Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Julie Berg

Multistate models in survival analysis using INLA, applied on data for resuscitation after cardiac arrest

Master's thesis in Mathematical Sciences
Supervisor: Sara Martino
Co-supervisor: Turid Follestad, Eirik Skogvoll
February 2022

NTNU
Norwegian University of Science and Technology

Julie Berg

Multistate models in survival analysis using INLA, applied on data for resuscitation after cardiac arrest

Master's thesis in Mathematical Sciences
Supervisor: Sara Martino
Co-supervisor: Turid Follestad, Eirik Skogvoll
February 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Abstract

Survival analysis is the study of data that describes the time to a particular event. Including several events are often needed to create a more accurate model of the world and these models are referred to as multistate models. As survival data often includes censored data, the hazard function plays a central role in survival analysis. Heterogeneity in the distributions can be accounted for by including covariates in a model for the hazard, and we wish to estimate the effects of these covariates. In addition, we use Weibull and exponential baseline hazards. We use a Bayesian analysis approach for our multistate models, which can be presented as latent Gaussian models (LGMs) by assigning Gaussian priors to the latent field. The inferential tool named integrated nested Laplace approximations (INLA) is used for inference. INLA can be adapted and applied to complex multistate models, making Bayesian analysis fast and accurate.

Sammendrag

Levetidsanalyse er studiet av data som beskriver tiden til en bestemt hendelse. Inkludering av flere hendelser er ofte nødvendig for å lage en mer nøyaktig modell av verden, og disse modellene blir referert til som multistate-modeller. Siden levetidsdata ofte inkluderer sensurerte data, spiller farefunksjonen en sentral rolle i levetidsanalyse. Heterogenitet i fordelingene kan gjøres rede for ved å inkludere kovariater i farefunksjonen, og vi ønsker å estimere effektene av disse kovariatene. I tillegg bruker vi Weibull og eksponentielle grunnlinjefarer. Vi bruker en Bayesiansk analysetilnærming for våre multistate-modeller, som kan presenteres som latente Gaussiske modeller (LGMs) ved å tilordne Gaussiske priors til det latente feltet. Inferensverktøyet kalt integrerte nestede Laplace-tilnærminger (INLA) brukes til å utføre inferens. INLA kan tilpasses og brukes på komplekse multistate-modeller, noe som gjør Bayesiansk analyse rask og nøyaktig.

Contents

Chapter 1

Introduction

Survival analysis is the study of data that describes the time to a certain event. By *event* we mean occurrences in the life of individuals that are of interest in scientific studies, for example in medicine, biology, or econometrics. Such events can be death, cancer diagnosis, divorce, the birth of a child, falling asleep, anything that can be of scientific interest. If an individual experiences an event of interest, we say that the survival time is *observed* (or *uncensored*), while if the event does not happen, we say it is *censored*. In short, censoring refers to incomplete data by means of unobserved event times. The survival analysis techniques resemble regression analysis, with the important distinction that the outcome variable time is always non-negative and often censored. In fact, ordinary statistical methods like regression analysis cannot handle right-censored (and left-truncated) survival data. This is why survival times requires a different statistical theory which is built upon two basic concepts; namely the *survival function* $S(t)$ and the *hazard rate* $h(t)$. The survival function gives the expected proportion of individuals for which the event has not yet happened by time t. In contrast, the hazard rate gives the limiting probability of experiencing the event of interest in the short time interval

$[t, t + dt]$, given that an individual has not yet experienced the event of interest by time t. Therefore, the survival function specifies an unconditional probability, while the hazard rate is defined by means of a conditional probability.

Usually, survival analysis deals with one event, but some extensions allow us to study situations where one individual is at risk of experiencing one event or several events. When there are two or more events that a patient is at risk of experiencing, we call it a *competing risks* model. If recurrent events can happen to the patient, we call it a *transient* model. We use the umbrella term *multistate models* for both the competing risks model and the transient model since they are survival models with multiple states. The motivation for this thesis is the survival data on cardiac arrest in children and adolescents. The way these patients experience different heart palpitations can be formulated as a transient model. The patients in this study are observed to make transitions within the transient model, where some states are recurring. Additionally, we adapt the *cause-specific hazards analysis*, which means competing events may be coded as a censoring event as long as this is done for every competing event type in turn.

A common goal of survival analysis is to assess the effects of several factors on survival. This is also a goal of ours, in addition to estimating the hazard functions. We assume parametric baseline hazards for each transition and include covariates through the Cox model. Since we specify the baseline hazard, which is not done in the standard Cox model, built-in functions in R cannot be used for inference. Additionally, adding prior distributions to the model parameters, thus entering the Bayesian world, makes inference more complex. Usually, inference for Bayesian models has to rely on the Markov-Chain Monte-Carlo techniques, which can be slow to converge and challenging to implement. That is why we will be characterizing our survival models as *latent Gaussian models* (LGMs), which enables us to use the *integrated nested Laplace approximation* (INLA) method proposed by Rue, Martino, and Chopin (2009). This scheme based on Bayesian

inference is used for its computational efficiency and the ability to incorporate various modeling elements available in the accompanying `R-INLA` package (r-inla.org).

The simple survival model will first be presented in Chapter 2 to introduce the topic of survival analysis. We then explore more complex multistate models in Chapter 3. We base our simulations on the paper by Niekerk, Bakka, and Rue (2019), where they show how to apply the INLA inferential scheme to a competing risks model with three competing states and longitudinal observations. A simplified competing risks simulation study is presented in Chapter 5.1. Since transient models can be viewed as nested competing risks, it was necessary to work through and understand the method on a competing risks model before attempting a more complex, transient model. A simulation study on a transient model is presented in Chapter 5.2 in preparation for Chapter 6, where we tackle real-life data on cardiac arrest patients. Chapter 7 contains the discussion of this thesis.

Chapter 2

Simple Survival Analysis

Survival analysis is the study of data that describes the time to an event. An *event* can be defined as an experience of interest and can be of either positive or negative connotation. Typically, survival analysis is used in the biomedical field, where the survival times are associated with negative events, such as death or disease occurrence. However, the survival endpoint can also refer to a positive event such as skill learned or tumor response. In the first part of this chapter, we will consider standard single endpoint survival data, which we will refer to as *simple* survival analysis. Then, in Chapter 3, we will consider survival models with several endpoints and recurring events. Recurring events will be the main focus of this thesis, but it is essential to present the basic theory before extending it to more complex models.

The simple survival model can be represented as in Figure 2.1. From Figure 2.1 we see that state 0 is the initial state, and state 1 is a terminal event. A terminal state will sometimes be referred to as "death." We are interested in the time T that the individual spends in state 0 before reaching state 1. The time T

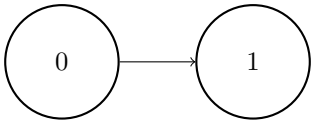

Figure 2.1: Simple survival model

is the response variable and must be a non-negative continuous or discrete random variable. The survival function $S(t)$ denotes the probability that a patient survives longer than time t and is mathematically expressed as

$$S(t) = \Pr(T > t) = 1 - F(t), \tag{2.1}$$

where $F(t)$ is the cumulative distribution function of T. The survival function is monotonically non-increasing and right continuous. At time $t = 0$, none of the patients have experienced the event, so $S(0) = 1$. Often we assume that over time, all individuals will experience the event of interest, so we write $S(\infty) = \lim_{t\to\infty} S(t) = 0$. This, however, is not necessarily true for all clinical studies. Some events do not necessarily happen to all individuals, like divorce or testicular cancer, so the random variable T may be infinite. In these situations, we say that the survival function $S(t)$ will decrease toward a positive value as t goes to infinity, $S(\infty) = \lim_{t\to\infty} S(t) = a,\ a > 0$.

We define the cumulative distribution function (CDF), which can be viewed as the complement to the survival function,

$$F(t) = P(T \leq t) = \int_0^t f(s)\, ds \tag{2.2}$$

where $f(\cdot)$ is the probability density function (PDF). We can derive the CDF and the PDF from each other from the expression

$$F'(t) = f(t). \tag{2.3}$$

The following result is used extensively in survival analysis,

$$f(t) = h(t) \cdot S(t), \tag{2.4}$$

where $h(t)$ is the hazard function. The hazard function is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathrm{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \tag{2.5}$$

that is, the instantaneous rate of failure for the patient in the interval $[t, t + \Delta t]$. We will discuss the hazard function more in Chapter 2.2

2.1 Censoring and truncation

The models used in survival analysis allow *censored* data, which the standard linear regression cannot handle. Censoring means that the event is not observed. There are three types of censoring; left-censoring, interval-censoring, and right-censoring. Left-censoring happens when the event has already occurred before the start of the study, but we do not know when it happened. For this reason, left-censored data is usually not included in a study. Interval-censoring is where the failure time is only known to have occurred within a specified time interval. Finally, right-censoring happens when the event has not yet occurred by the end of the study. For right-censored data, all we know is that the actual failure time is greater than a particular value, but not its exact value. We will only be working with right-censored data, as this is much more prevalent, and will refer to this only as *censoring* from now on. We also assume a random censorship model where the censoring time C is *independent* of the event time T, as described by Beyersmann, Allignol, and Schumacher (2012). For individual i an observed time is recorded as the event time T_i, while C_i denotes a censored time. Individual i's recorded response is presented as

$$(T_i \wedge C_i, \delta_i) \ . \tag{2.6}$$

which is called the data point y_i for that individual. The $\wedge$ notation in Equation (2.6) means that patient i's recorded time is either an event time T_i or a censoring time C_i, whichever comes first. The δ_i is the indicator of the event happening or not, taking the value of 1 if the event happened (uncensored) and 0 if it did not (censored). δ can be written mathematically as $\delta = \mathbb{1}(T \leq C)$. If, for example, patient i experiences the event of interest at time $T_i = 5$, then that patient's recorded response is $(T_i = 5, \delta_i = 1)$. If, however, patient i has not experienced the event of interest at last check-up time $t = 5$, then promptly drops out of the study, that patient's response is $(C_i = 5, \delta_i = 0)$. Usually, the study is closed at a predetermined fixed time, so an individual's event time can be the predetermined censoring time as the study is closed. Unobserved failure times give us incomplete observations. However, instead of throwing away information about a person due to the lack of an event time, we use all the data we have up until the time of censoring, as it provides valuable information.

Another type of "incomplete" data comes in the form of *truncation*. There is only one type of truncation, *left*-truncation. We say that data is left-truncated for the patients who enter the study later than time origin 0. One may also use the term *delayed entry* about such data. We denote individual i's left-truncated/delayed entry time by L_i. This delayed entry time L_i is vital to account for and include in the data, as the hazard for an event happening at time t may be different at time origin 0 and entry time L_i. However, if the hazard function is constant, left-truncation does not need to be accounted for. Different hazards will be explained in Chapter 2.2. If an individual experiences the event before their left-truncation time $(T_i \leq L_i)$, this individual will never enter the study. Individual i is in the study if there is a left-truncation time L_i (the individual has entered the study), and the maximum recorded time $T_i \wedge C_i$ has not yet happened. This can mathematically be expressed as

$$L < t \leq T \wedge C.$$

If an individual i enters the study late, its recorded response by the end of the study is

$$([L_i, T_i \wedge C_i], \delta_i). \tag{2.7}$$

2.2 Hazard function

As we have discussed, one cannot always know all failure times when doing a clinical study. This is because the true event times are censored, hidden from us. Luckily, *hazards* remain undisturbed by censoring and plays a central role in survival analysis. This is why survival analysis is hazard-based, and below, the hazard function will be presented and explained.

