rexp(1,curr.est[3])
        while (pos.est.3 >= 1) {
                pos.est.3 <- rexp(1,curr.est[3])
        }
        cand.est[3] <- pos.est.3
}
if (lamb == 4) {
        pos.est.4 <- rexp(1,curr.est[4])
        while (pos.est.4 >= 1) {
                pos.est.4 <- rexp(1,curr.est[4])
        }
cand.est[4] <- pos.est.4
}
k.curr <- curr.est[2]*curr.est[4] + curr.est[1]*curr.est[4] + curr.est[3]*curr.est[1]
p1.curr <- (curr.est[2]*curr.est[4]/k.curr)
p2.curr <- (curr.est[1]*curr.est[4]/k.curr)
k.cand <- cand.est[2]*cand.est[4] + cand.est[1]*cand.est[4] + cand.est[3]*cand.est[1]
```

```
        p1.cand <- (cand.est[2]*cand.est[4]/k.cand)
        p2.cand <- (cand.est[1]*cand.est[4]/k.cand)
        post.curr <- (p1.curr^(m1-0.5))*(p2.curr^(m2-0.5))*((1-p1.curr-p2.curr)^(n-m1-m2-0.5))
        post.cand <- (p1.cand^(m1-0.5))*(p2.cand^(m2-0.5))*((1-p1.cand-p2.cand)^(n-m1-m2-0.5))
        q.curr.giv.cand <- (1/(cand.est[1]*cand.est[2]*cand.est[3]*cand.est[4]))*exp((-curr.est[1]/cand.est[1])-
            (curr.est[2]/cand.est[2])-(curr.est[3]/cand.est[3])-(curr.est[4]/cand.est[4]))
        q.cand.giv.curr <- (1/(curr.est[1]*curr.est[2]*curr.est[3]*curr.est[4]))*exp((-cand.est[1]/curr.est[1])-
            (cand.est[2]/curr.est[2])-(cand.est[3]/curr.est[3])-(cand.est[4]/curr.est[4]))
        post.cand.q.curr <- post.cand*q.curr.giv.cand
        post.curr.q.cand <- post.curr*q.cand.giv.curr
        a <- min(post.cand.q.curr/post.curr.q.cand,1)
        u <- runif(1,0,1)
        if (is.nan(a)){
        }
        else {
            if (u <= a){
                posterior <- rbind(posterior,cand.est)
                uniq.est <- uniq.est + 1
            }
            else {
                posterior <- rbind(posterior,curr.est)
                reject <- reject + 1
            }
        }
        i <- i + 1
        lamb <- lamb + 1
    }
    return(list(posterior = posterior,num.rejected = reject, tot.iter = i))
}

posterior.mdi <- function(m1,m2,m3,init.est,iter){
```

```
posterior <- NULL
posterior <- rbind(posterior,init.est)
uniq.est <- 1
i <- 1
reject <- 0
lamb <- 1
while (uniq.est < iter){
    curr.est <- posterior[i,]
    cand.est <- curr.est
    if (lamb > 4) {
        lamb <- 1
    }
    if (lamb == 1) {
        pos.est.1 <- rexp(1,curr.est[1])
        while (pos.est.1 >= 1) {
            pos.est.1 <- rexp(1,curr.est[1])
        }
        cand.est[1] <- pos.est.1
    }
    if (lamb == 2) {
        pos.est.2 <- rexp(1,curr.est[2])
        while (pos.est.2 >= 1) {
            pos.est.2 <- rexp(1,curr.est[2])
            }
        cand.est[2] <- pos.est.2
    }
    if (lamb == 3) {
        pos.est.3 <- rexp(1,curr.est[3])
        while (pos.est.3 >= 1) {
            pos.est.3 <- rexp(1,curr.est[3])
```

```
        }
        cand.est[3] <- pos.est.3
}
if (lamb == 4) {
        pos.est.4 <- rexp(1,curr.est[4])
        while (pos.est.4 >= 1) {
                pos.est.4 <- rexp(1,curr.est[4])
}
        cand.est[4] <- pos.est.4
}
k.curr <- curr.est[2]*curr.est[4] + curr.est[1]*curr.est[4] + curr.est[3]*curr.est[1]
p1.curr <- (curr.est[2]*curr.est[4]/k.curr)
p2.curr <- (curr.est[1]*curr.est[4]/k.curr)
p3.curr <- (curr.est[3]*curr.est[1]/k.curr)
k.cand <- cand.est[2]*cand.est[4] + cand.est[1]*cand.est[4] + cand.est[3]*cand.est[1]
p1.cand <- (cand.est[2]*cand.est[4]/k.cand)
p2.cand <- (cand.est[1]*cand.est[4]/k.cand)
p3.cand <- (cand.est[3]*cand.est[1]/k.cand)
post.curr <- (p1.curr^(m1+p1.curr))*(p2.curr^(m2+p2.curr))*(p3.curr^(m3+p3.curr))
post.cand <- (p1.cand^(m1+p1.cand))*(p2.cand^(m2+p2.cand))*(p3.cand^(m3+p3.cand))
q.curr.giv.cand <- (1/(cand.est[1]*cand.est[2]*cand.est[3]*cand.est[4]))*exp(-curr.est[1]/cand.est[1]
        -curr.est[2]/cand.est[2]-curr.est[3]/cand.est[3]-curr.est[4]/cand.est[4])
q.cand.giv.curr <- (1/(curr.est[1]*curr.est[2]*curr.est[3]*curr.est[4]))*exp(-cand.est[1]/curr.est[1]
        -cand.est[2]/curr.est[2]-cand.est[3]/curr.est[3]-cand.est[4]/curr.est[4])
comb.curr <- post.curr * q.cand.giv.curr
comb.cand <- post.cand * q.curr.giv.cand
a <- min(comb.cand/comb.curr,1)
u <- runif(1,0,1)
if (is.nan(a)){
}
```

```
        else {
            if (u <= a){
                posterior <- rbind(posterior,cand.est)
                uniq.est <- uniq.est + 1
            }
            else {
                posterior <- rbind(posterior,curr.est)
                reject <- reject + 1
            }
        }
        i <- i + 1
        lamb <- lamb + 1
    }
return(list(posterior = posterior,num.rejected = reject, tot.iter = i))
}
```

```
cal.MSE <- function(data.mat, pop.par) {
    MSE <- function(x,pop.par) {
        par.num <- x[1]
        x <- x[-1]
        sum((x-pop.par[par.num])^2)/length(x)
    }
    MSE.eq <- apply(rbind(1:dim(data.mat)[[2]],data.mat),2,MSE,pop.par=pop.par)
    return(MSE.eq)
}
```

The following program code uses the above functions to generate posterior distributions for Bayesian multi-state models based on using limiting probabilities in the likelihood. Data sets for different transition matrixes and data scenarios are generated, the Bayesian models are fitted and the posterior distributions generated, summarised and the results collated.

```
sim.prior.doen <- c(1,2,3,4)
sim.scn.doen <- c(1,2,3,4)
```

```
sim.q.mat <- c(1,2,3)
for (sim.prior in sim.prior.doen){
for (sim.scn in sim.scn.doen){
for (q.mat in sim.q.mat){
    top.dir.name <- paste('d:\\CJBMuller\\My Documents\\Navorsing\\PhD\\Multi-State Models\\
        Sagteware en Rekenaar Werk\\Finale Simulasie\\',sep='')
    next.dir.name <- paste('3 State\\Hfstk 5\\Q',q.mat,'.Sc',sim.scn,'.Pr',sim.prior,'.',sep='')
    dir.name <- paste(top.dir.name,next.dir.name,sep='')
    # Set type of model by the transition matrix
    if (q.mat == 1) {
        sim.qmatrix <- rbind(c(0, 0.5, 0),c(0.5, 0, 0.5),c(0, 0.5, 0))
        pop.par <- c(0.5,0.5,0.5,0.5)
    }
    if (q.mat == 2) {
        sim.qmatrix <- rbind(c(0, 0.25, 0),c(0.75, 0, 0.25),c(0, 0.75, 0))
        pop.par <- c(0.25,0.75,0.25,0.75)
    }
    if (q.mat == 3) {
        sim.qmatrix <- rbind(c(0, 0.75, 0),c(0.25, 0, 0.75),c(0, 0.25, 0))
        pop.par <- c(0.75,0.25,0.75,0.25)
    }
    # Set Data scenario
    times <- seq(0,24,1)
    if (sim.scn == 1) {
        num.pat <- 25
        missing.perc <- 0.1
    }
    if (sim.scn == 2) {
        num.pat <- 25
        missing.perc <- 0.5
```

```
}
if (sim.scn == 3) {
        num.pat <- 50
        missing.perc <- 0.1
}
if (sim.scn == 4) {
        num.pat <- 50
        missing.perc <- 0.5
}
num.obs.pat <- round(length(times)*(1-missing.perc),0)
#Prior 52
if (sim.prior == 1){
        prior.lambda.52 <- c(0.2,0.2,0.2,0.2)
}
if (sim.prior == 2){
        prior.lambda.52 <- c(0.8,0.8,0.8,0.8)
}
if (sim.prior == 3){
        prior.lambda.52 <- c(0.2,0.8,0.2,0.8)
}
if (sim.prior == 4){
        prior.lambda.52 <- c(0.8,0.2,0.8,0.2)
}
s.trans.names <- list("1->2","2->1","2->3","3->2")
# Create matrixes for final answers
comb.post.mean <- NULL
comb.post.med <- NULL
comb.post.sd <- NULL
comb.post.perc <- NULL
comb.post.perc.len <- NULL
```

```
comb.post.HPD <- NULL
comb.post.HPD.len <- NULL
comb.post.MSE <- NULL
cov.perc.511 <- c(0,0,0,0)
cov.HPD.511 <- c(0,0,0,0)
cov.perc.512 <- c(0,0,0,0)
cov.HPD.512 <- c(0,0,0,0)
# Number of repitions for the type of data
num.rep <- 4
iter.51 <- 5500
for (big.rep in 1:num.rep) {
    # Simulate dataset based on tran matrix
    sim.data <- simulate.data.nocov(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,
                vis.dif=1)
    sim.data.cov <- simulate.data.cov(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,
                vis.dif=1,cov.effect)
    # Sec 5.1: Prior op Limiting Prob
    num.vis <- apply(statetable.msm(state,subject,data=sim.data),2,sum)
    m1 <- num.vis[1]
    m2 <- num.vis[2]
    m3 <- num.vis[3]
    init.est <- c(0.5,0.5,0.5,0.5)
    post.511 <- posterior.jeff.ind.lamb(m1,m2,m3,init.est,iter.51)
    post.511 <- post.511[501:5500,]
    post.511.mean <- apply(post.511[[1]],2,mean)
    post.511.med <- apply(post.511[[1]],2,median)
    post.511.sd <- apply(post.511[[1]],2,sd)
    post.511.perc.l <- apply(post.511[[1]],2,quantile,probs=c(0.025))
    post.511.perc.u <- apply(post.511[[1]],2,quantile,probs=c(0.975))
    post.511.perc.len <- post.511.perc.u - post.511.perc.l
```

```
post.511.HPD <- HPDinterval(as.mcmc(post.511[[1]]))
post.511.HPD.l <- post.511.HPD[,1]
post.511.HPD.u <- post.511.HPD[,2]
post.511.HPD.len <- post.511.HPD[,2] - post.511.HPD[,1]
post.511.MSE <- cal.MSE(post.511[[1]],pop.par)
cov.perc.511 <- cov.perc.511 + ((post.511.perc.l <= pop.par) & (pop.par <= post.511.perc.u))
cov.HPD.511 <- cov.HPD.511 + ((post.511.HPD.l <= pop.par) & (pop.par <= post.511.HPD.u))
post.512 <- posterior.mdi.ind.lamb(m1,m2,m3,init.est,iter.51)
post.512 <- post.512[501:5500,]
post.512.mean <- apply(post.512[[1]],2,mean)
post.512.med <- apply(post.512[[1]],2,median)
post.512.sd <- apply(post.512[[1]],2,sd)
post.512.perc.l <- apply(post.512[[1]],2,quantile,probs=c(0.025))
post.512.perc.u <- apply(post.512[[1]],2,quantile,probs=c(0.975))
post.512.perc.len <- post.512.perc.u - post.512.perc.l
post.512.HPD <- HPDinterval(as.mcmc(post.512[[1]]))
post.512.HPD.l <- post.512.HPD[,1]
post.512.HPD.u <- post.512.HPD[,2]
post.512.HPD.len <- post.512.HPD[,2] - post.512.HPD[,1]
post.512.MSE <- cal.MSE(post.512[[1]],pop.par)
cov.perc.512 <- cov.perc.512 + ((post.512.perc.l <= pop.par) & (pop.par <= post.512.perc.u))
cov.HPD.512 <- cov.HPD.512 + ((post.512.HPD.l <= pop.par) & (pop.par <= post.512.HPD.u))
# Posterior Summary of techniques
post.mean <- post.511.mean
post.med <- post.511.med
post.sd <- post.511.sd
post.perc.l <- post.511.perc.)
post.perc.u <- post.511.perc.u
post.perc.len <- post.511.perc.len
post.HPD.l <- post.511.HPD.l
```

```
        post.HPD.u <- post.511.HPD.u
        post.HPD.len <- post.511.HPD.len
        post.MSE <- post.511.MSE
        post.mean <- cbind(big.rep,post.mean)
        post.med <- cbind(big.rep,post.med)
        post.sd <- cbind(big.rep,post.sd)
        post.perc <- cbind(big.rep,post.perc)
        post.perc.len <- cbind(big.rep,post.perc.len)
        post.HPD <- cbind(big.rep,post.HPD)
        post.HPD.len <- cbind(big.rep,post.HPD.len)
        post.MSE <- cbind(big.rep,post.MSE)
        comb.post.mean <- rbind(comb.post.mean,post.mean)
        comb.post.med <- rbind(comb.post.med,post.med)
        comb.post.sd <- rbind(comb.post.sd,post.sd)
        comb.post.perc <- rbind(comb.post.perc,post.perc)
        comb.post.perc.len <- rbind(comb.post.perc.len,post.perc.len)
        comb.post.HPD <- rbind(comb.post.HPD,post.HPD)
        comb.post.HPD.len <- rbind(comb.post.HPD.len,post.HPD.len)
        comb.post.MSE <- rbind(comb.post.MSE,post.MSE)
}
post.cov.perc.count <- cov.perc.511
post.cov.HPD.count <- cov.HPD.511
post.cov.perc <- post.cov.perc.count/num.rep
post.cov.HPD <- post.cov.HPD.count/num.rep
post.cov.count <- rbind(post.cov.perc.count,post.cov.HPD.count)
post.cov <- rbind(post.cov.perc,post.cov.HPD)
rows.511 <- c(1,5,9,13)
rows.512 <- c(2,6,10,14)
mean.511 <- apply(comb.post.mean[rows.511,-1],2,mean)
med.511 <- apply(comb.post.med[rows.511,-1],2,mean)
```

```
    stdev.511 <- apply(comb.post.sd[rows.511,-1],2,mean)
    mse.511 <- apply(comb.post.MSE[rows.511,-1],2,mean)
    mean.512 <- apply(comb.post.mean[rows.512,-1],2,mean)
    med.512 <- apply(comb.post.med[rows.512,-1],2,mean)
    stdev.512 <- apply(comb.post.sd[rows.512,-1],2,mean)
    mse.512 <- apply(comb.post.MSE[rows.512,-1],2,mean)
    comb.summary <- rbind(mean.511,med.511,stdev.511,mse.511,mean.512,med.512,stdev.512,mse.512)
    file.name <- paste(dir.name,'comb.summary.csv',sep='')
    write.csv(round(comb.summary,5),file=file.name)
}
}
}
```

A.3 Using the likelihood with transition rates

The following functions and program were used in Sections 4.3.2.2 and 4.3.2.3.

```
posterior.theta <- function(pat.data,init.est,iter,varcov.est,lambda){
    posterior <- NULL
    posterior <- rbind(posterior,init.est)
    uniq.est <- 1
    tot.it <- 1
    c.1 <- 1
    c.2 <- 1
    c.3 <- 1
    c.4 <- 1
    reject <- 0
    rej.c1 <- 0
    rej.c2 <- 0
    rej.c3 <- 0
    rej.c4 <- 0
    while (uniq.est < iter){
```

```
current.est <- posterior[tot.it,]
candidate.est <- current.est
pos.est.1 <- abs(rnorm(1,mean=current.est[1], sd=varcov.est[1]))
while (pos.est.1 >= 1){
        pos.est.1 <- abs(rnorm(1,mean=current.est[1], sd=varcov.est[1]))
}
candidate.est[1] <- pos.est.1
trans.matrix.current <- rbind(c(0,current.est[1],0),c(current.est[2],0,current.est[3]),
        c(0,current.est[4],0))
trans.matrix.candidate <- rbind(c(0,candidate.est[1],0),c(candidate.est[2],0,candidate.est[3]),
        c(0,candidate.est[4],0))
like.current <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.current,
      fixedpars=TRUE)$minus2loglik/-2)
like.candidate <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.candidate,
        fixedpars=TRUE)$minus2loglik/-2)
post.current <- like.current*exp(-trans.matrix.current[1,2]*lambda[1])*exp(-trans.matrix.current[2,1]*
        lambda[2])*exp(-trans.matrix.current[2,3]*lambda[3])*exp(-trans.matrix.current[3,2]*lambda[4])
post.candidate <- like.candidate*exp(-trans.matrix.candidate[1,2]*lambda[1])*exp(-trans.matrix.candidate[2,1]*
      lambda[2])*exp(-trans.matrix.candidate[2,3]*lambda[3])*exp(-trans.matrix.candidate[3,2]*lambda[4])
a <- min(post.candidate/post.current,1)
u <- runif(1,0,1)
if (is.nan(a)){
}
else {
        tot.it <- tot.it + 1
        if (u <= a){
                posterior <- rbind(posterior,candidate.est)
                uniq.est <- uniq.est + 1
                rej.c1 <- 0
        }
```

```
        else {
                posterior <- rbind(posterior,current.est)
                reject <- reject + 1
                rej.c1 <- rej.c1 + 1
        }
}
current.est <- posterior[tot.it,]
candidate.est <- current.est
pos.est.2 <- abs(rnorm(1,mean=current.est[2], sd=varcov.est[2]))
while (pos.est.2 >= 1){
        pos.est.2 <- abs(rnorm(1,mean=current.est[2], sd=varcov.est[2]))
}
candidate.est[2] <- pos.est.2
trans.matrix.current <- rbind(c(0,current.est[1],0),c(current.est[2],0,current.est[3]),
        c(0,current.est[4],0))
trans.matrix.candidate <- rbind(c(0,candidate.est[1],0),c(candidate.est[2],0,candidate.est[3]),
        c(0,candidate.est[4],0))
like.current <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.current,
        fixedpars=TRUE)$minus2loglik/-2)
like.candidate <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.candidate,
        fixedpars=TRUE)$minus2loglik/-2)
post.current <- like.current*exp(-trans.matrix.current[1,2]*lambda[1])*exp(-trans.matrix.current[2,1]*
        lambda[2])*exp(-trans.matrix.current[2,3]*lambda[3])*exp(-trans.matrix.current[3,2]*lambda[4])
post.candidate <- like.candidate*exp(-trans.matrix.candidate[1,2]*lambda[1])*exp(-trans.matrix.candidate[2,1]*
        lambda[2])*exp(-trans.matrix.candidate[2,3]*lambda[3])*exp(-trans.matrix.candidate[3,2]*lambda[4])
a <- min(post.candidate/post.current,1)
u <- runif(1,0,1)
if (is.nan(a)){
}
else {
```

```
        tot.it <- tot.it + 1
        if (u <= a){
                posterior <- rbind(posterior,candidate.est)
                uniq.est <- uniq.est + 1
                rej.c2 <- 0
                }
        else {
                posterior <- rbind(posterior,current.est)
                reject <- reject + 1
                rej.c3 <- rej.c3 + 1
        }
}
current.est <- posterior[tot.it,]
candidate.est <- current.est
pos.est.3 <- abs(rnorm(1,mean=current.est[3], sd=varcov.est[3]))
while (pos.est.1 >= 1){
        pos.est.3 <- abs(rnorm(1,mean=current.est[3], sd=varcov.est[3]))
}
candidate.est[3] <- pos.est.3
trans.matrix.current <- rbind(c(0,current.est[1],0),c(current.est[2],0,current.est[3]),
        c(0,current.est[4],0))
trans.matrix.candidate <- rbind(c(0,candidate.est[1],0),c(candidate.est[2],0,candidate.est[3]),
        c(0,candidate.est[4],0))
like.current <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.current,
      fixedpars=TRUE)$minus2loglik/-2)
like.candidate <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.candidate,
        fixedpars=TRUE)$minus2loglik/-2)
post.current <- like.current*exp(-trans.matrix.current[1,2]*lambda[1])*exp(-trans.matrix.current[2,1]*
        lambda[2])*exp(-trans.matrix.current[2,3]*lambda[3])*exp(-trans.matrix.current[3,2]*lambda[4])
post.candidate <- like.candidate*exp(-trans.matrix.candidate[1,2]*lambda[1])*exp(-trans.matrix.candidate[2,1]*
```