Assume a patient has survived up to time t. The instantaneous rate of failure for the patient in the interval $[t, t + \Delta t]$ is called the hazard function, and is expressed as a limit in the following way:

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathrm{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{2.8}$$

Here, we let Δt approach zero, giving the interpretation of the hazard of dying *at* time t. The hazard function has the following relations with the PDF, CDF, and survival function:

$$h(t) = \frac{f(t)}{S(t)} = \frac{F'(t)}{S(t)} = -\frac{S'(t)}{S(t)}. \tag{2.9}$$

where the penultimate equality is the result of $F(t) = 1 - S(t) \implies \frac{d}{dt}F(t) = \frac{d}{dt}1 - \frac{d}{dt}S(t) \implies F'(t) = -S'(t)$. Note the conditional part of Equation (2.8), which represents the fact that the hazard rate is defined by means of a conditional

probability. Also note that the hazard rate can essentially be any non-negative function.

The hazard function can alternatively be expressed as the *cumulative* hazard function,

$$\begin{aligned} H(t) &= \int_0^t h(s)\,ds = -\int_0^t \frac{d}{ds}[\ln S(s)]\,ds \\ &= -[\ln S(s)]_0^t = -\ln S(t). \end{aligned} \tag{2.10}$$

From this expression we can find another relation between the survival function and the cumulative hazard function,

$$S(t) = \mathrm{e}^{-H(t)}. \tag{2.11}$$

Details about how we extend the notation when working with multistate models will be presented in Chapter 3.

2.3 Kaplan-Meier

The Kaplan-Meier (KM) method is a non-parametric method of estimating the probability of surviving until time t. We say that the patients currently at risk of experiencing an event under observation are included in the *risk set*. The KM estimation of the survival curve is defined as

$$\hat{S}(t) = \prod_{i:t_i<t} \left(1 - \frac{d_i}{n_i}\right), \tag{2.12}$$

where n_i is the number of subjects in the risk set at time t_i, and d_i is the number of individuals that experience the event of interest at this time. In Equation (2.12), information on all patients (censored and uncensored) is used and combined in the risk set to estimate the KM curve. They are included in the risk set if they have not already experienced the event of interest at

time t_i. A new risk set is calculated every time an individual experiences the event of interest or is included in the study by a left-truncated time. If an individual experiences the event of interest, they are removed from the risk set, and the recorded number of the events increases by one. If a person drops out of the study, dies from another reason than what is being studied, or for any other reason is lost to follow up, they are removed from the subsequent risk set without the number of the event increasing. Allowing left-truncation implies that the risk set will not only decline over time but also increase when new individuals enter the study. The data structure needed to calculate the KM estimate is ordered failure times.

We conclude this section by presenting an exemplary KM curve in Figure 2.2. As can be seen, there are two KM curves. The dataset `myeloid` of package `survival` generated these KM curves. This simulated dataset is based on a trial in acute myeloid leukemia, where two treatments are being compared; treatment A and treatment B. The vertical "tacks" represent a censored time. We will not further detail KM analysis, as we will not demonstrate it in our simulated studies or real-life data analysis.

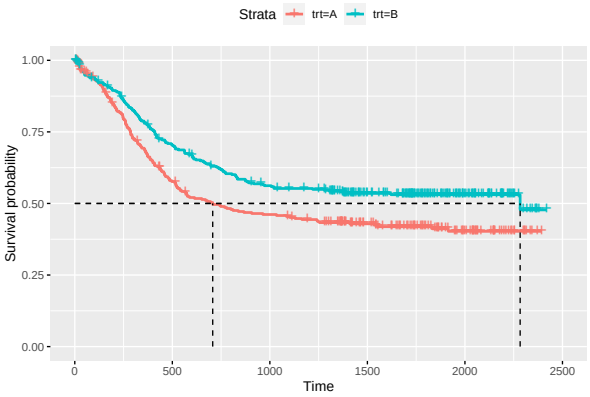

Figure 2.2: Two Kaplan-Meier curves showing the different survival proportion for the two different treatments A and B

2.4 Cox Regression

In a homogeneous population, the distribution of the time to event, described by the hazard, is the same for each individual. Therefore, we can use a standard function for the hazard, which assumes all individuals have the same hazard. However, it is often unreasonable to assume all individuals are subject to the same risk. It is rather reasonable to assume some individuals have a higher hazard of falling ill than others, based on some factors. This means we assume a heterogeneous population and should not assume the same hazard functions for all individuals. Heterogeneity in the distributions can be accounted for by including covariates in the model for the hazard. In this model, individuals with the same value of the covariates will have the same distribution, but usually, the values differ for each individual. Therefore, we call the hazard function a *proportional* hazard function, as we assume the exact shape of the baseline hazard for all individuals, but the hazards might be scaled for different values of the covariates.

For this reason, these models are the most common in survival analysis. These covariates can sometimes help explain the survival of the patients. We are often interested in how several covariates affect the survival probability, and we want to use both categorical and continuous covariates in our model. However, the Kaplan-Meier curve is only suitable for stratifying the data with one categorical covariate. In the relative risk model, we assume that the covariates $\boldsymbol{z}_i(t)$ for individual i is related to the hazard rate $h(t)$ by the following relation:

$$h_i(t) = h_0(t) \cdot r(\boldsymbol{\beta}, \boldsymbol{Z}_i(t)), \tag{2.13}$$

where $h_0(t)$ is the baseline hazard, $r(\boldsymbol{\beta}, \boldsymbol{Z}_i(t))$ is the risk function, $\boldsymbol{\beta}$ is the vector of regression coefficients and $\boldsymbol{Z}_i(t)$ is the vector of covariates for patient i. The baseline hazard function $h_0(t)$ corresponds to the hazard of experiencing the event if all covariates z_i equal zero. The baseline hazard can be a function

of time t, which means the baseline hazard may vary with time or be constant concerning time.

The commonly used form of the relative risk function is the exponential function. This is called the *Cox regression model*, also called the proportional hazards model, introduced by Cox (1972). For each individual i, the Cox model takes the form

$$\begin{aligned} h_i(t) &= h_0(t) \cdot r(\boldsymbol{\beta}, \boldsymbol{Z}_i(t)) \\ &= h_0(t) \cdot \exp\{\eta_i\} \end{aligned} \tag{2.14}$$

Different possible baseline hazard functions will be presented in Chapter 2.5, but the general case for the Cox model is to not assume a form for $h_0(t)$. We get the basic Cox model when $\eta_i = \boldsymbol{\beta}^T \boldsymbol{Z}_i$, where $\boldsymbol{\beta}$ is the vector of regression parameters and $\boldsymbol{Z}_i$ is the vector of observed covariates. By letting the predictor η_i take the structured additive form

$$\eta_i = \beta_0 + \sum_{k=1}^{n_\beta} \beta_k Z_{ki} + \sum_{j=1}^{n_f} w_{ij} f^{(j)}(u_{ij}) + \epsilon_i \ , \tag{2.15}$$

we are able to include non-linear effects through the covariates u using functions $\left\{f^{(j)}(\cdot)\right\}$, for example group-specific random effects (frailties). Here β_0 is the intercept, and the $\{\beta_k\}$s are the regression coefficients for the fixed effects $\mathbf{Z}$. The w_{ij} are known weights defined for each observed data point y_i. Finally, the ϵ_is are unstructured random effects. Frailties are useful to include in survival analysis when we suspect dependent survival times in certain clusters or unobserved heterogeneity (see Balan and Putter (2020) and Niekerk, Bakka, and Rue (2019)). A reason for including random effects as frailties is that family members may share environmental and genetic characteristics which affect their survival times, or we have recurrent events for several individuals, so it may not be appropriate to treat them as independent observations.

Compared to other relative risk models, Cox's model is commonly used because of its simple interpretation of the hazard rate ratio and assumption-free baseline hazard function. It is one of the most natural forms since it produces only positive values. A fundamental assumption for the Cox model is that the hazard functions for the patients should be proportional and not cross, hence the name *proportional* hazards model. We are primarily interested in estimating the regression coefficients $\boldsymbol{\beta}$ in Cox regression to see if one group has a relatively higher risk than another group. We assume equal baseline hazard functions for all groups, which is why this term cancels out when comparing two patients:

$$\begin{aligned}\frac{h_1(t)}{h_2(t)} &= \frac{h_0(t) \cdot \exp\{\eta_1\}}{h_0(t) \cdot \exp\{\eta_2\}} = \frac{\exp\{\eta_1\}}{\exp\{\eta_2\}} \\ &= \exp\left\{\sum_{k=1}^{p} \beta_k Z_{1k} - \sum_{k=1}^{p} \beta_k Z_{2k}\right\} \\ &= \exp\{\beta_1(Z_{11} - Z_{21}) + \cdots + \beta_p(Z_{1p} - Z_{2p})\} \qquad (2.16)\end{aligned}$$

Say we wish to estimate the effects of a unit increase in one of the covariates, e.g., covariate Z_1. Suppose that Z_1 is `Age` for an easy understanding. To quantify the effect of an increase of one year, we add 1 to the covariate value Z_1 of patient 1, so patient 1 is one year older than patient 2. The rest of the covariates are equal for patients 1 and 2. The effect of a unit increase in the covariate Z_1 is:

$$\frac{h_1(t)}{h_2(t)} = \exp\{\beta_1(Z_{11} + 1 - Z_{21}) + \cdots + \beta_p(Z_{1p} - Z_{2p})\} = \exp(\beta_1). \qquad (2.17)$$

All the other effects reduce to zero, as we can only truly compare the effect of one covariate if all the others are equal. Hence, the covariates have a multiplicative effect on the hazard. The Cox model does not estimate the survival function itself because we

do not specify the baseline hazard function other than that it must be non-negative. We will not further detail the basic proportional hazards Cox model, as we will not assume an unspecified baseline hazard in any of our simulation studies. Instead, we will be assuming a parametric baseline hazard, the details on which will be presented in the following section. Still, when comparing patients with the same parametric baseline hazard, the interpretation of the effects of the covariates are the same as for the basic Cox proportional hazards model.

2.5 Parametric baseline hazards models

From what we saw, we do not get a specification of the intercept, as the baseline hazard $h_0(t)$ is unspecified. It does, however, allow us to estimate all the other coefficients, so if the interest lies in hazard ratios, Cox's model works well. If we want a more predictive model, we should put a parametric or non-parametric assumption on the baseline hazard function to improve upon the model. Specific parametric models have been used repeatedly throughout the literature on failure time data; the *exponential* and *Weibull* models (see Niekerk, Bakka, and Rue (2019) and Martino, Akerkar, and Rue (2011)). We will be using both of these distributions in our simulations and the model building for the real-life data.

The baseline hazard is an important topic for this thesis, as its assumed shape will affect the simulations of event times and the estimated hazard shape of the real-life data. We define the baseline hazard as the function which captures everything equal for the patients, had their covariates been zero. Since Equation (2.15) includes an intercept term, which is equal for all patients, we will include the intercept term in the baseline hazard formulation. The only part which is genuinely individual for each patient is the fixed effects Z_i. This may not be the standard notation for baseline hazards, but we found it helpful to separate the hazard function into two; one part with information

about everything equal for all patients, and the other part where information may differ between patients. Details on how this is written mathematically is presented below for each of the two parametric models.

2.5.1 Exponential baseline model

The *one-parameter exponential model* is obtained when the hazard function is constant over the range of T. A constant hazard means that the instantaneous failure rate is independent of t, and is often referred to as the *memoryless property* of the exponential distribution. The risk of failure in a time interval of specified length conditioned on having survived up to the interval is the same for all intervals regardless of how long the patient has been in the study. Since we want to include explanatory covariates in our models, we will put the exponential assumption in the baseline hazard, which can be seen in Equation (2.18):

$$h_0(t) = \lambda_0, \ \lambda_0 > 0 \ , \tag{2.18}$$

where $\lambda_0 = \exp(\beta_0)$. The complete hazard will look like

$$h(t) = \exp(\eta), \tag{2.19}$$

where λ_0 of Equation (2.18) is included in η in Equation (2.19) as an intercept term. We will be using the exponential baseline model in the simulation studies as well as in the real-life data analysis.