```
        lambda[2])*exp(-trans.matrix.candidate[2,3]*lambda[3])*exp(-trans.matrix.candidate[3,2]*lambda[4])
a <- min(post.candidate/post.current,1)
u <- runif(1,0,1)
if (is.nan(a)){
}
else {
        tot.it <- tot.it + 1
        if (u <= a){
                posterior <- rbind(posterior,candidate.est)
                uniq.est <- uniq.est + 1
                rej.c3 <- 0
        }
        else {
                posterior <- rbind(posterior,current.est)
                reject <- reject + 1
                rej.c3 <- rej.c3 + 1
        }
}
current.est <- posterior[tot.it,]
candidate.est <- current.est
pos.est.4 <- abs(rnorm(1,mean=current.est[4], sd=varcov.est[4]))
while (pos.est.4 >= 1){
        pos.est.4 <- abs(rnorm(1,mean=current.est[4], sd=varcov.est[4]))
}
candidate.est[4] <- pos.est.4
trans.matrix.current <- rbind(c(0,current.est[1],0),c(current.est[2],0,current.est[3]),
        c(0,current.est[4],0))
trans.matrix.candidate <- rbind(c(0,candidate.est[1],0),c(candidate.est[2],0,candidate.est[3]),
        c(0,candidate.est[4],0))
like.current <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.current,
```

```
            fixedpars=TRUE)$minus2loglik/-2)
        like.candidate <- exp(msm(state ~time,subject=subject,data=pat.data,qmatrix=trans.matrix.candidate,
            fixedpars=TRUE)$minus2loglik/-2)
        post.current <- like.current*exp(-trans.matrix.current[1,2]*lambda[1])*exp(-trans.matrix.current[2,1]*
            lambda[2])*exp(-trans.matrix.current[2,3]*lambda[3])*exp(-trans.matrix.current[3,2]*lambda[4])
        post.candidate <- like.candidate*exp(-trans.matrix.candidate[1,2]*lambda[1])*exp(-trans.matrix.candidate[2,1]*
            lambda[2])*exp(-trans.matrix.candidate[2,3]*lambda[3])*exp(-trans.matrix.candidate[3,2]*lambda[4])
        a <- min(post.candidate/post.current,1)
        u <- runif(1,0,1)
        if (is.nan(a)){
        }
        else {
            tot.it <- tot.it + 1
            if (u <= a){
                posterior <- rbind(posterior,candidate.est)
                uniq.est <- uniq.est + 1
                rej.c4 <- 0
            }
            else {
                posterior <- rbind(posterior,current.est)
                reject <- reject + 1
                rej.c4 <- rej.c4 + 1
            }
        }
    }
    return(list(posterior = posterior,num.rejected = reject,tot.iter = tot.it))
}

posterior.theta.cov <- function(pat.data,init.est,iter,varcov.trans.rate,varcov.cov.eff,prior.lambda,prior.mu,prior.sig){
    posterior <- NULL
    posterior <- rbind(posterior,init.est)
```

```
uniq.est <- 1
i <- 1
reject <- 0
while (i < iter){
    current.est <- posterior[i,]
    current.est.trans.rate <- current.est[1:4]
    candidate.est.trans.rate <- abs(rmnorm(mean = current.est.trans.rate,varcov =
        diag(varcov.trans.rate,nrow=4,ncol=4)))
    trans.matrix.current <- rbind(c(0,current.est.trans.rate[1],0),c(current.est.trans.rate[2],0,
        current.est.trans.rate[3]),c(0,current.est.trans.rate[4],0))
    trans.matrix.candidate <- rbind(c(0,candidate.est.trans.rate[1],0),c(candidate.est.trans.rate[2],0,
        candidate.est.trans.rate[3]),c(0,candidate.est.trans.rate[4],0))
    current.est.cov.eff <- current.est[5:8]
    candidate.est.cov.eff <- rmnorm(mean = current.est.cov.eff,varcov = diag(varcov.cov.eff,nrow=4,ncol=4))
    like.current <- exp(msm(state ~time,covariates=~x,subject=subject,data=pat.data,qmatrix=
        trans.matrix.current,covinits=list(x=current.est.cov.eff),fixedpars=TRUE)$minus2loglik/-2)
    like.candidate <- exp(msm(state ~time,covariates=~x,subject=subject,data=pat.data,qmatrix=
        trans.matrix.candidate,covinits=list(x=candidate.est.cov.eff),fixedpars=TRUE)$minus2loglik/-2)
    post.current <- like.current*exp(-trans.matrix.current[1,2]*prior.lambda[1])*exp(-trans.matrix.current[2,1]*
        prior.lambda[2])*exp(-trans.matrix.current[2,3]*prior.lambda[3])*exp(-trans.matrix.current[3,2]*
        prior.lambda[4])*exp(-0.5*((current.est.cov.eff[1]-prior.mu[1])^2/prior.sig[1]))*exp(-0.5*
        ((current.est.cov.eff[2]-prior.mu[2])^2/prior.sig[2]))*exp(-0.5*((current.est.cov.eff[3]-prior.mu[3])^2/
        prior.sig[3]))*exp(-0.5*((current.est.cov.eff[4]-prior.mu[4])^2/prior.sig[4]))
    post.candidate <- like.candidate*exp(-trans.matrix.candidate[1,2]*prior.lambda[1])*
        exp(-trans.matrix.candidate[2,1]*prior.lambda[2])*exp(-trans.matrix.candidate[2,3]*prior.lambda[3])*
        exp(-trans.matrix.candidate[3,2]*prior.lambda[4])*exp(-0.5*((candidate.est.cov.eff[1]-prior.mu[1])^2/
        prior.sig[1]))*exp(-0.5*((candidate.est.cov.eff[2]-prior.mu[2])^2/prior.sig[2]))*exp(-0.5*((candidate.est.
        cov.eff[3]-prior.mu[3])^2/prior.sig[3]))*exp(-0.5*((candidate.est.cov.eff[4]-prior.mu[4])^2/prior.sig[4]))
    a <- min(post.candidate/post.current,1)
    u <- runif(1,0,1)
```

```
            if (is.nan(a)){
            }
            else {
                  if (u <= a){
                        candidate.est <- c(candidate.est.trans.rate,candidate.est.cov.eff)
                        posterior <- rbind(posterior,candidate.est)
                        uniq.est <- uniq.est + 1
                  }
            else {
                  posterior <- rbind(posterior,current.est)
                  reject <- reject + 1
                  }
            }
            i <- i + 1
      }
      return(list(posterior = posterior,num.rejected = reject))
}

cal.MSE <- function(data.mat, pop.par) {
      MSE <- function(x,pop.par) {
            par.mum <- x[1]
            x <- x[-1]
            sum((x-pop.par[par.num])^2)/length(x)
      }
      MSE.eq <- apply(rbind(1:dim(data.mat)[[2]],data.mat),2,MSE,pop.par=pop.par)
      return(MSE.eq)
}
```

The following program code uses the above functions to generate posterior distributions for Bayesian multi-state models based on using the transition rates in the likelihood. Data sets for different transition matrixes and data scenarios are generated, the Bayesian models are fitted and the posterior distributions generated, summarised and the results collated.

```
sim.prior.doen <- c(1,2,3,4)
sim.scn.doen <- c(1,2,3,4)
sim.q.mat <- c(1,2,3)
for (sim.prior in sim.prior.doen){
for (sim.scn in sim.scn.doen){
for (q.mat in sim.q.mat){
    top.dir.name <- paste('d:\\CJBMuller\\My Documents\\Navorsing\\PhD\\Multi-State Models\\
        Sagteware en Rekenaar Werk\\Finale Simulasie\\',sep='')
    next.dir.name <- paste('3 State\\Hfstk 5\\Q',q.mat,'.Sc',sim.scn,'.Pr',sim.prior,'.',sep='')
    dir.name <- paste(top.dir.name,next.dir.name,sep='')
    # Set type of model by the transition matrix
    if (q.mat == 1) {
        sim.qmatrix <- rbind(c(0, 0.5, 0),c(0.5, 0, 0.5),c(0, 0.5, 0))
        pop.par <- c(0.5,0.5,0.5,0.5)
    }
    if (q.mat == 2) {
        sim.qmatrix <- rbind(c(0, 0.25, 0),c(0.75, 0, 0.25),c(0, 0.75, 0))
        pop.par <- c(0.25,0.75,0.25,0.75)
    }
    if (q.mat == 3) {
        sim.qmatrix <- rbind(c(0, 0.75, 0),c(0.25, 0, 0.75),c(0, 0.25, 0))
        pop.par <- c(0.75,0.25,0.75,0.25)
    }
    # Set Data scenario
    times <- seq(0,24,1)
    if (sim.scn == 1) {
        num.pat <- 25
        missing.perc <- 0.1
    }
    if (sim.scn == 2) {
```

```
        num.pat <- 25
        missing.perc <- 0.5
}
if (sim.scn == 3) {
        num.pat <- 50
        missing.perc <- 0.1
}
if (sim.scn == 4) {
        num.pat <- 50
        missing.perc <- 0.5
}
num.obs.pat <- round(length(times)*(1-missing.perc),0)
#Prior 52
if (sim.prior == 1){
        prior.lambda.52 <- c(0.2,0.2,0.2,0.2)
}
if (sim.prior == 2){
        prior.lambda.52 <- c(0.8,0.8,0.8,0.8)
}
if (sim.prior == 3){
        prior.lambda.52 <- c(0.2,0.8,0.2,0.8)
}
if (sim.prior == 4){
        prior.lambda.52 <- c(0.8,0.2,0.8,0.2)
}
cov.effect <- c(-0.7,-0.7,-0.7,-0.7)
pop.par.cov <- c(pop.par,cov.effect)
sim.covinits <- c(-0.7,-0.7,-0.7,-0.7)
s.trans.names <- list("1->2","2->1","2->3","3->2")
trans.names <- list("1->2","2->1","2->3","3->2","B1->2","B2->1","B2->3","B3->2")
```

```
trans.names.int <- list("1->2 L","1->2 U","2->1 L","2->1 U","2->3 L","2->3 U","3->2 L","3->2 U",
    "B1->2 L","B1->2 U","B2->1 L","B2->1 U","B2->3 L","B2->3 U","B3->2 L","B3->2 U")
# Create matrixes for final answers
comb.post.mean <- NULL
comb.post.med <- NULL
comb.post.sd <- NULL
comb.post.perc <- NULL
comb.post.perc.len <- NULL
comb.post.HPD <- NULL
comb.post.HPD.len <- NULL
comb.post.MSE <- NULL
comb.post.mean.522 <- NULL
comb.post.med.522 <- NULL
comb.post.sd.522 <- NULL
comb.post.perc.522 <- NULL
comb.post.HPD.522 <- NULL
comb.post.MSE.522 <- NULL
cov.perc.521 <- c(0,0,0,0)
cov.HPD.521 <- c(0,0,0,0)
cov.perc.522 <- c(0,0,0,0,0,0,0,0)
cov.HPD.522 <- c(0,0,0,0,0,0,0,0)
# Number of repitions for the type of data
num.rep <- 4
iter.521 <- 5500
iter.522 <- 5500
for (big.rep in 1:num.rep) {
    # Simulate dataset based on tran matrix
    sim.data <- simulate.data.nocov(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,
            vis.dif=1)
    sim.data.cov <- simulate.data.cov(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,
```

```
            vis.dif=1,cov.effect)
# Sec 5.2.1: Prior of Transition Rates sonder Kovariate
init.est <- c(0.5,0.5,0.5,0.5)
varcov.est <- c(0.025,0.025,0.025,0.025)
post.521 <- posterior.theta.one.one(sim.data,init.est,iter.521,varcov.est,prior.lambda.52)
post.521 <- post.521[501:5500,]
post.521.mean <- apply(post.521[[1]],2,mean)
post.521.med <- apply(post.521[[1]],2,median)
post.521.sd <- apply(post.521[[1]],2,sd)
post.521.perc.l <- apply(post.521[[1]],2,quantile,probs=c(0.025))
post.521.perc.u <- apply(post.521[[1]],2,quantile,probs=c(0.975))
post.521.perc.len <- post.521.perc.u - post.521.perc.l
post.521.HPD <- HPDinterval(as.mcmc(post.521[[1]]))
post.521.HPD.l <- post.521.HPD[,1]
post.521.HPD.u <- post.521.HPD[,2]
post.521.HPD.len <- post.521.HPD[,2] - post.521.HPD[,1]
post.521.MSE <- cal.MSE(post.521[[1]],pop.par)
cov.perc.521 <- cov.perc.521 + ((post.521.perc.l <= pop.par) & (pop.par <= post.521.perc.u))
cov.HPD.521 <- cov.HPD.521 + ((post.521.HPD.l <= pop.par) & (pop.par <= post.521.HPD.u))
# Sec 5.2.2: Prior of Transition Rates met Kovariate
init.est <- c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5)
varcov.trans.rate <- c(0.01,0.01,0.01,0.01)
varcov.cov.eff <- c(0.01,0.01,0.01,0.01)
prior.mu <- c(0.5,0.5,0.5,0.5)
prior.sig <- c(1000,1000,1000,1000)
post.522 <- posterior.theta.cov(pat.data=sim.data.cov,init.est=init.est,iter=iter.522,varcov.trans.rate=
            varcov.trans.rate,varcov.cov.eff=varcov.cov.eff,prior.lambda=prior.lambda.52,prior.mu=
            prior.mu,prior.sig=prior.sig)
post.522 <- post.52[501:5500,]
post.522.mean <- apply(post.522[[1]],2,mean)
```

```
post.522.med <- apply(post.522[[1]],2,median)
post.522.sd <- apply(post.522[[1]],2,sd)
post.522.perc.l <- apply(post.522[[1]],2,quantile,probs=c(0.025))
post.522.perc.u <- apply(post.522[[1]],2,quantile,probs=c(0.975))
post.522.perc.len <- post.522.perc.u - post.522.perc.l
post.522.HPD <- HPDinterval(as.mcmc(post.522[[1]]))
post.522.HPD.l <- post.522.HPD[,1]
post.522.HPD.u <- post.522.HPD[,2]
post.522.HPD.len <- post.522.HPD[,2] - post.522.HPD[,1]
post.522.MSE <- cal.MSE(post.522[[1]],pop.par.cov)
cov.perc.522 <- cov.perc.522 + ((post.522.perc.l <= pop.par.cov) & (pop.par.cov <= post.522.perc.u))
cov.HPD.522 <- cov.HPD.522 + ((post.522.HPD.l <= pop.par.cov) & (pop.par.cov <= post.522.HPD.u))
# Posterior Summary of techniques
post.mean <- rbind(post.521.mean)
post.med <- rbind(post.521.med)
post.sd <- rbind(post.521.sd)
post.perc.l <- rbind(post.521.perc.l)
post.perc.u <- rbind(post.521.perc.u)
post.perc.len <- rbind(post.521.perc.len)
post.HPD.l <- rbind(post.521.HPD.l)
post.HPD.u <- rbind(post.521.HPD.u)
post.HPD.len <- rbind(post.521.HPD.len)
post.MSE <- rbind(post.521.MSE)
post.mean <- cbind(post.mean,0,0,0,0)
post.med <- cbind(post.med,0,0,0,0)
post.sd <- cbind(post.sd,0,0,0,0)
post.perc <- cbind(post.perc.l[,1],post.perc.u[,1],post.perc.l[,2],post.perc.u[,2],post.perc.l[,3],post.perc.u[,3]
              ,post.perc.l[,4],post.perc.u[,4],0,0,0,0,0,0,0,0)
post.perc.len <- cbind(post.perc.len,0,0,0,0)
post.HPD <- cbind(post.HPD.l[,1],post.HPD.u[,1],post.HPD.l[,2],post.HPD.u[,2],post.HPD.l[,3],post.HPD.u[,3]
```

```
                ,post.HPD.l[,4],post.HPD.u[,4],0,0,0,0,0,0,0,0)
post.HPD.len <- cbind(post.HPD.len,0,0,0,0)
post.MSE <- cbind(post.MSE,0,0,0,0)
post.mean <- rbind(post.mean,post.522.mean)
post.med <- rbind(post.med,post.522.med)
post.sd <- rbind(post.sd,post.522.sd)
post.perc.522 <- cbind(post.522.perc.l[1],post.522.perc.u[1],post.522.perc.l[2],post.522.perc.u[2],
                post.522.perc.l[3],post.522.perc.u[3],post.522.perc.l[4],post.522.perc.u[4],post.522.perc.l[5],
                post.522.perc.u[5],post.522.perc.l[6],post.522.perc.u[6],post.522.perc.l[7],post.522.perc.u[7],
                post.522.perc.l[8],post.522.perc.u[8])
post.perc <- rbind(post.perc,post.perc.522)
post.perc.len <- rbind(post.perc.len,post.522.perc.len)
post.HPD.522 <- cbind(post.522.HPD.l[1],post.522.HPD.u[1],post.522.HPD.l[2],post.522.HPD.u[2],
                post.522.HPD.l[3],post.522.HPD.u[3],post.522.HPD.l[4],post.522.HPD.u[4],post.522.HPD.l[5],
                post.522.HPD.u[5],post.522.HPD.l[6],post.522.HPD.u[6],post.522.HPD.l[7],post.522.HPD.u[7],
                post.522.HPD.l[8],post.522.HPD.u[8])
post.HPD <- rbind(post.HPD,post.HPD.522)
post.HPD.len <- rbind(post.HPD.len,post.522.HPD.len)
post.MSE <- rbind(post.MSE,post.522.MSE)
post.mean <- cbind(big.rep,post.mean)
post.med <- cbind(big.rep,post.med)
post.sd <- cbind(big.rep,post.sd)
post.perc <- cbind(big.rep,post.perc)
post.perc.len <- cbind(big.rep,post.perc.len)
post.HPD <- cbind(big.rep,post.HPD)
post.HPD.len <- cbind(big.rep,post.HPD.len)
post.MSE <- cbind(big.rep,post.MSE)
comb.post.mean <- rbind(comb.post.mean,post.mean)
comb.post.med <- rbind(comb.post.med,post.med)
comb.post.sd <- rbind(comb.post.sd,post.sd)
```

```
        comb.post.perc <- rbind(comb.post.perc,post.perc)
        comb.post.perc.len <- rbind(comb.post.perc.len,post.perc.len)
        comb.post.HPD <- rbind(comb.post.HPD,post.HPD)
        comb.post.HPD.len <- rbind(comb.post.HPD.len,post.HPD.len)
        comb.post.MSE <- rbind(comb.post.MSE,post.MSE)
    }
    post.cov.perc.count <- rbind(cbind(rbind(cov.perc.521),0,0,0,0),cov.perc.522)
    post.cov.HPD.count <- rbind(cbind(rbind(cov.HPD.521),0,0,0,0),cov.HPD.522)
    post.cov.perc <- post.cov.perc.count/num.rep
    post.cov.HPD <- post.cov.HPD.count/num.rep
    post.cov.count <- rbind(post.cov.perc.count,post.cov.HPD.count)
    post.cov <- rbind(post.cov.perc,post.cov.HPD)
    rows.521 <- c(1,5,9,13)
    rows.522 <- c(2,6,10,14)
    mean.521 <- apply(comb.post.mean[rows.521,-1],2,mean)
    med.521 <- apply(comb.post.med[rows.521,-1],2,mean)
    stdev.521 <- apply(comb.post.sd[rows.521,-1],2,mean)
    mse.521 <- apply(comb.post.MSE[rows.521,-1],2,mean)
    mean.522 <- apply(comb.post.mean[rows.522,-1],2,mean)
    med.522 <- apply(comb.post.med[rows.522,-1],2,mean)
    stdev.522 <- apply(comb.post.sd[rows.522,-1],2,mean)
    mse.522 <- apply(comb.post.MSE[rows.522,-1],2,mean)
    comb.summary <- rbind(mean.521,med.521,stdev.521,mse.521,mean.522,med.522,stdev.522,mse.522)
    file.name <- paste(dir.name,'comb.summary.csv',sep='')
    write.csv(round(comb.summary,5),file=file.name)
}
}
}
```

A.4 Imputing all unknown observations

The following functions and program were used in Section 5.4.2.1.

```
Cal.Dirichlet.Stage <- function(Est1,Est2){
    pos.stages <- c(1:length(Est1))
    mean.est <- (Est1+Est2)/2
    sum.sq <- sum(mean.est^2)
    sum.sq.diff <- sum((Est1-Est2)^2)
    est.C <- (2*(1-sum.sq)/sum.sq.diff)
    prior.mult.par <- est.C*mean.est
    prior.mult.par[which(prior.mult.par == 0)] <- 1e-50
    prior.prob <- rdirichlet(1,prior.mult.par)
    ran.multi <- rmultinom(1,1,prior.prob)
    stage <- pos.stages[ran.multi == 1]
    return(stage)
}
```

```
Fill.Dirichlet.Stages <- function(pat.data,Pat.ID,Stage,Indicator,Est1,Est2){
    pat.ids <- unique(pat.data[,Pat.ID])
    new.data <- NA
    for (j in pat.ids) {
        temp.data <- pat.data[pat.data[,Pat.ID] == j,]
        num.obs <- dim(temp.data)[[1]]
        observed <- which(temp.data[,Indicator] > 0,0)
        first.obs <- observed[1]
        last.obs <- observed[length(observed)]
        for (i in (first.obs+1):num.obs) {
            if (temp.data[i,Indicator] == 0) {
                if (temp.data[i-1,Stage] == 1) {
                    new.stage <- Cal.Dirichlet.Stage(Est1[1,],Est2[1,])
                    if (new.stage != 3) temp.data[i,Stage] <- new.stage
```

```
                    }
                    else if (temp.data[i-1,Stage] == 2) temp.data[i,Stage] <- Cal.Dirichlet.Stage(Est1[2,],Est2[2,])
                    else if (temp.data[i-1,Stage] == 3) {
                        new.stage <- Cal.Dirichlet.Stage(Est1[3,],Est2[3,])
                        if (new.stage != 1) temp.data[i,Stage] <- new.stage
                    }
                }
            }
            rev.observed <- observed[length(observed):1]
            for (l in (1:(length(rev.observed)-1))) {
                begin.obs <- rev.observed[l]
                end.obs <- rev.observed[l+1] + 1
                if (begin.obs != end.obs) {
                    test.stage <- temp.data[begin.obs,Stage]
                    for (k in (begin.obs-1):end.obs) {
                    if (abs(temp.data[k,Stage] - test.stage) > 1)
                        temp.data[k,Stage] <- NA
                    else
                        break
                    }
                }
            }
            new.data <- rbind(new.data,temp.data)
        }
        return.data <- new.data[is.na(new.data[,Stage])==0,]
        return(return.data)
}
```

The following program code uses the above functions to imput all unknown observations in a multi-state data sets based on prior probability vectors. Data sets for different transition matrixes and data scenarios are generated, the Bayesian imputation is performed, a Markov

model is fitted and the posterior distributions are generated, summarised and the results collated.