2.5.2 Weibull baseline model

The *Weibull model* is a generalization of the exponential model and allows for a power dependence of the hazard on time. The baseline hazard will look like

$$h_0(t) = \alpha t^{\alpha-1}\lambda_0, \quad \alpha, \lambda_0 > 0, \tag{2.20}$$

where α is a hyperparameter that decides the shape, and $\lambda_0 = \exp(\beta_0)$. We are not necessarily interested in estimating α, and it shall be referred to as a nuisance parameter or *hyperparameter*. Basing our model on the Cox proportional hazard model and using the Weibull baseline hazard function, our model is now:

$$h(t) = \alpha t^{\alpha-1} \cdot \exp\{\eta\} \tag{2.21}$$

where λ_0 from the baseline hazard of Equation (2.20) is included in η as an intercept term, like in the exponential model presented above. Using a Weibull baseline hazard assumes that all patients have the same time-varying baseline hazard. By including $\exp(\eta_i)$, which is different for each individual $i = 1, \ldots, N$, we say that the shape of the hazard for each patient is similar up to a constant. The shape remains the same but multiplied by the constant term $\exp(\eta_i)$.

2.5.3 Likelihood function for censored survival data

As with most statistical models, we want to determine how likely our data is given our parameters. The likelihood function expresses this. Given N patients with lifetimes governed by a survival function $S(t)$ with associated density function $f(t)$ and hazard function $h(t)$, we can express each patient's contribution to the likelihood function. If patient i experiences the event of interest at time t_i, then its contribution to the likelihood function is

$$L_i = f(t_i) = h(t_i) \cdot S(t_i).$$

However, if patient i is censored, all we know is that its lifetime exceeds t_i. The contribution of a censored observation to the likelihood function is

$$L_i = S(t_i).$$

We want to combine both cases in the likelihood contribution and do so by introducing the event indicator δ_i, which takes the value 1 if patient i experiences the event of interest or 0 otherwise;

$$L_i = [h(t_i) \cdot S(t_i)]^{\delta_i} [S(t)]^{1-\delta_i}. \tag{2.22}$$

From Equation (2.22) we see that $S(t)$ is included whether δ_i equals 0 or 1, so we can safely move $S(t)$ outside any parentheses. The complete likelihood function for censored data is then

$$L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} [h(t_i)]^{\delta_i} S(t). \tag{2.23}$$

Taking the log and recalling that $S(t) = \exp(-H(t))$ we get

$$\log L = \sum_{i=1}^{N} [\delta_i \cdot \log h(t_i) - H(t)]. \tag{2.24}$$

Given a Weibull hazards model, the log-likelihood is as in Equation (2.25).

$$\begin{aligned} \log L &= \sum_{i=1}^{N} \left[\delta_i \log(\alpha t_i^{\alpha-1} \exp(\eta_i)) - t_i^{\alpha} \exp(\eta_i)\right] \\ &= \sum_{i=1}^{N} [\delta_i (\log \alpha + (\alpha - 1) \log t_i + \eta_i) - t_i^{\alpha} \exp(\eta_i)] \end{aligned} \tag{2.25}$$

Given an exponential hazards model, the log-likelihood is as in Equation (2.26).

$$\begin{aligned}\log L &= \sum_{i=1}^{N} \left[\delta_i \log(k \cdot \exp(\eta_i)) - kt \cdot \exp(\eta_i)\right] \\ &= \sum_{i=1}^{N} \left[\delta_i (\log k + \eta_i) - kt \cdot \exp(\eta_i)\right] \end{aligned} \tag{2.26}$$

2.5.4 Semi-parametric model

As this thesis was motivated by the real-life data on cardiac arrest among adolescents where not much data was available, we propose a fully parametric approach. However, one can assume a non-parametric baseline hazard with the parametric Cox model, resulting in a semi-parametric model. A semi-parametric approach relies more on the observed data, meaning we would have needed a substantial amount of data to get good results. It is, however, possible to use a semi-parametric approach for survival data with INLA. The method relies on splitting the time axis into a finite partition and assuming the baseline hazard constant in each time interval. This model is referred to as *the Cox model with piecewise log-constant baseline hazard*. It also typically involves using a random-walk model as a prior for smooth realizations. More details can be found in Martino, Akerkar, and Rue (2011).

Chapter 3

Multistate models

Up until now, we have thought of the event as only *one* absorbing state. However, there may be several events a patient is exposed to in the real world that we are interested in modeling. The events may be absorbing or non-absorbing states, meaning terminal or non-terminal events. In the case where there are several terminal states, we call the models *competing risks models* and these models are the simplest multistate models. We are also interested in the models that have a terminal event following a non-terminal event, and we wish to study the dependency between these. In addition, there may be several non-terminal events between which the patient may transition, which we call *transient models*. Exploring these transient models will be this thesis's primary goal, as there already exists a paper on competing risks with example simulations (see Niekerk, Bakka, and Rue (2019)). However, the competing risks model will still be presented to form a baseline understanding of Bayesian inference for survival analysis with competing risks. We also present a competing risks simulation study in Chapter 5.1. This way, we get an easier transition and a more intuitive understanding of Bayesian inference on the lesser worked-through transient models.

3.1 Competing risks

Competing risks (Hoel (1972), Moeschberger and David (1971), Altshuler (1970)) is the simplest extension from the single event model we first start learning about in survival analysis. In competing risks models, we model several competing events. Transitions between the initial state and the competing risks states are considered. There is no moving between the events, as the events are all terminal. The patients are at risk of multiple terminal events, assuming that these events are mutually exclusive. That means the patient is at risk of any of these events until one event occurs, then the risk of all other events is zero. Here, the risk of each event is dependent on the assumption of surviving all other events.

Figure 3.1 graphically explains how a patient can go from the initial state 0 to one of two terminal states 1 or 2, with hazard functions h_{01} and h_{02}, respectively.

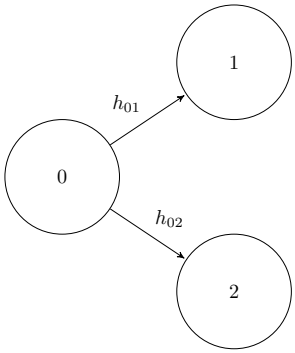

Figure 3.1: Two competing risks

This figure can of course be extended to include J terminal states, creating J cause-specific hazards h_{0j}, $j = 1, \ldots, J$. The one-sided arrows show that once a patient has ended up in one of the terminal states, they cannot return to the initial state. Note that the notation h_{lj} has been used, where the l placement refers to the *from* state and the j placement refers to the *to*

state. As there is no other starting state than 0 in a competing risks model, $l = 0$. Note that we could have used the notation h_1 and h_2 for the cause-specific hazards of going to state 1 or state 2, respectively. It will, however, be useful with specific notation when moving onto transient models where one may go from different states than 0. Thus, we incorporate the same notation for competing risks as for transient models.

An important topic in multistate modeling is *censoring by a competing event*. A difficulty in multistate theory is the assumption of independent censoring. Since we are modeling several risks compared to a simple survival model, censoring becomes informative. If an individual experiences event type 1, it cannot experience event type 2 and vice versa. This is because both events are terminal. The event $0 \to 1$ is then observed for that specific individual, while $0 \to 2$ will be censored. The censoring is *because* event 1 happened first, and hence it is no longer independent. Beyersmann, Allignol, and Schumacher (2012) explain that there are primarily two different approaches to competing risks analysis; (1) the cause-specific hazard approach (Prentice et al. (1978)) and (2) the subdistribution approach (Fine and Gray (1999)). Since we incorporate covariates in our models, we call them the *proportional* cause-specific approach and the *proportional* subdistribution approach. Both modeling approaches have advantages and disadvantages. The proportional cause-specific approach assumes right-censoring and left-truncation to be independent, where the censoring and truncation can depend on covariates included in the model. The data is incorporated through a predictor term for each cause-specific hazard function. In contrast, the proportional subdistribution approach assumes that the competing risks data are subject to random censoring only, where the censoring may not depend on covariates. Here, the data enters the model through a predictor term in each of the *cumulative incidence functions* (CIFs), the marginal cumulative probability of experiencing the specific event. Discussions about which approach to take is an ongoing debate. Taking the approach of Niekerk, Bakka, and Rue (2019) along with the algorithm of Beyersmann, Allignol, and Schumacher (2012),

we use the proportional cause-specific hazard analysis method. More on this topic will be discussed in Chapter 7. For now, all we need to know is that the method of cause-specific hazards works and is the way we will be conducting all analyzes in this thesis.

The notation for competing risks observations is similar to what we saw in Chapter 2, but we expand the δ_i indicator to δ_{ij} to include information about the specific state the patient has moved to. We shall see it used in Equation (3.12). The competing risks process moves out of the initial state 0 at time T,

$$T := \inf\{t > 0 \mid X_t \neq 0\}. \tag{3.1}$$

In the two-state competing risk model in Figure 3.1, the competing risks process is either in state 1 or in state 2 at event time T and is denoted by the *cause of failure*

$$X_T \in \{1, 2\}. \tag{3.2}$$

We account for right-censoring and left-truncation as explained in Chapter 2 in the following manner,

$$([L, T \wedge C], \mathbb{1}(T \leq C) \cdot X_T), \tag{3.3}$$

where the status indicator $\mathbb{1}(T \leq C) \cdot X_T \in \{0, 1, 2\}$ equals 0 if the observation was censored.

The cause-specific hazards are defined in a similar manner as the hazard function defined in Chapter 2.2 (Moore (2016)), except now we are including the risk that exactly cause j is happening:

$$h_{0j}(t) = \lim_{\Delta t \to 0} \frac{\mathrm{P}(t \leq T < t + \Delta t, X_T = j \mid T \geq t)}{\Delta t}. \tag{3.4}$$

We shall be using the parametric form of the Cox model for our cause-specific hazard functions, so we write

$$h_{0j}(t \mid \eta^{0j}, \boldsymbol{\theta}^{0j}) = h_{0j}^{0} \cdot \exp(\eta^{0j}), \tag{3.5}$$

where η^{0j} is as in Equation (2.15), and the vector of hyperparameters $\boldsymbol{\theta}^{j}$ depends on the parametric form of the baseline cause-specific hazard h_{0j}^{0} (e.g. the shape parameter α of the Weibull model) (Niekerk, Bakka, and Rue (2019)).

The cause-specific hazard function for individual i is written

$$h_{0j;i}(t \mid \eta_i^{0j}, \boldsymbol{\theta}^{0j}) = h_{0j}^{0} \cdot \exp(\eta_i^{0j}). \tag{3.6}$$

We often do not write the hazard function as $h_{0j}(t \mid \eta^{0j}, \boldsymbol{\theta}^{0j})$ simply because it is quite extensive. However, in this thesis the hazard function always conditions on the η function and the hyperparameters ($\boldsymbol{\theta}^{0j}$) even when we only write $h_{0j}(t)$. Assuming the basic form of the linear predictor of Equation (2.15) and p covariates, for each individual i, the linear predictor for transition $0 \to j$ is

$$\eta_i^{0j} = \beta_0^{0j} + \beta_1^{0j} Z_1 + \cdots + \beta_p^{0j} Z_p. \tag{3.7}$$

Using the competing risks model from Figure 3.1 as an example, there are two linear predictors as there are two cause-specific hazard functions. Each hazard function may have different parameter estimates, which is why we must be specific with the notation.