```
sim.prior.doen <- c(1,2,3,4,5)
sim.scn.doen <- c(1,2,3,4)
sim.q.mat <- c(1,2,3,4,5,6)
for (sim.prior in sim.prior.doen){
for (sim.scn in sim.scn.doen){
for (q.mat in sim.q.mat) {
    top.dir.name <- paste('d:\\CJBMuller\\My Documents\\Navorsing\\PhD\\Multi-State Models\\
        Sagteware en Rekenaar Werk\\Finale Simulasie\\',sep='')
    next.dir.name <- paste('4 State\\Hfstk 6\\Covariates\\Q',q.mat,'.Sc',sim.scn,'.Pr',sim.prior,'.',sep='')
    dir.name <- paste(top.dir.name,next.dir.name,sep='')
    cov.effect.x <- c(-0.7,-0.7,-0.7,-0.7,-0.7,-0.7)
    cov.effect.y <- c(0.01,0.01,0.01,0.01,0.01,0.01)
    # Set type of model by the transition matrix
    if (q.mat == 1) {
        sim.qmatrix <- rbind(c(0, 0.5, 0),c(0.5, 0, 0.5),c(0, 0.5, 0))
        pop.par <- c(0.5,0.5,0.5,0.5,cov.effect.x)
    }
    if (q.mat == 2) {
        sim.qmatrix <- rbind(c(0, 0.25, 0),c(0.75, 0, 0.25),c(0, 0.75, 0))
        pop.par <- c(0.25,0.75,0.25,0.75,cov.effect.x)
    }
    if (q.mat == 3) {
        sim.qmatrix <- rbind(c(0, 0.75, 0),c(0.25, 0, 0.75),c(0, 0.25, 0))
        pop.par <- c(0.75,0.25,0.75,0.25,cov.effect.x)
    }
    if (q.mat == 4) {
        sim.qmatrix <- rbind(c(0, 0.5, 0, 0),c(0.5, 0, 0.5, 0),c(0, 0.5, 0, 0.5),c(0, 0, 0.5, 0))
        pop.par <- c(0.5,0.5,0.5,0.5,0.5,0.5,cov.effect.x,cov.effect.y)
```

```
}
if (q.mat == 5) {
    sim.qmatrix <- rbind(c(0, 0.25, 0, 0),c(0.75, 0, 0.25, 0),c(0, 0.75, 0, 0.25),c(0, 0, 0.75, 0))
    pop.par <- c(0.25,0.75,0.25,0.75,0.25,0.75,cov.effect.x,cov.effect.y)
}
if (q.mat == 6) {
    sim.qmatrix <- rbind(c(0, 0.75, 0, 0),c(0.25, 0, 0.75, 0),c(0, 0.25, 0, 0.75),c(0, 0, 0.25, 0))
    pop.par <- c(0.75,0.25,0.75,0.25,0.75,0.25,cov.effect.x,cov.effect.y)
}
# Set Data scenario
# Info on time frame 0 to 24 months, number of patients, max number of observations per patient
times <- seq(0,24,1)
if (sim.scn == 1) {
    num.pat <- 25
    missing.perc <- 0.1
}
if (sim.scn == 2) {
    num.pat <- 25
    missing.perc <- 0.5
}
if (sim.scn == 3) {
    num.pat <- 50
    missing.perc <- 0.1
}
if (sim.scn == 4) {
    num.pat <- 50
    missing.perc <- 0.5
}
num.obs.pat <- round(length(times)*(1-missing.perc),0)
# Prior
```

```
if (sim.prior == 1){
    est1 <- matrix(c(0.50,0.50,0,0, 1/3,1/3,1/3,0, 0,1/3,1/3,1/3, 0,0,0.50,0.50),
        nrow=4,ncol=4,byrow=T)
    est2 <- matrix(c(0.55,0.45,0,0, 0.30,0.40,0.30,0, 0,0.30,0.40,0.30, 0,0,0.45,0.55),
        nrow=4,ncol=4,byrow=T)
}
if (sim.prior == 2){
    est1 <- matrix(c(0.80,0.20,0,0, 0.10,0.80,0.10,0, 0,0.10,0.80,0.10, 0,0,0.20,0.80),
        nrow=4,ncol=4,byrow=T)
    est2 <- matrix(c(0.90,0.10,0,0, 0.05,0.90,0.05,0, 0,0.05,0.90,0.05, 0,0,0.10,0.90),
        nrow=4,ncol=4,byrow=T)
}
if (sim.prior == 3){
    est1 <- matrix(c(0.20,0.80,0,0, 0.40,0.20,0.40,0, 0,0.40,0.20,0.40, 0,0,0.80,0.20),
        nrow=4,ncol=4,byrow=T)
    est2 <- matrix(c(0.10,0.90,0,0, 0.45,0.10,0.45,0, 0,0.45,0.10,0.45, 0,0,0.90,0.10),
        nrow=4,ncol=4,byrow=T)
}
if (sim.prior == 4){
    est1 <- matrix(c(0.20,0.80,0,0, 0.10,0.10,0.80,0, 0,0.10,0.10,0.80, 0,0,0.20,0.80),
        nrow=4,ncol=4,byrow=T)
    est2 <- matrix(c(0.30,0.70,0,0, 0.15,0.15,0.70,0, 0,0.15,0.15,0.70, 0,0,0.30,0.70),
        nrow=4,ncol=4,byrow=T)
}
if (sim.prior == 5){
    est1 <- matrix(c(0.80,0.20,0,0, 0.80,0.10,0.10,0, 0,0.80,0.10,0.10, 0,0,0.80,0.20),
        nrow=4,ncol=4,byrow=T)
    est2 <- matrix(c(0.70,0.30,0,0, 0.70,0.15,0.15,0, 0,0.70,0.15,0.15, 0,0,0.70,0.30),
        nrow=4,ncol=4,byrow=T)
}
```

```
trans.names <- list("1->2","2->1","2->3","3->2","3->4","4->3","X 1->2","X 2->1","X 2->3","X 3->2",
    "X 3->4","X 4->3","Y 1->2","Y 2->1","Y 2->3","Y 3->2","Y 3->4","Y 4->3")
trans.names.int <- list("1->2 L","1->2 U","2->1 L","2->1 U","2->3 L","2->3 U","3->2 L","3->2 U","3->4 L",
    "3->4 U","4->3 L","4->3 U")
fin.row.names <- list("61","62")
# Create matrixes for final answers
comb.post.mean <- NULL
comb.post.med <- NULL
comb.post.sd <- NULL
comb.post.perc <- NULL
comb.post.perc.len <- NULL
comb.post.HPD <- NULL
comb.post.HPD.len <- NULL
comb.post.MSE <- NULL
cov.perc.61 <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
cov.HPD.61 <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
# Number of repitions for the type of data
num.rep <- 4
fill.rep <- 5000
for (big.rep in 1:num.rep) {
    # Simulate dataset based on tran matrix
    sim.data <- simulate.data.cov.two(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,vis.dif=1,
        cov.effect.x,cov.effect.y)
    sim.data.big <- Create.Sim.Data.Cov(sim.data,Pat.ID=1,Visit=6,Time=2,Stage=3,Indicator=7)
    # Fit ML model to simulated data
    sim.model <- msm(state ~time,subject=subject,data=sim.data,qmatrix=sim.qmatrix,covariates=~x + y)
    sim.model.est <- qmatrix.msm(sim.model, covariates = list(x=0,y=0))[[1]]
    # Sec 6.1: Fill data Dirichlet from two prior vectors
    post.61 <- NULL
    se.not.found <- 0
```

```
i <- 1
while (i <= fill.rep) {
        i <- i + 1
        sim.data.fill <- Fill.Dirichlet.Stages(sim.data.big,Pat.ID=1,Stage=3,Indicator=7,est1,est2)
        fill.model <- msm(state ~time,subject=subject,data=sim.data.fill,qmatrix=sim.model.est,
                covariates=~x + y)
        fill.model.est <- qmatrix.msm(fill.model, covariates = list(x=0,y=0))[[1]]
        fill.model.est <- c(fill.model.est[1,c(2)],fill.model.est[2,c(1,3)],fill.model.est[3,c(2,4)],
                fill.model.est[4,c(3)])
        fill.model.x.est <- c(fill.model$Qmatrices$x[1,c(2)],fill.model$Qmatrices$x[2,c(1,3)],fill.model$Qmatrices
                $x[3,c(2,4)],fill.model$Qmatrices$x[4,c(3)])
        fill.model.y.est <- c(fill.model$Qmatrices$y[1,c(2)],fill.model$Qmatrices$y[2,c(1,3)],fill.model$Qmatrices
                $y[3,c(2,4)],fill.model$Qmatrices$y[4,c(3)])
        if (fill.model$foundse) {
                post.61 <- rbind(post.61,cbind(t(fill.model.est),t(fill.model.x.est),t(fill.model.y.est)))
        }
else {
        se.not.found <- se.not.found + 1
        i <- i - 1
        }
}
dimnames(post.61)[[2]] <- trans.names
post.61.mean <- apply(post.61,2,mean)
post.61.med <- apply(post.61,2,median)
post.61.sd <- apply(post.61,2,sd)
post.61.perc.l <- apply(post.61,2,quantile,probs=c(0.025))
post.61.perc.u <- apply(post.61,2,quantile,probs=c(0.975))
post.61.perc.len <- post.61.perc.u - post.61.perc.l
post.61.HPD <- HPDinterval(as.mcmc(post.61))
post.61.HPD.l <- post.61.HPD[,1]
```

```
post.61.HPD.u <- post.61.HPD[,2]
post.61.HPD.len <- post.61.HPD[,2] - post.61.HPD[,1]
post.61.MSE <- cal.MSE(post.61,pop.par)
cov.perc.61 <- cov.perc.61 + ((post.61.perc.l <= pop.par) & (pop.par <= post.61.perc.u))
cov.HPD.61 <- cov.HPD.61 + ((post.61.HPD.l <= pop.par) & (pop.par <= post.61.HPD.u))
# Posterior Summary of techniques
post.mean <- post.61.mean
post.med <- post.61.med
post.sd <- post.61.sd
post.perc.l <- post.61.perc.l
post.perc.u <- post.61.perc.u
post.perc.len <- post.61.perc.len
post.HPD.l <- post.61.HPD.l
post.HPD.u <- post.61.HPD.u
post.HPD.len <- post.61.HPD.len
post.MSE <- post.61.MSE
post.mean <- cbind(big.rep,post.mean)
post.med <- cbind(big.rep,post.med)
post.sd <- cbind(big.rep,post.sd)
post.perc <- cbind(big.rep,post.perc.l[,1],post.perc.u[,1],post.perc.l[,2],post.perc.u[,2],post.perc.l[,3],post.perc.u[,3]
      ,post.perc.l[,4],post.perc.u[,4],post.perc.l[,5],post.perc.u[,5],post.perc.l[,6],post.perc.u[,6])
post.perc.len <- cbind(big.rep,post.perc.len)
post.HPD <- cbind(big.rep,post.HPD.l[,1],post.HPD.u[,1],post.HPD.l[,2],post.HPD.u[,2],post.HPD.l[,3],
      post.HPD.u[,3],post.HPD.l[,4],post.HPD.u[,4],post.HPD.l[,5],post.HPD.u[,5],post.HPD.l[,6],post.HPD.u[,6])
post.HPD.len <- cbind(big.rep,post.HPD.len)
post.MSE <- cbind(big.rep,post.MSE)
comb.post.mean <- rbind(comb.post.mean,post.mean)
comb.post.med <- rbind(comb.post.med,post.med)
comb.post.sd <- rbind(comb.post.sd,post.sd)
comb.post.perc <- rbind(comb.post.perc,post.perc)
```

```
        comb.post.perc.len <- rbind(comb.post.perc.len,post.perc.len)
        comb.post.HPD <- rbind(comb.post.HPD,post.HPD)
        comb.post.HPD.len <- rbind(comb.post.HPD.len,post.HPD.len)
        comb.post.MSE <- rbind(comb.post.MSE,post.MSE)
    }
    post.cov.perc.count <- cov.perc.61
    post.cov.HPD.count <- cov.HPD.61
    post.cov.perc <- post.cov.perc.count/num.rep
    post.cov.HPD <- post.cov.HPD.count/num.rep
    post.cov.count <- rbind(post.cov.perc.count,post.cov.HPD.count)
    post.cov <- rbind(post.cov.perc,post.cov.HPD)
    mean.61 <- apply(comb.post.mean[,-1],2,mean)
    med.61 <- apply(comb.post.med[,-1],2,mean)
    stdev.61 <- apply(comb.post.sd[,-1],2,mean)
    mse.61 <- apply(comb.post.MSE[,-1],2,mean)
    comb.summary.61 <- rbind(mean.61,med.61,stdev.61,mse.61)
    file.name <- paste(dir.name,'comb.summary.csv',sep='')
    write.csv(round(comb.summary.61,5),file=file.name)
}
}
}
```

A.5 Estimating and imputing the transition point

The following functions and program were used in Section 5.4.2.2.

```
Cal.Dirichlet.Trans.Point.Betas.Her <- function(times,info,beta.pos,beta.her) {
    num.int <- info['End Visit'] - info['Begin Visit']
    end.time <- times[info['End Visit']]
    elaps.times <- c(0)
    for (i in (info['Begin Visit']+1):info['End Visit']) {
        elaps.times <- c(elaps.times,times[i] - times[info['Begin Visit']])
```

```
}
elaps.times.max <- elaps.times[length(elaps.times)]
c.weight <- 5
theta <- 0
if (info['Begin Stage'] < info['End Stage']) change <- paste(info['Begin Stage'],info['Begin Stage']+1,sep='')
if (info['Begin Stage'] > info['End Stage']) change <- paste(info['Begin Stage'],info['Begin Stage']-1,sep='')
if (change == "12") lambda.parm <- 1
if (change == "21") lambda.parm <- 2
if (change == "23") lambda.parm <- 3
if (change == "32") lambda.parm <- 4
if (change == "34") lambda.parm <- 5
if (change == "43") lambda.parm <- 6
if (change == "45") lambda.parm <- 7
beta.use <- beta.pos[lambda.parm,]
trans.func <- function(t,pr.loc,pr.shape){1-exp(-(t/pr.loc)^pr.shape)}
prior.loc <- elaps.times.max
for (i in (1:beta.her)){
    prior.k <- abs(rnorm(1,beta.use[1],beta.use[2]))
    alfa <- c()
    if (num.int != 0){
        for (i in 1:(num.int)) {
            alfa.temp <- abs(c.weight*(trans.func(elaps.times[i],prior.loc,prior.k) -
                trans.func(elaps.times[(i+1)],prior.loc,prior.k)))
            alfa <- c(alfa,alfa.temp)
        }
    }
    else {
        alfa <- 2
    }
    theta <- theta + rdirichlet(1,abs(alfa))
```

```
}
gem.theta <- theta/beta.her
prob.NA <- sum(is.na(gem.theta))
if (prob.NA > 0) {
        aantal.theta <- length(gem.theta)
        if (aantal.theta > 1) {
                for (h in 1:aantal.theta){
                        if (is.na(gem.theta[h])) {
                                gem.theta[h] <- 0
                        }
                }
                if (sum(gem.theta) == 0) {
                        gem.theta[1] <- 1
                }
        }
        if (aantal.theta == 1) {
                gem.theta <- 1
        }
}
ran.multi <- rmultinom(1,1,gem.theta)
trans.time <- which(ran.multi == 1)
if (info['Begin Stage'] < info['End Stage']) new.stage <- info['Begin Stage']+1
if (info['Begin Stage'] > info['End Stage']) new.stage <- info['Begin Stage']-1
answer <- matrix(nrow=1,ncol=2)
answer[1,1] <- info['Begin Visit'] + trans.time
answer[1,2] <- new.stage
if (info["Begin Visit"] == info["End Visit"]){
        answer[1,1] <- info["Begin Visit"]
}
return(answer)
```

```
}

Dirichlet.Trans.Point.Betas.Her <- function(big.data,Pat.ID,Stage,Time,Indicator,beta.pos,beta.her) {
    update.data <- big.data
    pat.ids <- unique(big.data[,Pat.ID])
    for (i in pat.ids) {
        in.pat <- big.data[big.data[,Pat.ID]==i,]
        visit.times <- in.pat[,Time]
        num.visits <- length(visit.times)
        visit.id <- c(1:num.visits)
        obs.visit <- visit.id[in.pat[,Indicator] == 1]
        obs.stages <- in.pat[obs.visit,Stage]
        obs.times <- in.pat[obs.visit,Time]
        num.obs <- length(obs.stages)
        start.change.int <- c()
        for (j in 1:(num.obs-1)){
            if (obs.stages[j+1] != obs.stages[j])
                start.change.int <- c(start.change.int,j)
        }
        num.changes <- length(start.change.int)
        change.int <- matrix(ncol=9,nrow=num.changes)
        dimnames(change.int) <-list(c(),c('Begin Visit','End Visit','Begin Time','End Time','Begin Stage',
            'End Stage','Num Stage Changes','Direction','Order'))
        temp.change.int <- NULL
        if (num.changes > 0) {
            for (k in 1:num.changes) {
                change.int[k,'Begin Visit'] <- obs.visit[start.change.int[k]]
                change.int[k,'End Visit'] <- obs.visit[start.change.int[k]+1]
                change.int[k,'Begin Time'] <- obs.times[start.change.int[k]]
                change.int[k,'End Time'] <- obs.times[start.change.int[k]+1]
                change.int[k,'Begin Stage'] <- obs.stages[start.change.int[k]]
```

```
        change.int[k,'End Stage'] <- obs.stages[start.change.int[k]+1]
        change.int[k,'Num Stage Changes'] <- change.int[k,5] - change.int[k,6]
        if (change.int[k,'Num Stage Changes'] > 0) change.int[k,'Direction'] <- -1
        if (change.int[k,'Num Stage Changes'] < 0) change.int[k,'Direction'] <- 1
        change.int[k,'Num Stage Changes'] <- abs(change.int[k,'Num Stage Changes'])
        change.int[k,'Order'] <- k
}
test.num.changes <- sum(change.int[,'Num Stage Changes'])
if (test.num.changes != num.changes) {
        for (l in 1:num.changes) {
                new.change.point.temp <- NULL
                if (change.int[l,'Num Stage Changes'] == 1){
                        temp.change.int <- rbind(temp.change.int,change.int[l,])
                }
                else {
                        for (m in 1:(change.int[l,'Num Stage Changes']-1)) {
                                new.change.point.temp <- rbind(new.change.point.temp,change.int[l,])
                        }
                        for (m in 1:(change.int[l,'Num Stage Changes']-1)) {
                                new.change.point.temp[m,'Begin Visit'] <- NA
                                new.change.point.temp[m,'Begin Time'] <- NA
                                if (new.change.point.temp[m,'Direction'] == -1){
                                        new.change.point.temp[m,'Begin Stage'] <- change.int[l,"Begin Stage"]
                                                - m
                                        new.change.point.temp[m,'End Stage'] <- new.change.point.temp
                                                [m,"Begin Stage"] - 1
                                }
                                if (new.change.point.temp[m,'Direction'] == 1) {
                                        new.change.point.temp[m,'Begin Stage'] <- change.int[l,"Begin Stage"]
                                                + m
```

```
                    new.change.point.temp[m,'End Stage'] <- new.change.point.temp
                        [m,"Begin Stage"] + 1
                }
                new.change.point.temp[m,'Num Stage Changes'] <-
                    change.int[l,"Num Stage Changes"] - m
                new.change.point.temp[m,'Order'] <- new.change.point.temp[m,'Order']
                    + 0.5
            }
            if (change.int[l,'Direction'] == -1) change.int[l,'End Stage'] <-
                change.int[l,'Begin Stage'] - 1
            if (change.int[l,'Direction'] == 1) change.int[l,'End Stage'] <-
                change.int[l,'Begin Stage'] + 1
            temp.change.int <- rbind(temp.change.int,change.int[l,],new.change.point.temp)
        }
    }
    change.int <- temp.change.int
    num.changes <- test.num.changes
}
pred.change <- matrix(ncol=2,nrow=num.changes)
for (k in 1:num.changes) {
    if (change.int[k,'Num Stage Changes'] == 1)
    pred.change[k,] <- Cal.Dirichlet.Trans.Point.Betas.Her(visit.times,change.int[k,],beta.pos,
        beta.her)
    if (change.int[k,'Num Stage Changes'] > 1) {
        temp.change <- Cal.Dirichlet.Trans.Point.Betas.Her(visit.times,change.int[k,],
            beta.pos,beta.her)
        change.int[k+1,'Begin Visit'] <- temp.change[1]
        change.int[k+1,'Begin Time'] <- visit.times[temp.change[1]]
        pred.change[k,] <- temp.change
    }
```

```
                }
                dimnames(pred.change)[[2]] <- c("New Visit","New Stage")
                update.data[update.data[,Pat.ID]==i,][pred.change[,"New Visit"],Stage] <- pred.change[,"New Stage"]
            }
        }
        return.data <- update.data[is.na(update.data[,Stage])==0,]
        return(return.data)
}
```

The following program code uses the above functions to imput the transition rate between two known observations in a multi-state data sets based on prior transition functions. Data sets for different transition matrixes and data scenarios are generated, the Bayesian imputation is performed, a Markov model is fitted and the posterior distributions are generated, summarised and the results collated.