We write $H_{0j}(t)$ for the cumulative cause-specific hazards

$$H_{0j}(t) := \int_0^t h_{0j}(u)\, du. \tag{3.8}$$

All cause-specific hazards should completely determine the stochastic behavior of the competing risks process. Adding all the cause-specific hazards from Equation (3.5) gives the all-cause hazard due to the additivity of probabilities. Mathematically, this is expressed as

$$h_{0\cdot}(t) = \sum_{j=1}^{J} h_{0j}(t), \tag{3.9}$$

and the interpretation is the hazard of experiencing *any* event at time t. We also write

$$\begin{aligned} H_{0\cdot}(t) &:= \int_0^t h_{0\cdot}(u)\, du \\ &= \sum_{j=1}^{J} H_{0j}(t), \end{aligned} \tag{3.10}$$

for the cumulative all-cause hazard.

The survival function $P(T > t)$, or $S(t)$, of the waiting time T in the initial state 0 is a function of all the cause-specific hazard functions and is written

$$\begin{aligned} S(t) = P(T > t) &= \exp(-\int_0^t h_{0\cdot}(u)\, du) \\ &= \exp(-H_{0\cdot}(t)). \end{aligned} \tag{3.11}$$

The interpretation of the survival function for a competing risk is the probability of surviving *any* event past time t. The likelihood function for competing risks models is similar to that of a simple survival model, but we expand it to take into account J competing risks for N individuals. We denote the cause-specific hazards for cause $j = 1, \ldots, J$ as in Equation (3.5). The likelihood function for the observed event times $\mathbf{t} = \{t_1, \ldots, t_N\}$ with event indicators $\delta_{ij},\ i = 1, \ldots, N,\ j = 1, \ldots, J$ is:

$$L(\boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{t}) = \pi(\mathbf{t}|\boldsymbol{\eta}, \boldsymbol{\theta})$$
$$= \prod_{i=1}^{N}\prod_{j=1}^{J} h_{0j;i}(t_i|\eta_i^{0j}, \boldsymbol{\theta}^{0j})^{\delta_{i;0j}} \exp\left(-\int_0^{t_i} h_i^{0j}(u|\eta_i^{0j}, \boldsymbol{\theta^{0j}})du\right), \tag{3.12}$$

where $d_{i;0j}$ is an indicator if individual i died from cause $0 \to j$ (Niekerk, Bakka, and Rue (2019)).

We simulate a competing risks model with two competing risks for patients with fixed covariates in Chapter 5.1. We show how well the predefined parameter values are estimated using INLA as our inferential scheme and how well it captures the true hazard function. INLA will be presented in Chapter 4 where we will discuss how it can be used in survival models.

3.2 Semi-Competing risks

In the statistical literature, data that arise when the observation of the time to some non-terminal event is subject to some terminal event are referred to as *semi-competing risks* data. Figure 3.2 is a semi-competing risks model. Compared to the competing risks model, the semi-competing risks model includes a non-terminal state and allows a transition from the non-terminal state 1 to the terminal state 2. This type of model is also known as the Illness-Death model, where the patients are at risk of falling ill, dying without illness, or dying after illness, and is of great importance in biostatistics.

The general Illness-Death model differentiates between three hazards; the hazard of illness, the hazard of death without illness, and the hazard of death with illness. We will not look further into this multistate model, but it is a stepping stone to working with transient models.

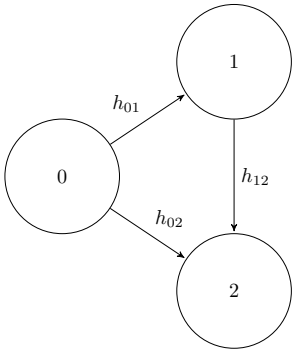

Figure 3.2: Semi-competing risks model, three states

3.3 Transient models

While competing risks refer to several competing terminal events, models with non-terminal events out of which one can transition are referred to as transient models (Kay (1982), Scheike and Zhang (2007)). For example, Figure 3.3 is a typical transient model, sometimes referred to as the Illness-Death model *with* recovery. We will use the model in Figure 3.3 when simulating a transient model as well as analyzing the real-life data.

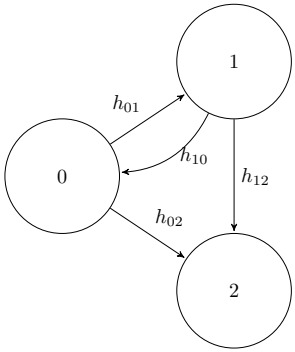

Figure 3.3: Transient model, three states

Here, the disease is curable as transitions between states 0 and 1 are possible in both directions. The clue with transient models is to view them as *nested* competing risks (Beyersmann, Allignol, and Schumacher (2012)). Instead of viewing the model in its full complexity, only consider the individual's possible risks given their state. Take, for example, Figure 3.3, and let us say we are following patient i who begins in state 0. They risk going to state 1 or state 2 (like in the case of the competing risk in Figure 3.1). Say they end up in state 1, which is a transient state. The risk they now face is going back to state 0 or to state 2. This is simply another competing risks situation. Patient i can go between these two competing risks situations until they end up in terminal state 2, out of which there are no transitions.

Even though the notation for transient models is similar to that of competing risks, we present them as there are slight differences. Now, the notation for the hazard functions is

$$h_{lj}(t \mid \eta^{lj}, \boldsymbol{\theta}^{lj}) = h_{lj}^0 \cdot \exp\{\eta^{lj}\}. \tag{3.13}$$

In the case of Figure 3.3, there are four different cause-specific hazard functions. The cumulative cause-specific hazards can be calculated similarly as in Equation 3.8. Since only two of the three states in our transient model are transient states, we have two all-cause hazards; one all-cause for the hazard out of state 0 and one out of state 1.

$$h_{0\cdot}(t) = \sum_{j=1}^{J} h_{0j}(t), \tag{3.14}$$

$$h_{1\cdot}(t) = \sum_{j=1}^{J} h_{1j}(t). \tag{3.15}$$

The cumulative all-cause hazards and survival functions are found similarly as in Equations (3.10) and (3.11), respectively.

Usually, a multistate model is assumed to be a *time-inhomogeneous Markov process* (Beyersmann, Allignol, and Schumacher (2012)), where the future development of the process only depends on the current state. We also take into account the time t_0 the individual entered their current state. This is sometimes referred to as a *clock-forward* model, as opposed to a *clock-reset* model. The clock-reset model, also known as the time-homogeneous Markov model, resets the time when a transition has been made and only depends on the length of the time interval, not the interval itself. This is best explained by an example. Consider a patient who starts in state 0 at time 0. This patient is either transitioning to state 1 or 2, which is a competing risk experiment with transitions hazards $h_{01}(t)$ and $h_{02}(t)$. If the patient enters state 2, the experiment stops as state 2 is a terminal state. However, if the patient enters state 1 at time t_0, i.e., $X_{t_0} = 1$, a second experiment follows. Another competing risks experiment is carried out, with transition hazards $h_{10}(t)$ and $h_{12}(t)$, $t \geq t_0$, i.e., in the clock-forward model we only consider the transient hazards for values t with $t \geq t_0$.

Note that movements within a multistate model generate 'internal' left-truncation (Beyersmann, Allignol, and Schumacher (2012)), which does not happen in a competing risks model. Say, individual i moves from state 0 to state 1 at time $t_i = 0.5$. Since state 1 is transient, individual i experiences the competing risks of state 0 and state 2 but does not start from time 0 but rather from time $t_i = 0.5$. From this viewpoint, individual i has entered the study at a delayed time, and time $t_i = 0.5$ will be the left-truncation time, referred to as $L_i = 0.5$.

Another difference between a competing risks model and a transient model is the number of transitions/hazard functions. In a competing risk model, the initial state is not considered a competing state. Since there is only one transition per competing state, the number of competing states and transitions are the same. In a transient model, we have additional transitions. Using Figure 3.3 as a reference, we have two additional transitions compared

to Figure 3.1; the transition from state 1 to 0 and the transition from state 1 to state 2. This turns state 0 into a competing state once a patient is in state 1. Just as in the competing risks chapter, the notation h_{lj} has been used to prevent any confusion about which transition we write about. Again, the l placement refers to the *from* state and the j placement refers to the *to* state.

The likelihood for a transient time-inhomogeneous Markov model is almost the same as for a competing risks model, except we use the notation lj and integrate from t_0 instead of 0:

$$L(\boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{t}) = \pi(\mathbf{t}|\boldsymbol{\eta}, \boldsymbol{\theta})$$
$$= \prod_{i=1}^{N} \prod_{lj} h_{lj;i}(t_i|\eta_i^{lj}, \boldsymbol{\theta}^{lj})^{\delta_{i;lj}} \exp\left(- \int_{t_0}^{t_i} h_{lj;i}(u|\eta_i^{lj}, \boldsymbol{\theta}^{lj}) du \right). \tag{3.16}$$

Pay special attention to lj here, as it refers to the specific transitions in a transient model. For example, in the model from Figure 3.3, we take the product over $lj \in \{0 \to 1, 0 \to 2, 1 \to 0, 1 \to 2\}$, meaning four different transitions with four possibly different baseline hazard assumptions. In Chapter 5.2, we simulate a three-state transient model for patients with fixed covariates and show how well our chosen inferential scheme handles this complex multistate model in terms of parameter estimation and computing time.

Beyersmann, Allignol, and Schumacher (2012) raises the concern about dependency in transient models, as an individual may contribute more than one observed transition of the same type. The concern is that the individual contributes a "cluster of dependent data" to the analysis. However, they note that the concern disappears if we follow the time-dynamic perspective of the generating Algorithm 2 of Chapter 5.2 and since we rely on the time-inhomogeneous Markov assumption. We can then safely proceed with our simulations in Chapter 5.

Chapter 4

Inferential scheme

This chapter is heavily based on the papers by Martino and Riebler (2019) and Martino, Akerkar, and Rue (2011).

We wish to use Bayesian inference to analyze multistate time-to-event data. However, Bayesian inference often relies on Markov chain Monte Carlo (MCMC) techniques which require interaction from the user to diagnose convergence and accuracy of the estimates and often carry a high computational cost. *The Integrated nested Laplace approximations* (INLA) approach is a deterministic paradigm for Bayesian inference in latent Gaussian models (LGMs) introduced in Rue, Martino, and Chopin (2009). The main benefit of using INLA instead of the much-used MCMC techniques is computational. Furthermore, INLA is fast even for large, complex models and does not suffer from slow convergence and poor mixing. As we shall see, we can present the latent field in our hierarchical model as a latent Gaussian field, resulting in a latent Gaussian model (LGM) and thus use the INLA scheme to approximate the posterior distribution for the model parameters. The methodology of INLA will be briefly explained in this chapter, but the interested reader is directed to Rue, Martino, and Chopin (2009) and Martino and Riebler (2019) for more details.

4.1 Bayesian inference and hierarchical models

Bayesian inference is a method of estimating model parameters, different from the *frequentist* inference method usually taught in entry-level statistics courses. In the frequentist world it is usually assumed that data is generated from some fixed, unknown parameter. Let us call this parameter $\mathbf{x}$, e.g., $\mathbf{x} = (\mu, \sigma)$ in the Gaussian distribution or $\mathbf{x} = \lambda$ in the Poisson distribution. A typical inference method for estimating the value of $\mathbf{x}$ based on data for frequentists is the maximum likelihood estimation (MLE), which uses the likelihood of the observed data. The likelihood gives the chance that each possible parameter value of $\mathbf{x}$ produces the data that has been observed. The likelihood is used in both frequentist and Bayesian statistics and is expressed as

$$L(\mathbf{x}|\text{data}) = f(\text{data}|\mathbf{x}) = \prod_{i=1}^{N} f(\text{data}_i|\mathbf{x}) \ , \tag{4.1}$$

where the i.i.d. assumption holds, N is the number of data points, and data_i is each data point. The MLE is the $\mathbf{x}$ that maximizes the likelihood. This method uses only the observed data available, giving little room for adding previous knowledge about the model parameters. Frequentist statistics also depend on events being repeatable and interpret probability as the long-run frequency of repeatable experiments. However, in Bayesian statistics, the parameters $\mathbf{x}$ are random variables with a distribution. In the Bayesian world, we are often interested in finding the *posterior* distribution $\pi(\mathbf{x}|\text{data})$ of our parameters given the data, which we can use to obtain different summary statistics of interest. The posterior distribution is a probability distribution representing our updated beliefs about the parameters after seeing the data. If we already have some prior knowledge about the distribution and parameters that generate the data *before* seeing the data, it can be used to improve the posterior distribution. The prior

knowledge about parameters usually comes from experience or past experiments and is denoted $\pi(\mathbf{x})$. However, we can also use Bayesian inference if we do not have any prior knowledge. We can include *uninformative* priors in the calculation of the posterior distribution and still be in the Bayesian world (Gamerman and Lopes (2006)).