```
sim.prior.doen <- c(1,2,3,4,5)
sim.scn.doen <- c(1,2,3,4)
sim.q.mat <- c(1,2,3)
for (sim.prior in sim.prior.doen){
for (sim.scn in sim.scn.doen){
for (q.mat in sim.q.mat) {
        top.dir.name <- paste('d:\\CJBMuller\\My Documents\\Navorsing\\PhD\\Multi-State Models\\
                Sagteware en Rekenaar Werk\\Finale Simulasie\\',sep='')
        next.dir.name <- paste('4 State\\Hfstk 6\\Covariates\\Q',q.mat,'.Sc',sim.scn,'.Pr',sim.prior,'.',sep='')
        dir.name <- paste(top.dir.name,next.dir.name,sep='')
        cov.effect.x <- c(-0.7,-0.7,-0.7,-0.7,-0.7,-0.7)
        cov.effect.y <- c(0.01,0.01,0.01,0.01,0.01,0.01)
        # Set type of model by the transition matrix
        if (q.mat == 1) {
            sim.qmatrix <- rbind(c(0, 0.5, 0),c(0.5, 0, 0.5),c(0, 0.5, 0))
            pop.par <- c(0.5,0.5,0.5,0.5,cov.effect.x)
        }
```

```
if (q.mat == 2) {
    sim.qmatrix <- rbind(c(0, 0.25, 0),c(0.75, 0, 0.25),c(0, 0.75, 0))
    pop.par <- c(0.25,0.75,0.25,0.75,cov.effect.x)
}
if (q.mat == 3) {
    sim.qmatrix <- rbind(c(0, 0.75, 0),c(0.25, 0, 0.75),c(0, 0.25, 0))
    pop.par <- c(0.75,0.25,0.75,0.25,cov.effect.x)
}
if (q.mat == 4) {
    sim.qmatrix <- rbind(c(0, 0.5, 0, 0),c(0.5, 0, 0.5, 0),c(0, 0.5, 0, 0.5),c(0, 0, 0.5, 0))
    pop.par <- c(0.5,0.5,0.5,0.5,0.5,0.5,cov.effect.x,cov.effect.y)
}
if (q.mat == 5) {
    sim.qmatrix <- rbind(c(0, 0.25, 0, 0),c(0.75, 0, 0.25, 0),c(0, 0.75, 0, 0.25),c(0, 0, 0.75, 0))
    pop.par <- c(0.25,0.75,0.25,0.75,0.25,0.75,cov.effect.x,cov.effect.y)
}
if (q.mat == 6) {
    sim.qmatrix <- rbind(c(0, 0.75, 0, 0),c(0.25, 0, 0.75, 0),c(0, 0.25, 0, 0.75),c(0, 0, 0.25, 0))
    pop.par <- c(0.75,0.25,0.75,0.25,0.75,0.25,cov.effect.x,cov.effect.y)
}
# Set Data scenario
# Info on time frame 0 to 24 months, number of patients, max number of observations per patient
times <- seq(0,24,1)
if (sim.scn == 1) {
    num.pat <- 25
    missing.perc <- 0.1
}
if (sim.scn == 2) {
    num.pat <- 25
    missing.perc <- 0.5
```

```
}
if (sim.scn == 3) {
        num.pat <- 50
        missing.perc <- 0.1
}
if (sim.scn == 4) {
        num.pat <- 50
        missing.perc <- 0.5
}
num.obs.pat <- round(length(times)*(1-missing.perc),0)
# Prior
# Set the shape parameter of the Weibull transition function
possible.shape <- rbind(c(0.3,0.05),c(1.1,0.05),c(8,0.75))
if (sim.prior == 1){
        shape.pos <- rbind(possible.shape[1,],possible.shape[1,],possible.shape[1,],possible.shape[1,],
                possible.shape[1,],possible.shape[1,])
}
if (sim.prior == 2){
        shape.pos <- rbind(possible.shape[2,],possible.shape[2,],possible.shape[2,],possible.shape[2,],
                possible.shape[2,],possible.shape[2,])
}
if (sim.prior == 3){
        shape.pos <- rbind(possible.shape[3,],possible.shape[3,],possible.shape[3,],possible.shape[3,],
                possible.shape[3,],possible.shape[3,])
}
if (sim.prior == 4){
        shape.pos <- rbind(possible.shape[1,],possible.shape[3,],possible.shape[1,],possible.shape[3,],
                possible.shape[1,],possible.shape[3,])
}
if (sim.prior == 5){
```

```
        shape.pos <- rbind(possible.shape[3,],possible.shape[1,],possible.shape[3,],possible.shape[1,],
            possible.shape[3,],possible.shape[1,])
}
trans.names <- list("1->2","2->1","2->3","3->2","3->4","4->3","X 1->2","X 2->1","X 2->3","X 3->2",
        "X 3->4","X 4->3","Y 1->2","Y 2->1","Y 2->3","Y 3->2","Y 3->4","Y 4->3")
trans.names.int <- list("1->2 L","1->2 U","2->1 L","2->1 U","2->3 L","2->3 U","3->2 L","3->2 U","3->4 L",
        "3->4 U","4->3 L","4->3 U")
fin.row.names <- list("61","62")
# Create matrixes for final answers
comb.post.mean <- NULL
comb.post.med <- NULL
comb.post.sd <- NULL
comb.post.perc <- NULL
comb.post.perc.len <- NULL
comb.post.HPD <- NULL
comb.post.HPD.len <- NULL
comb.post.MSE <- NULL
cov.perc.62 <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
cov.HPD.62 <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
# Number of repitions for the type of data
num.rep <- 4
fill.rep <- 5000
tp.rep <- 5000
shape.in.tp.rep <- 100
for (big.rep in 1:num.rep) {
        # Simulate dataset based on tran matrix
        sim.data <- simulate.data.cov.two(num.pat,times,num.obs.pat,sim.qmatrix=sim.qmatrix,vis.dif=1,
            cov.effect.x,cov.effect.y)
        sim.data.big <- Create.Sim.Data.Cov(sim.data,Pat.ID=1,Visit=6,Time=2,Stage=3,Indicator=7)
        # Fit ML model to simulated data
```

```
sim.model <- msm(state ~time,subject=subject,data=sim.data,qmatrix=sim.qmatrix,covariates=~x + y)
sim.model.est <- qmatrix.msm(sim.model, covariates = list(x=0,y=0))[[1]]
# Sec 6.2: Calculate trans point
post.62 <- NULL
i <- 1
se.not.found <- 0
while (i <= tp.rep){
    i <- i + 1
    sim.data.tp <- Dirichlet.Trans.Point.Betas.Her(sim.data.big,Pat.ID=1,Stage=3,Time=2,Indicator=7,
    beta.pos=shape.pos,beta.her=shape.in.tp.rep)
    tp.model <- msm(state ~time,subject=subject,data=sim.data.tp,qmatrix=sim.model.est,
        covariates=~x + y)
    tp.model.est <- qmatrix.msm(tp.model, covariates = list(x=0,y=0))[[1]]
    tp.model.est <- c(tp.model.est[1,c(2)],tp.model.est[2,c(1,3)],tp.model.est[3,c(2,4)],
        tp.model.est[4,c(3)])
    tp.model.x.est <- c(tp.model$Qmatrices$x[1,c(2)],tp.model$Qmatrices$x[2,c(1,3)],tp.model$Qmatrices
        $x[3,c(2,4)],tp.model$Qmatrices$x[4,c(3)])
    tp.model.y.est <- c(tp.model$Qmatrices$y[1,c(2)],tp.model$Qmatrices$y[2,c(1,3)],tp.model$Qmatrices
        $y[3,c(2,4)],tp.model$Qmatrices$x[4,c(3)])
    if (tp.model$foundse) {
        post.62 <- rbind(post.62,cbind(t(tp.model.est),t(tp.model.x.est),t(tp.model.y.est)))
}
else {
    se.not.found <- se.not.found + 1
    i <- i - 1
    }
}
dimnames(post.62)[[2]] <- trans.names
post.62.mean <- apply(post.62,2,mean)
post.62.med <- apply(post.62,2,median)
```

```
post.62.sd <- apply(post.62,2,sd)
post.62.perc.l <- apply(post.62,2,quantile,probs=c(0.025))
post.62.perc.u <- apply(post.62,2,quantile,probs=c(0.975))
post.62.perc.len <- post.62.perc.u - post.62.perc.l
post.62.HPD <- HPDinterval(as.mcmc(post.62))
post.62.HPD.l <- post.62.HPD[,1]
post.62.HPD.u <- post.62.HPD[,2]
post.62.HPD.len <- post.62.HPD[,2] - post.62.HPD[,1]
post.62.MSE <- cal.MSE(post.62,pop.par)
cov.perc.62 <- cov.perc.62 + ((post.62.perc.l <= pop.par) & (pop.par <= post.62.perc.u))
cov.HPD.62 <- cov.HPD.62 + ((post.62.HPD.l <= pop.par) & (pop.par <= post.62.HPD.u))
# Posterior Summary of techniques
post.mean <- post.62.mean
post.med <- post.62.med
post.sd <- post.62.sd
post.perc.l <- post.62.perc.l
post.perc.u <- post.62.perc.u
post.perc.len <- post.62.perc.len
post.HPD.l <- post.62.HPD.l
post.HPD.u <- post.62.HPD.u
post.HPD.len <- post.62.HPD.len
post.MSE <- post.62.MSE
post.mean <- cbind(big.rep,post.mean)
post.med <- cbind(big.rep,post.med)
post.sd <- cbind(big.rep,post.sd)
post.perc <- cbind(big.rep,post.perc.l[,1],post.perc.u[,1],post.perc.l[,2],post.perc.u[,2],post.perc.l[,3],
      post.perc.u[,3],post.perc.l[,4],post.perc.u[,4],post.perc.l[,5],post.perc.u[,5],post.perc.l[,6],post.perc.u[,6])
post.perc.len <- cbind(big.rep,post.perc.len)
post.HPD <- cbind(big.rep,post.HPD.l[,1],post.HPD.u[,1],post.HPD.l[,2],post.HPD.u[,2],post.HPD.l[,3],
      post.HPD.u[,3],post.HPD.l[,4],post.HPD.u[,4],post.HPD.l[,5],post.HPD.u[,5],post.HPD.l[,6],post.HPD.u[,6])
```