Using the general Bayesian formula, we get that the expression for the posterior distribution is

$$\pi(\mathbf{x}|\text{data}) = \frac{f(\text{data}|\mathbf{x}) \cdot \pi(\mathbf{x})}{\pi(\text{data})}, \tag{4.2}$$

where the denominator $\pi(\text{data}) = \int f(\text{data}|\mathbf{x}) \cdot \pi(\mathbf{x}) \; d\mathbf{x}$ and can be seen as a normalizing constant. Sometimes $\pi(\text{data})$ is difficult to calculate, and we ignore it by saying that the posterior distribution is proportional to the likelihood times the prior;

$$\pi(\mathbf{x}|\text{data}) \propto f(\text{data}|\mathbf{x}) \cdot \pi(\mathbf{x}) \; . \tag{4.3}$$

Hierarchical models are common in Bayesian statistics and are structured, typically, in three levels. The first level is the likelihood, the second level is called a *latent* field $\mathbf{x}$, and the third level is the hyperpriors. Latent means to be hidden or concealed, and the field $\mathbf{x}$ is called latent since it is only observed through the data $\mathbf{y}$. Letting the data be represented by $\mathbf{y}$, the latent field by $\mathbf{x}$ and any hyperparameters by $\boldsymbol{\theta}$, we define the hierarchical model as:

1. the likelihood $\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$,
2. the prior distribution $\pi(\mathbf{x}|\boldsymbol{\theta})$,
3. the hyperprior distribution $\pi(\boldsymbol{\theta})$.

From the expressions above, we find the joint posterior distribution

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})} \tag{4.4}$$

As in Equation (4.3), we can write the posterior as a proportion,

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{4.5}$$

Equation (4.5) is usually not available in closed form, which is why we need computational tools to approximate it. MCMC techniques are usually used in the Bayesian setting, but we propose the INLA scheme to perform the calculations faster.

4.2 Latent Gaussian models and INLA

All hierarchical models have a latent field. The characteristic of LGMs is that the latent field $\mathbf{x}$ is assumed to have Gaussian density conditional on some hyperparameters $\boldsymbol{\theta}$, so that $\mathbf{x}|\boldsymbol{\theta} \sim \mathcal{N}(0, \mathbf{Q}^{-1}(\boldsymbol{\theta}))$ (Rue, Martino, and Chopin (2009)). Here, $\mathbf{Q}$ is the precision (inverse of covariance) matrix of the latent Gaussian field. The data $\mathbf{y}$ are assumed to be conditionally independent given the latent field $\mathbf{x}$ and, possibly, some additional hyperparameters in the likelihood. The data $\mathbf{y}$ is actually only dependent on the latent field $\mathbf{x}$ through the linear predictor η_i. The general form of the linear predictor is presented in Equation (2.15), but we write it here again,

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

The linear predictor η_i connects the data to the latent field. Hence, we can also add 'complicated' components in the models (e.g., frailty effects), which results only in a trivial change in the Gaussian part of the model. The LGM is completed by also assuming some prior density for the hyperparameters $\boldsymbol{\theta}$. The hyperparameter α from the Weibull likelihood is assigned a

penalized complexity (PC) prior (Simpson et al. (2017) and van Niekerk, Bakka, and Rue (2021)), which is a robust prior easily accessed in `R-INLA`. This prior penalizes departure from a base model and, for this reason, earned its name.

To sum up, the three levels of the LGM are

$$\begin{aligned}\mathbf{y}|\mathbf{x},\boldsymbol{\theta}_1 &\sim \prod_i \pi(y_i|\eta_i,\boldsymbol{\theta}_1)\\ \mathbf{x}|\boldsymbol{\theta}_2 &\sim \mathcal{N}(0,\mathbf{Q}^{-1}(\boldsymbol{\theta}_2))\\ \boldsymbol{\theta} = [\boldsymbol{\theta}_1,\boldsymbol{\theta}_2] &\sim \pi(\boldsymbol{\theta})\end{aligned}$$

The posterior distribution then reads:

$$\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{y})} \tag{4.6}$$

where $\pi(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ is referred to as the likelihood, $\pi(\mathbf{x}|\boldsymbol{\theta})$ is the distribution of the latent field given the hyperparameters, $\pi(\boldsymbol{\theta})$ is the prior density of the hyperparameters and $\pi(\mathbf{y})$ is the marginal likelihood.

Again, we can write the posterior density as a proportion,

$$\pi(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) \propto \prod_i \pi(y_i|\eta_i,\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{4.7}$$

where we can take the product over all individual likelihood contributions since they are assumed to be conditionally independent of each other.

For INLA to be effective, we require the latent field $\mathbf{x}$ to be endowed with some conditional independence. That is, if two elements of the field are conditionally independent given all the others, then the corresponding entry of the precision matrix is equal to zero:

$$x_i \perp x_j | \mathbf{x}_{-ij} \iff \mathbf{Q}_{ij} = 0,$$

where the notation "$-ij$" refers to "all elements other than i and j." Hence, the latent field takes the form of a Gaussian Markov Random Field (GMRF) (Rue and Held (2005))

$$\pi(\mathbf{x}|\boldsymbol{\theta}) \propto |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\right\} \sim N(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta})). \tag{4.8}$$

Given the structure of the GMRF the precision matrix $\mathbf{Q}(\boldsymbol{\theta})$ is sparse, therefore making computations considerably faster. INLA does not estimate $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$, but rather $\pi(x_l|\mathbf{y})$ and $\pi(\theta_k|\mathbf{y})$ i.e. the posterior marginals for the latent parameters and hyper-parameters, respectively. Theoretically, these are

$$\pi(x_l|\mathbf{y}) = \int \pi(x_l|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})\, d\theta \tag{4.9}$$

$$\pi(\theta_k|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\, d\theta_{-k} \tag{4.10}$$

To be able to integrate these expressions, INLA approximates $\pi(\boldsymbol{\theta}|\mathbf{y})$ for a certain θ_k and use the new approximated $\tilde{\pi}(\theta_k|\mathbf{y})$ instead. This is where the Laplace approximation part of INLA enters. Then, posterior marginals for the latent variables $\tilde{\pi}(x_l|\mathbf{y})$ are computed via numerical integration:

$$\begin{aligned}\tilde{\pi}(x_l|\mathbf{y}) &= \int \tilde{\pi}(x_l|\boldsymbol{\theta}, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \\ &\approx \sum_{k=1}^{K} \tilde{\pi}(x_l|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k \end{aligned} \tag{4.11}$$

where $\boldsymbol{\theta}_k$ are points accurately chosen in the $\boldsymbol{\theta}$ space and Δ_k are integration weights.

Summing up, INLA can be applied to LGMs that fulfill the following assumptions:

1. Each data point y_i depends only on one of the elements of the latent field $\mathbf{x}$, namely the linear predictor η_i, so the likelihood can be written as

$$\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1 \sim \prod_i \pi(y_i|\eta_i, \boldsymbol{\theta}_1)$$

2. The size of the hyperparameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$ is small (<20). This is necessary for the integral in (4.10) to be computationally feasible, which is used further in (4.11). The size of the latent field $\mathbf{x}$ can be large ($10^3 - 10^5$).

3. The latent field $\mathbf{x}$ is endowed with some conditional independence properties, resulting in a GMRF which has a sparse precision matrix $\mathbf{Q}(\boldsymbol{\theta}_2)$.

4. The linear predictor η_i depends linearly on the unknown smooth functions of covariates.

5. The inferential interest lies in the univariate posterior marginals $\pi(x_i|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$.

4.3 Using INLA in Survival Analysis

We have seen the theory behind Bayesian inference, hierarchical models, LGMs, and INLA in the sections above. In this section, we want to connect the theory behind INLA with the topic at hand; survival analysis. The Cox model has become the default choice when dealing with continuous time-to-event data (Andersen and Gill (1982), Wulfsohn and Tsiatis (1997), He et al. (2016)), which we discussed in Chapter 2. Using Bayesian methods with Cox-type models allows the use of the full likelihood to estimate all unknown parameters in the model jointly. Martino,

Akerkar, and Rue (2011) show that many of the Cox-type models can be seen as LGMs, which will be summarized below.

We write the likelihood for the ith observation as $\pi(y_i|\eta_i, \boldsymbol{\theta}_1)$, to express how it depends on some structured additive predictor η_i and, possibly, some hyperparameters $\boldsymbol{\theta}_1$. Since we want to allow left-truncation and right-censoring, we note that an observation y_i looks like $([L_i, T_i \wedge C_i], \mathbb{1}(T \leq C) \cdot X_T)$ as was seen in Equation (3.3). The predictor η_i is as in Equation (2.15).

Niekerk, Bakka, and Rue (2019) showed that a competing risks model fit in the LGM framework. Here, we show how the competing risks model in Figure 3.1 and the transient model in Figure 3.3 fit in the LGM framework, using the five steps from the section above. From the likelihood functions in Equations (3.12) and (3.16) the first assumption is fulfilled, as the likelihood functions only depend on $\mathbf{x}$ through η_i. The second assumption is fulfilled since the size of $\boldsymbol{\theta}$ is small. We only use a Weibull and an exponential baseline hazard, so the only hyperparameter from the likelihood included is the α from Equation (2.21), as the exponential model does not have any hyperparameters. The latent field for a competing risks model is presented as

$$\mathbf{x} = \{\{\beta_k^{lj}\}, \{f^{(lj)}(\cdot)\}, \{\eta_i^{lj}\}\}, \tag{4.12}$$

which is the same way we write the latent field for a transient model. Recall that a transient model is a nested competing risks model. Note that the number of lj transitions is different for transient models compared to competing risks models. We need to include all η_i^{lj} functions in the latent field, e.g., for the transient model in Figure 3.3 we include the functions $\eta_i^{01}, \eta_i^{02}, \eta_i^{10}$ and η_i^{12}. The latent field in Equation (4.12) is an LGM by assigning vague Gaussian priors. The third assumption is also fulfilled as we assume conditional independence between the elements of $\mathbf{x}$. The fourth assumption is fulfilled as we use the linear predictor η_i from Equation (2.15), which shows the linear dependency on the unknown smooth functions. Lastly, the fifth and final assumption is that the interest lies in the

univariate posteriors of every element in the latent field and the hyperparameters, which is fulfilled. If we wish to sample from the joint posterior, there is a way to do that. We know that $\pi(x, \theta) = \pi(x|\theta)\pi(\theta)$, so we can first sample from the marginal posterior of the hyperparameter, $\theta* \sim \pi(\theta)$, and use those samples in the posterior of the latent field $x \sim \pi(x|\theta*)$. That way, it is like sampling from the joint posterior. We do this using the function `inla.posterior.sample()`.

Above, we reasoned how multistate models could be viewed as LGMs, making it possible to use INLA to analyze Bayesian models. Moreover, INLA provides fast and accurate approximations to the posterior marginals compared to the usual MCMC method. The following chapter shows the process of simulating survival data and how we use INLA as the inferential scheme for these models.

Chapter 5

Simulation studies

This chapter contains two simulation studies. The simulations are done in R, and the R-scripts used in this thesis are publicly available at https://github.com/juliib/Multistate_models_with_INLA_master. We simulate in order to show that the proposed methodology works. That is, the methodology can recover the underlying data-generating mechanism even in the presence of right-censoring and left-truncation. Niekerk, Bakka, and Rue (2019) shows how to simulate a competing risks joint model with longitudinal data, with INLA as the inferential scheme. Using that paper as guidance, we also present a competing risks simulation before presenting a more complex multistate model; the transient model. Our modeling approach is the proportional cause-specific hazards model, which was explained in Chapter 3.

5.1 Competing risks simulation

The first simulation study we will look at is the competing risks model. Having seen a competing risks simulation analyzed with the INLA scheme will hopefully make understanding the transient model simulation easier. The initial state S_0 is 0, where

all patients start at time origin $t_0 = 0$. Every individual stays in state 0 until another event occurs, whichever comes first of state 1 or state 2. Figure 5.1 shows a visual representation of the model. The circles represent the possible states and the pointed arrow show which transition is possible to which state. Since the arrows are not double-sided, it shows that both events 1 and 2 are terminal events.

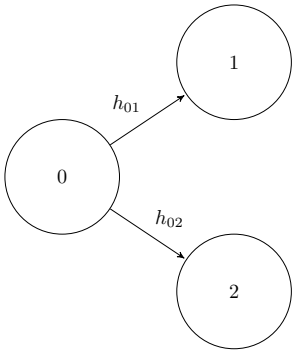

Figure 5.1: Two competing risks

In Figure 5.1 we see the cause-specific hazard functions h_{01} and h_{02}. The cause-specific hazards are Cox models with different baseline hazards. We assume an exponential baseline for transition $0 \to 1$ and a Weibull baseline for transition $0 \to 2$,

$$h_{01} = \exp(\eta^{01}) \tag{5.1}$$

$$h_{02} = h_{02}^0 \exp(\eta^{02}) \tag{5.2}$$

In Equations (5.1) and (5.2), $\eta^{01} = \beta_0^{01} + \boldsymbol{\beta}^{01}\mathbf{Z}$ and $\eta^{02} = \beta_0^{02} + \boldsymbol{\beta}^{02}\mathbf{Z}$. The chosen parameters are $\beta_0^{01} = 1$, $\boldsymbol{\beta}^{01} = [2, 0.4]$ and $\beta_0^{02} = 2$, $\boldsymbol{\beta}^{02} = [0.6, 1]$. $\mathbf{Z} = [Z_1, Z_2]^T$, where $Z_1 \sim N(0, 1)$ and Z_2 is Bernoulli($p = 0.5$) . Since transition $0 \to 2$ is Weibull distributed, the baseline hazard takes the form $h_{02}^0 = \alpha t^{\alpha-1}\lambda_0$ where $\alpha_2 = 0.8$. As we discussed in Chapter 2 we include every

term that is equal for all patients in the baseline. That is, we include $\lambda_0 = \exp(\beta_0^{01})$ in the baseline for h_{01} and $\lambda_0 = \exp(\beta_0^{02})$ in the baseline for h_{02}. A plot of the baseline hazards can be seen in Figure 5.2.

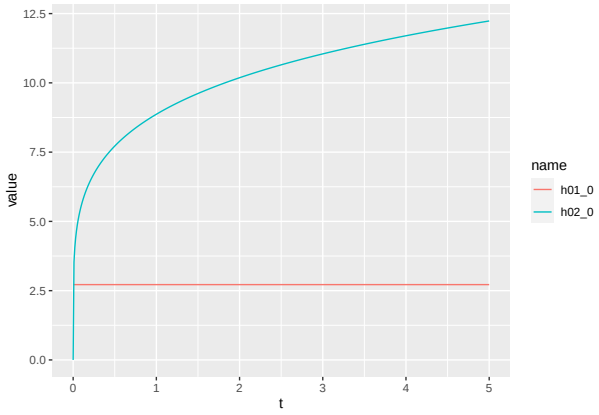

Figure 5.2: Baseline hazards for the competing risks model

The cumulative hazards are expressed as

$$H_{01}(t) = \int_0^t h_{01}(u)\ du = t \cdot \exp(\eta^{01})$$
$$H_{02}(t) = \int_0^t h_{02}(u)\ du = t^{\alpha} \cdot \exp(\eta^{02})$$

We plot the cumulative baseline hazards in Figure 5.3. The cumulative all-cause hazard is expressed as

$$\begin{aligned} H_{0\cdot}(t) &= H_{01}(t) + H_{02}(t) \\ &= t \cdot \exp(\eta^{01}) + t^{\alpha} \cdot \exp(\eta^{02}) \end{aligned}$$

To sample from our distribution, we shall use the `inversion method`, a popular technique to create samples from a distribution

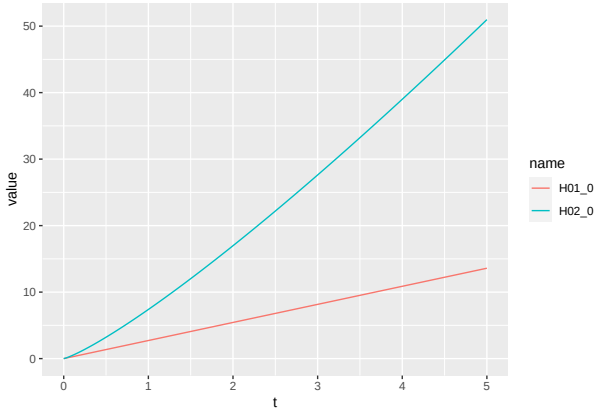

Figure 5.3: Cumulative baseline hazards for the competing risks model

of a continuous variable. We use that $F(t) = 1 - \exp(-H_{0\cdot}(t)) = u$, where $u \sim \text{Unif}[0, 1]$. Details on the inversion method for competing risks can be found in Beyersmann, Allignol, and Schumacher (2012). If we cannot find the inverse of $H_{0\cdot}(t)$ analytically, we can use the numerical inversion `uniroot` function in R. The `uniroot` function searches the interval from `lower` to `upper` for a root of the function $H_{0\cdot}^{-1}(-\ln(1-u))$. Thus, the collection of simulated $t's$ replicates our target distribution of the event times T. We simulate $N = 300$ patients with the algorithm presented in detail in Algorithm 1, and present the following transition matrix:

	0	1	2
0	F	T	T
1	F	F	F
2	F	F	F

This matrix shows that patients can go from state 0 to states 1 and 2 but not from state 0 to 0. Since states 1 and 2 are terminal states, we see that no transition out of those states is possible. The event time T decides how long a patient is in state 0.

Algorithm 1 Simulating event times for two competing risks

for each patient j **do**

1. Simulate values for the covariates $\boldsymbol{Z}^j$
2. Compute cause specific hazards $h_{01}^j(t)$, $h_{02}^j(t)$
3. Set starting state S_0 to 0
4. Compute the cumulative all-causes hazard $H_{0\cdot}(t) = H_{01}(t) + H_{02}(t)$
5. Use `uniroot` function or invert $H_{0\cdot}(t)$ to simulate an event time T from $F(t)$
6. Run a binomial experiment to decide the specific event that happens at time T. When the starting event state is $S_0 = 0$ the probability that the event at time T is to enter state r and not r' is:

$$P(X_T = r \mid S_0 = 0) = \frac{h_{0r}}{h_{0r} + h_{0r'}}$$

7. Save event time T, state r and covariates for patient j in a dataset

end for

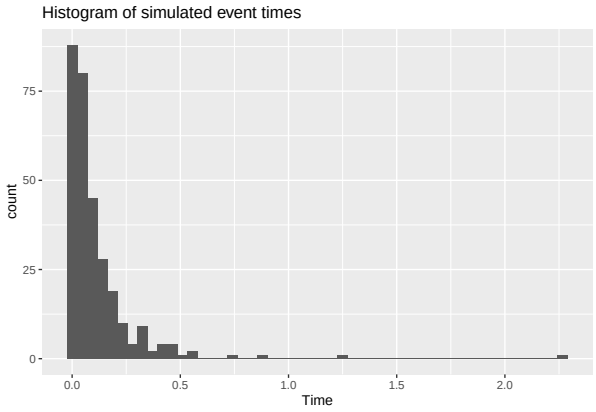

Figure 5.4: Histogram of simulated event times

ID	time	event	X_1	X_2
1	0.137	2	-0.841	1
2	0.007	1	1.384	0
3	0.1	2	-1.255	1
4	0.042	2	0.07	1
5	0.004	2	1.711	1
6	0.072	2	-0.603	1

Table 5.1: First six simulated patients with corresponding event times, event status and covariates.

	event	
	1	2
captured	107	193

Table 5.2: Overview of how many patients went to which state

The first six patients of our dataset are presented in Table 5.1. Notice that each patient only has one row of information since each patient can experience only one of the two competing risks. A histogram of all simulated event times can be seen in Figure 5.4. We look at how many patients each state captures in Table 5.2. State 1 captures 107 patients while state 2 captures 193 patients. For the implementation, we create two datasets, one per transition. Both datasets have the same amount of rows, keeping the information of all patients. The first dataset contains information on transition $0 \to 1$, creating a status variable specifying whether the specific transition has been observed. The same is done for transition $0 \to 2$. This way, we keep all observed transitions in their respective datasets, ready to build a data frame suitable to run `R-INLA`.

The parameter estimates can be seen in Table 5.3. We present the true value first, then the estimated mean, estimated 2.5%, and 97.5% credible intervals (CIs) along with the posterior mode.

	true	mean	0.025quant	0.975quant	mode
β_0^{01}	1	1.031	0.726	1.316	1.042
β_0^{02}	2	2.031	1.708	2.355	2.03
β_1^{01}	2	1.943	1.725	2.162	1.943
β_2^{01}	0.4	0.49	0.106	0.87	0.492
β_1^{02}	0.6	0.654	0.449	0.863	0.652
β_2^{02}	1	0.881	0.586	1.177	0.88
α	1.2	1.159	1.034	1.292	1.156

Table 5.3: True parameter values and estimated values for the competing risks model.

The model summary shows that the estimated CIs fit well with the true parameter values. For convenience, we plot the estimated CIs for the intercepts, fixed effects, and hyperparameter in Figure 5.5, where the red dots are the true values.

INLA also estimates the hazards, using the function `inla.posterior.sample()` to sample from the joint posterior. We sample 1000 samples from the joint posterior and calculate 1000 hazard estimates, which can be seen in Figure 5.6. As we can see, the true hazard function is well recovered by the estimations. As this is a simulation, new event times will be simulated each time we run the code. We repeated this simulation several times, which led to similar results.

5.2 Transient model simulation

We want to simulate data from the transient model of Figure 5.7. Here, states 1 and 2 are transient, while state 3 is absorbing. This simulation study is based on the theory we presented in Chapter 3.3, with the difference that the states are called 1, 2, and 3 instead of 0, 1, and 2. We define the following transition matrix for the problem

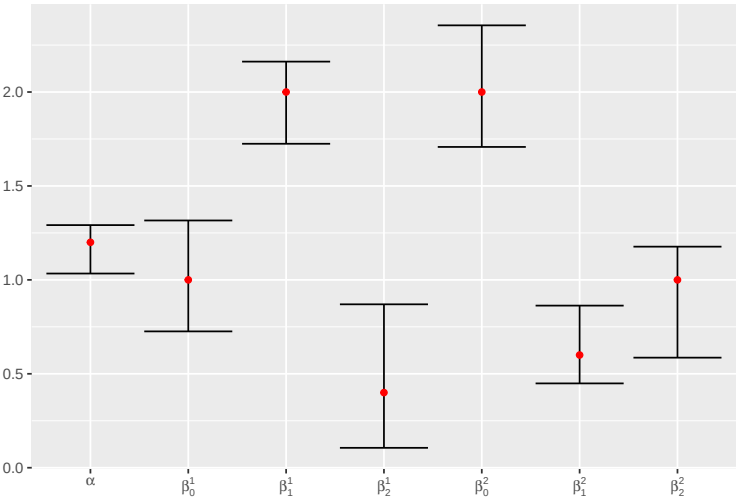

Figure 5.5: Estimated credible intervals for all parameters of the competing risks model. The red dot is the true value.