```
        post.HPD.len <- cbind(big.rep,post.HPD.len)
        post.MSE <- cbind(big.rep,post.MSE)
        comb.post.mean <- rbind(comb.post.mean,post.mean)
        comb.post.med <- rbind(comb.post.med,post.med)
        comb.post.sd <- rbind(comb.post.sd,post.sd)
        comb.post.perc <- rbind(comb.post.perc,post.perc)
        comb.post.perc.len <- rbind(comb.post.perc.len,post.perc.len)
        comb.post.HPD <- rbind(comb.post.HPD,post.HPD)
        comb.post.HPD.len <- rbind(comb.post.HPD.len,post.HPD.len)
        comb.post.MSE <- rbind(comb.post.MSE,post.MSE)
    }
    post.cov.perc.count <- cov.perc.62
    post.cov.HPD.count <- cov.HPD.62
    post.cov.perc <- post.cov.perc.count/num.rep
    post.cov.HPD <- post.cov.HPD.count/num.rep
    post.cov.count <- rbind(post.cov.perc.count,post.cov.HPD.count)
    post.cov <- rbind(post.cov.perc,post.cov.HPD)
    mean.62 <- apply(comb.post.mean[,-1],2,mean)
    med.62 <- apply(comb.post.med[,-1],2,mean)
    stdev.62 <- apply(comb.post.sd[,-1],2,mean)
    mse.62 <- apply(comb.post.MSE[,-1],2,mean)
    comb.summary.62 <- rbind(mean.62,med.62,stdev.62,mse.62)
    file.name <- paste(dir.name,'comb.summary.csv',sep='')
    write.csv(round(comb.summary.62,5),file=file.name)
}
}
}
```

References

Agresti, A. (2002). *Categorical Data Analysis (2^{nd} ed)*, New Yersey: Wiley-Interscience.

Aguirre-Hernández, R. and Farewell, V.T. (2002). A Pearson-type goodness-of-fit test for stationary and time-continuous Markov regression models. *Statistics in Medicine*, **21**, 1899-1911.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding (1993). *Statistical Models Based on Counting Processes (2^{nd} ed),* Springer Series in Statistics, New York: Springer.

Andersen, P.K., Abildstrom, S.Z. and Rosthoj, S. (2002). Competing Risks as a Multi-State Model. *Statistical Methods in Medical Research*, **11**, 203-215.

Berger, J.O. and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of the American Statistical Association*, **84**, 200-207.

Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society: Series B*, **41**, 113–147.

Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory*, New York: Wiley.

Box, G.E.P. and Tiao, G.C. (1992). *Bayesian Inference in Statistical Analysis*, New York: Wiley.

Casella, G. and Berger, R.L. (2001). *Statistical Inference (2^{nd} ed)*, Belmont: Duxbury Advanced Series.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167-174.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**, 327-336.

Congdon, P. (2002). *Bayesian statistical modelling*, New York: Wiley.

Data, G.S. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*, New York: Springer.

Data, G.S. and Sweeting, T.J. (2005). Probability matching priors. In *Handbook of Statistics*, **25**, *Bayesian Thinking: Modeling and Computation*, Dey, D. and Rao, C.R. (eds.), 91-114, Amsterdam: Elsevier.

De Groot, M.H. (1970). *Optimal Statistical Decisions*, New Jersey: John Wiley & Sons.

de Wreede, L.C., Fiocco, M. and Putter, H. (2010). The mstate Package for Estimation and Prediction in Non- and Semi-Parametric Multi-State and Competing Risks Models. *Computer Methods and Programs in Biomedicine*, **99**, 261-274.

Faddy, M.J. (1976). A note on the general time-dependent stochastic compartmental model. *Biometrics*, **32**, 443-448.

Ferguson, T.S. (1973). A Bayesian Analysis of some Nonparametric Problems. *The Annals of Statistics*, **1**, 209-230.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398-409.

Gentleman, R.C., Lawless, J.F., Lindsey, J.C. and Yan, P. (1994). Multi-State Markov Models for Analysing Incomplete Disease History Data with Illustrations for HIV Disease. *Statistics in Medicine*, **13**, 805-821.

Gill, J. (2008). *Bayesian Methods A Social and Behavioral Sciences Approach (2^{nd} ed)*, Boca Raton: Chapam & Hall/CRC.

Hamada, M.S., Wilson, A.G., Reese, C.S. and Martz, H.F. (2008). *Bayesian Reliability*, New York: Springer.

Hjort, N.L., Holmes, C., Müller, P. and Walker, S.G. editors (2010). *Bayesian Nonparametrics.* Cambridge Series in Statistical and Probabilistic Mathematics, number **28**. Cambridge: Cambridge University Press.

Hollander, M. and Proschan, F. (1976). Testing to Determine the Underlying Distribution Using Randomly Censored Data. *Biometrics*, **35**, 393-401.

Howard, G., Coffey, C.S., and Cutter, G.R. (2005). Is Bayesian Analysis Ready for Use in Phase III Randomized Clinical Trials. *Stroke*, 36, 1622-1623.

Hsiao, C. (2003). *Analysis of Panel Data*, Cambridge: Cambridge University Press

Hyndman, R.J. (1996). Computing and Graphing Highest Density Regions. *The American Statistician*, **50**, 120-126.

Ibrahim, J.G., Chen, M.-H. and Sinha, D. (2001). *Bayesian Survival Analysis*, New York: Springer.

Jackson, C.H., Sharples, L.D., Thompson, S.G., Duffy, S.W. and Couto, E. (2003). Multistate Markov Models for Disease Progression with Classification Error, *The Statistician*, **52**, 193-209.

Jackson, C.H (2005). *Multi-state modelling with R: the msm package.* R msm Package Software Manual (http://cran.r-project.org/, Downloaded: 12/12/2011).

Jackson, C.H. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, **38**, 8.

Jaynes, E.T. (1980). Marginalization and Prior Probabilities. In *Bayesian Analysis in Econometrics and Statistics*, A. Zellner (ed.). Amsterdam: North Holland.

Kalbfleisch, J.D. and Lawless, J.F. (1985). The Analysis of Panel Data under a Markov Assumption. *Journal of the American Statistical Association*, **80**, 863-871.

Kass, R.E. and Wasserman, L. (1996). Formal rules of selecting prior distributions: a review and annotated bibliography. *Journal of the American Statistical Association*, **91**, 343-1370.

Kay, R. (1986). A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies. *Biometrics*, **42**, 855-865.

Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis Techniques for Censored and Truncated Data (2^{nd} Edition)*, New York: Springer.

Longini Jr., I.M., Clark, W.S., Byers, R.H., Ward, J.W., Darrow, W.W., Lemp, G.F. and Hethcote, H.W. (1989). Statistical Analysis of the Stages of HIV Infection Using a Markov Model. *Statistics in Medicine*, **8**, 831-843.

Marshall, G. and Jones, R.H. (1995). Multi-State Models and Diabetic Retinopathy. *Statistics in Medicine*, **14**, 1975-1983.

Martz, H.F. and Waller R.A. (1982). *Bayesian Reliability Analysis*, New York: Wiley.

Meira-Machado, L., Cadarso-Suárez, C. and de Uña-Álvarez, J. (2008). Inference in the Progressive Three-state Model. *International Journal of Mathematical Models and Methods in Applied Sciences*, **3**, 447-454.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C. and Anderson, P.K. (2009). Multi-state Models for the Analysis of Time-to-Event Data. *Statistical Methods in Medical Research*, **18**, 195-222.

Mostert, P.J., Bekker, A. and Yadavalli, V.S.S. (2004). A two-unit intermittently used cold standby system from a Bayesian viewpoint assuming a Dirichlet prior. *Pakistan Journal of Statistics*, **20**(2), 223-229.

Moyé, L.A. (2008). *Elementary Bayesian Biostatistics*, Boca Raton: Chapman & Hall/CRC.

Pérez-Ocón, R., Ruiz-Castro, J.E. and Gámiz-Pérez, M.L. (1998). A Multivariate Model to Measure the Effect of Treatments in Survival to Breast Cancer. *Biometrical Journal*, **40**, 703-715.

Pérez-Ocón, R., Ruiz-Castro, J.E. and Gámiz-Pérez, M.L. (2001). Non-homogeneous Markov models in the analysis of survival after breast cancer. *Journal of the Royal Statistical Society: Series C*, **50**, 111-124.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, Boston: Harvard University Press.

Robert, C. P. (2001). *The Bayesian Choice (2^{nd} ed)*, New York: Springer.

Ross, M. S. (2003). *Introduction to Probability Models, 8^{th} Edition.* San Diego: Academic Press.

Sharples, L.D., Jackson, C.H., Parameshwar, J., Wallwork, J, and Large, S.R. (2003). Diagnostic Accuracy of Coronary Angiography and Risk Factors for Post-Heart-Transplant Cardiac Allograft Vasculopathy. *Transplantation*, **76**(4), 679-682.

Sinha, D. and Dey, D.K. (1997). Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association*, **92**, 1195-1212.

Spiegelhalter, D.J., Adrams, K.R. and Myles, J.P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, New York: Wiley.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B*, **64**, 583-639.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations. *Journal of the American Statistical Association*, **71**, 897-902.

Sweeting, T.J. (2005). On the implementation of local probability matching priors for interest parameters. *Biometrika*, **92**, 47-57.

Titman, A.C. and Sharples, L.D. (2007). A general goodness-of-fit test for Markov and hidden Markov models. *Statistics in Medicine*, **27**, 2177-2195.

Titman, A.C. and Sharples, L.D. (2010). Model diagnostics for multi-state models. *Statistical Methods in Medical Research*, **19**, 621-651.

Welch, B.L. and Peers, H.W. (1963). On formulae for confidence points based on intergrals of weighted likelihoods. Journal of the Royal Statistical Society: Series B, **35**, 318-329. Cited in Data, G.S. and Sweeting, T.J. (2005). Probability matching priors. In *Handbook of Statistics*, **25**, *Bayesian Thinking: Modeling and Computation*, Dey, D. and Rao, C.R. (eds.), 91-114, Amsterdam: Elsevier.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* New York: Wiley.

Zellner, A. (1996). Models, prior information, and Bayesian analysis. *Journal of Econometrics*, **78**, 51-68.

Zellner, A. (1998). Past and Recent Results on Maximal Data Information Priors. *Journal of Statistical Reseach*, **32**, 1-22.

Zellner, A. and Min, C-K. (1995). Gibbs Sampler Convergence Criteria. *Journal of the American Statistical Association*, **90**, 921-927.