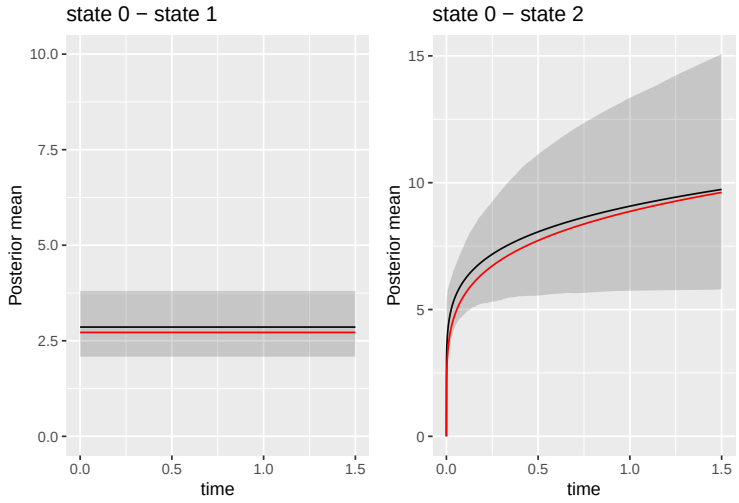

Figure 5.6: Estimated hazard functions for the competing risks model. The red line is the true hazard.

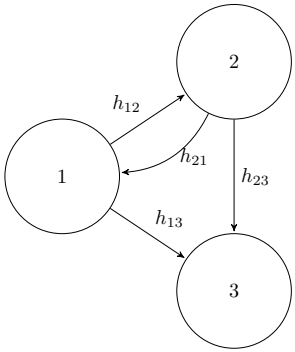

Figure 5.7: Transient model, three states

	1	2	3
1	F	T	T
2	T	F	T
3	F	F	F

This transition matrix is read similarly to the transition matrix from the Competing Risks simulation in Chapter 5.1. One can go between states 1 and 2 as many times as possible, but one can never leave if one enters state 3. This is why all transitions from state 3 are labeled `F`.

There are 4 hazards to be estimated. We use the Cox model and define them for each patient $j = 1, \ldots, N$ as

$$h_{12:j}(t) = h_{12}^0(t) \exp(\beta_{12}^0 + \beta_{12}^T \mathbf{Z}_j) \tag{5.3}$$

$$h_{13;j}(t) = h_{13}^0(t) \exp(\beta_{13}^0 + \beta_{13}^T \mathbf{Z}_j) \tag{5.4}$$

$$h_{21;j}(t) = h_{21}^0(t) \exp(\beta_{21}^0 + \beta_{21}^T \mathbf{Z}_j) \tag{5.5}$$

$$h_{23;j}(t) = h_{23}^0(t) \exp(\beta_{23}^0 + \beta_{23}^T \mathbf{Z}_j) \tag{5.6}$$

where $h_{lj}^0(t)$ are baseline hazards. The vector $\mathbf{Z}$ consists of two covariates; $Z_1 \sim N(0, 1)$ and $Z_2 \sim \text{Bernoulli}(p = 0.5)$. The chosen β_{lj} coefficients are: $\beta_{12}^0 = 1.3$, $\beta_{12} = [0.5, 0.5]$, $\beta_{13}^0 = 0.5$,

$\beta_{13} = [-0.5, 0.3]$, $\beta_{21}^0 = 1.1$, $\beta_{21} = [0.2, -0.2]$ and $\beta_{23}^0 = 0.9$, $\beta_{23} = [0.4, -0.1]$.

For the baseline hazards, we choose two Weibull and two exponential models, where one Weibull model is increasing, and the other is decreasing. We present the following baseline hazards;

$$h_{12}^0(t) = \alpha_1 t^{\alpha_1 - 1} \lambda_0^{12} \tag{5.7}$$
$$h_{13}^0(t) = \alpha_2 t^{\alpha_2 - 1} \lambda_0^{13} \tag{5.8}$$
$$h_{21}^0(t) = \lambda_0^{21} \tag{5.9}$$
$$h_{23}^0(t) = \lambda_0^{23} \tag{5.10}$$

where $\alpha_1 = 0.8$, $\alpha_2 = 1.2$, and $\lambda_0^{lj} = \exp(\beta_{lj}^0)$. Other choices of parameter values are, of course, possible, but we found that certain parameter values resulted in too extreme hazard functions, making all the patients experience the terminal event too early. The chosen values give reasonable hazard rates and simulate realistic transitions for each patient.

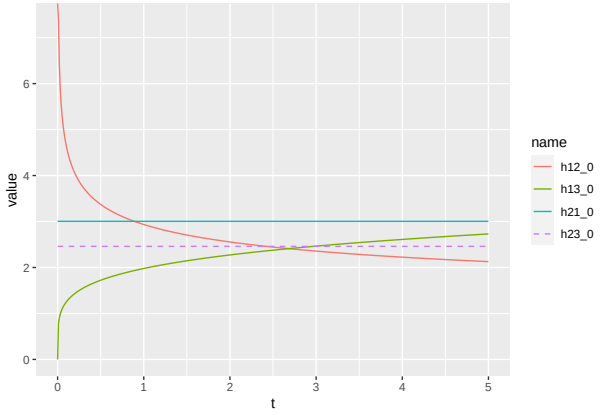

Figure 5.8: Cause-specific hazards for the transient model.

We plot the baseline hazards in Figure 5.8, while the cumulative baseline hazards are presented in Figure 5.9. The procedure of

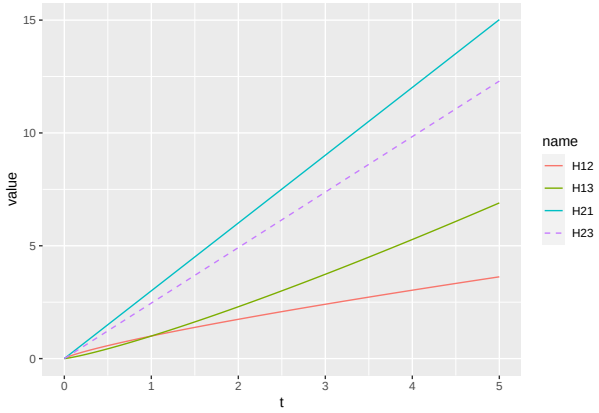

Figure 5.9: Cumulative cause-specific hazards for the transient model.

simulating $N = 300$ patients, with associated event times and event states can be found in Algorithm 2. A histogram of all event times can be found in Figure 5.10. We view the head of the dataset from the R output below. This dataset has all the information about a patient; `ID`, `Tstart`, `Tstop`, `duration`, `from`, `to`, `status`, `cov1` and `cov2`. We might notice in this dataset that every second line contains the same `Tstart` and `Tstop` time as the row above. We keep track of the events that did not happen to each patient, which we can see from their `status` equaling 0. This is called *extending* the dataset, and we call the dataset the *extended* dataset. In the competing risks simulation of Chapter 5.1, we created two datasets for the two transitions. It gives the same results whether we separate each transition in their own datasets or keep all the information in one big dataset with several rows per transition. They are simply two different ways of preparing a data frame suitable to run `R-INLA`.

```
##   id    Tstart     Tstop  duration from to status       cov1 cov2
## 1  1 0.0000000 0.6109676 0.6109676    2  1      1 -0.8178465    0
## 2  1 0.0000000 0.6109676 0.6109676    2  3      0 -0.8178465    0
## 3  1 0.6109676 0.8950126 0.2840450    1  2      1 -0.8178465    0
## 4  1 0.6109676 0.8950126 0.2840450    1  3      0 -0.8178465    0
## 5  1 0.8950126 1.7461798 0.8511672    2  1      1 -0.8178465    0
```

Algorithm 2 Simulating event times for a transient model with three states

for each patient j **do**

1. Simulate values for the covariates $\boldsymbol{Z}^j$
2. Compute cause specific hazards $h^j_{12}(t)$, $h^j_{13}(t)$, $h^j_{21}(t)$, $h^j_{23}(t)$
3. Simulate a starting state S_0 (1 or 2) at time t_0 and create two **if** statements
4. Compute the cumulative all-causes hazard from state S_0 with starting time t_0 (clock-forward model)
   - if $S_0 = 1$ then $H^j_{1.}(t, t_0) = H^j_{12}(t, t_0) + H^j_{13}(t, t_0)$
   - if $S_0 = 2$ then $H^j_{2.}(t, t_0) = H^j_{21}(t, t_0) + H^j_{23}(t, t_0)$
5. Use `uniroot` function or invert each cumulative all-cause hazard, simulate an event time T with hazard computed in step 4
   - if j is in state 1: $t = H^{-1}_{1.}(-\ln(1-u))$
   - if j is in state 2: $t = H^{-1}_{2.}(-\ln(1-u))$
6. Run a binomial experiment to decide the specific event that happens at time T. When the starting event state is $S_0 = 1$ the probability that the event at time T is to enter state j and not j' is:

$$P(X_T = j \mid S_0 = 1) = \frac{h_{1j}}{h_{1j} + h_{1j'}}$$

7. Save event time T, state j and covariates for patient j in a dataset

end for

```
## 6  1 0.8950126 1.7461798 0.8511672    2  3      0 -0.8178465    0
```

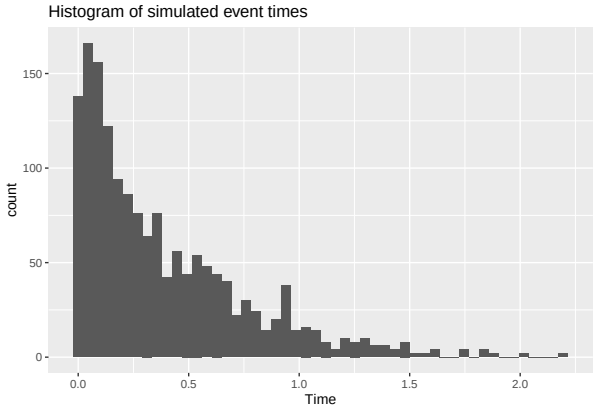

Figure 5.10: Histogram of simulated event times for the transient model.

Let us look at the event history of 12 randomly chosen patients in Figure 5.11. The vertical axis on the left of each graph shows the three possible states for each patient. Some patients are simulated to start in state 1, others in state 2. Some patients move directly to state 3, where they cannot leave, and others go back and forth between states 1 and 2 before finally ending up in state 3.

For our simulation to be successful, we need enough data per event. The following table shows the percentage of patients that move from states 1 and 2. It does not have to be a 50% transition rate, as long as we see a substantial amount of patients going to each state. There is reason to believe that fewer patients go from state 1 to 3, referring to a `healthy` $\rightarrow$ `dead` transition. This is shown in the transition table by transition $1 \rightarrow 2$ capturing most of the patients.

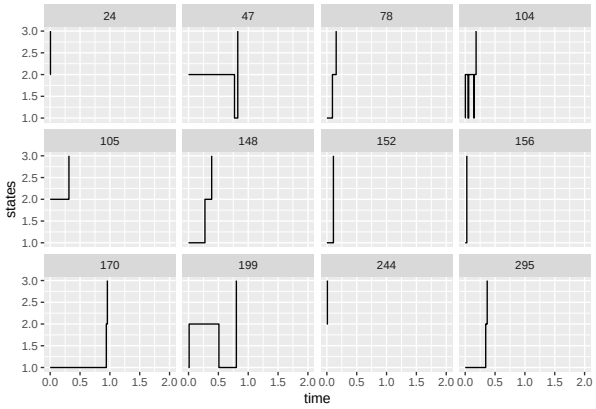

Figure 5.11: Event history of 12 randomly chosen patients

from	to	percent
1	2	71
1	3	29
2	1	53
2	3	47

The parameter estimates can be seen in Table 5.4. We present the true value first, then the estimated mean, estimated 2.5%, and 97.5% credible intervals (CIs) along with the posterior mode. The model summary shows that the true parameter values fit well within the estimated CIs. For convenience, we plot the estimated CIs for the intercepts, fixed effects, and hyperparameters in Figure 5.12, where the red dots are the true values.

We also sample the joint posterior to estimate the hazard functions and their 95% CIs, like we did in the competing risks simulation. The estimated hazard functions can be seen in Figure 5.13. Once again, we see that the estimation well captures the true hazard function.

	true	mean	0.025quant	0.975quant	mode
β_0^{12}	1.3	1.215	1.039	1.384	1.219
β_0^{13}	0.5	0.552	0.28	0.808	0.562
β_0^{21}	1.1	1.117	0.929	1.297	1.122
β_0^{23}	0.9	0.899	0.687	1.098	0.905
β_1^{12}	0.5	0.469	0.341	0.598	0.468
β_2^{12}	0.5	0.681	0.446	0.915	0.681
β_1^{13}	-0.5	-0.465	-0.7	-0.239	-0.461
β_2^{13}	0.3	0.307	-0.076	0.681	0.311
β_1^{21}	0.2	0.245	0.102	0.386	0.245
β_2^{21}	-0.2	-0.322	-0.599	-0.047	-0.321
β_1^{23}	0.4	0.461	0.309	0.613	0.46
β_2^{23}	-0.1	-0.201	-0.492	0.089	-0.201
α_1	0.8	0.781	0.714	0.851	0.779
α_2	1.2	1.15	1.001	1.31	1.146

Table 5.4: True parameter values and estimated values for the transient model.

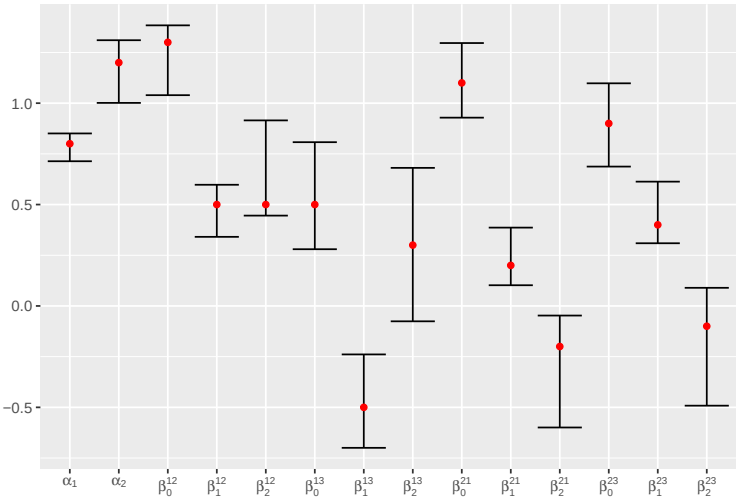

Figure 5.12: Estimated credible intervals for all parameters of the transient model. The red dot is the true value.

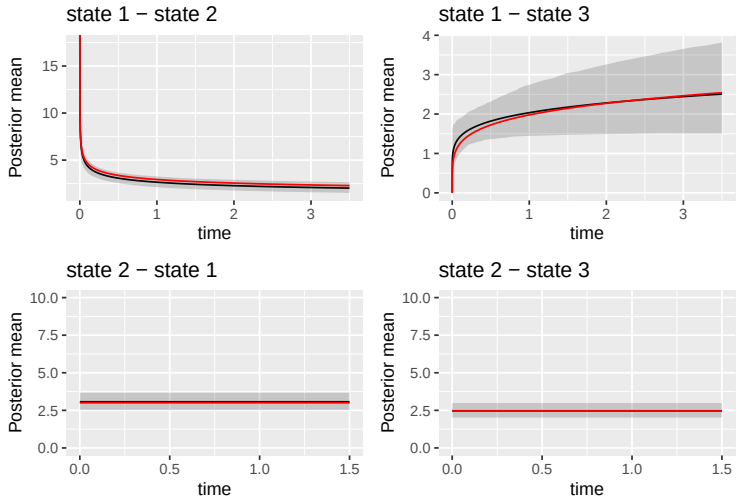

Figure 5.13: Estimated hazard functions for the transient model. The red line is the true hazard.

Chapter 6

Real life data analysis

The motivation behind this thesis is a dataset on cardiac arrest in children and adolescents presented in Nordseth et al. (2019) and Skogvoll et al. (2020). Here they consider a multi-state model where patients are observed to transition between four transient states, where some end up in a terminal state. Electrocardiogram (ECG) recordings were collected in patients who received cardiopulmonary resuscitation (CPR) at the Children's Hospital of Philadelphia (CHOP) between 2006 and 2013. During pediatric CPR, the states the patients may transition between are pulseless electrical activity (PEA), asystole, ventricular fibrillation/tachycardia (VF/VT), and return of spontaneous circulation (ROSC). The terminal state is death. Therefore, the estimation of hazard rates and covariate effects is of interest. To perform analysis, Nordseth et al. (2019) first uses the method of finding the Nelson-Aalen estimator of the cumulative hazard of specific state transitions over time. Then, numerical methods were used to find the derivative of the Nelson-Aalen estimator. Finally, continuous hazard rates for all relevant transitions were estimated using smoothing splines. Nordseth et al. (2019) do not investigate any possible state transitions with patient covariates, which they incorporate in Skogvoll et al. (2020). We

propose that the INLA scheme is well suited for this multistate survival model, where hazard functions can be directly accessed by adopting a parametric approach, and patient covariate effects are readily displayed. As multistate models can be viewed as nested competing risks, we show that INLA works well in the transient model scheme as well.

6.1 Exploring the dataset

Before we do any analysis, let us familiarize ourselves with the real-life data. The collected data consists of the following columns: `ID`, `ini`, `last`, `state`, `entry`, `to`, `time`, `Age`, `Gen`, `Mass` and `res`. `ID` contains the patients' ID, `ini` contains the initial state when cardiac arrest begins, and the patients' last recorded state is stored in `last`. The column `state` contains which state the patient was in at `entry`, and column `to` contains which state the patients moved to at which `time`. The rest of the columns in the dataset are `Age` (age of patients in years), `Gen` (gender of the patients coded as 0 or 1), `Mass` (weight of the patients in kg), and `res` (indicates whether the cause of cardiac arrest was "respiratory" or not). Cardiac arrest due to respiratory issues can be due to drowning, strangulation, airway obstruction, or progressive respiratory failure (like in pneumonia), as different from "circulatory" like in a primary cardiac disease or after profuse bleeding, or possible "other" causes.

To show the structure of the dataset, we present some example rows of the original dataset in Table 6.1. The original IDs have been renamed to protect any privacy. We choose not to include the columns `ini` and `last` due to lack of space in the table, and the information is already stored in the first row's `state` and the last row's `to` of each patient.

In addition to showing a few rows of the dataset, it is helpful to plot histograms of the covariates. The histograms of the covariates is presented in Figure 6.1. There are 74 patients recorded in this study. There are two patients whose covariates are missing in part, so we only have complete information on

ID	from	to	entry	time	Age	Gen	Mass	res
1	3	4	0	93.9	16.92	0	55	1
1	4	5	93.9	243.985	16.92	0	55	1
10	4	5	0	240.051	8.08	1	40	1
10	5	4	240.051	615.273	8.08	1	40	1
10	4	1	615.273	765.878	8.08	1	40	1
10	3	4	0	400.636	14.92	1	48	1

Table 6.1: Example rows of the original dataset.

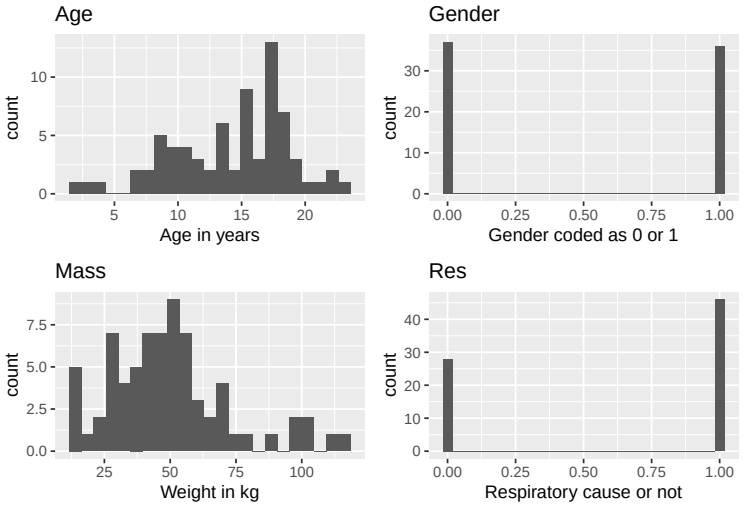

Figure 6.1: Histograms of the four covariates in the real life dataset

72 patients. One patient is missing the covariates `Age`, `Gen`, and `Mass`, while the other is only missing `Mass`. We will keep these patients in the dataset for now, but we make the reader aware of this deficiency. There are 37 patients with gender coded as 0 and 36 with gender coded as 1. One patient is, as previously stated, missing its gender coding. There are 28 patients with `res`=0, and 46 with `res`=1, i.e., 46 patients' cardiac arrest was due to respiratory issues and not circulatory.

The motivating dataset is a study done on children and adolescents. The World Health Organization (WHO) defines adolescents as those between 10 and 19 years of age, but as we can see from Figure 6.1 in the `Age` plot, there are patients older than that. Disregarding the one patient's missing age value, the sample average age is 14.11. The median age is 15. The sample average weight is 49.51, removing the two patients whose `Mass` value was missing. The median weight is 47.3. The lowest recorded weight is 11.7 kg, and the heaviest weight is 114.0 kg.

The states in this study are presented in Figure 6.2. Since it is a transient model, several two-headed arrows (transitions) go between each circle (state). We pay extra attention to the absorbing state 1, from which there are no transitions. There are no censored cases in this original dataset, as patients entered this study after first experiencing signs of a cardiac arrest—every patient experiences at least one transition to another state.

When providing advanced life support (ALS) in cardiac arrest, the patient may alternate between four clinical states: ventricular fibrillation/tachycardia (VF/VT), pulseless electrical activity (PEA), asystole, and return of spontaneous circulation (ROSC). At the end of the resuscitation efforts, either death has been declared or sustained ROSC obtained. PEA and asystole are related cardiac rhythms in that they are both life-threatening and unshockable cardiac rhythms ("Pulseless Electrical Activity Asystole" (n.d.)). There may be a subtle movement away from the baseline (drifting flat-line), but there is no perceptible cardiac electrical activity. A "flat-line" is reserved for the asystole definition, but PEA includes a flat line as well as any

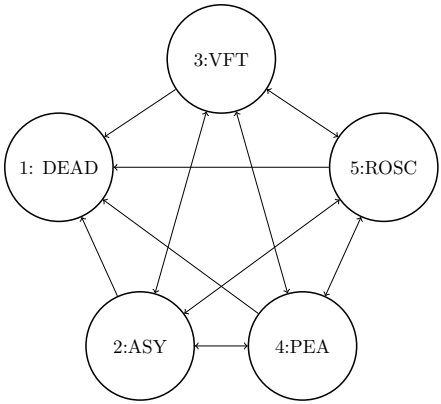

Figure 6.2: The original model. Transient states and a terminal state.

other wave. Ventricular fibrillation (VF) and pulseless ventricular tachycardia (VT) are life-threatening cardiac rhythms that result in ineffective ventricular contractions ("Ventricular Fibrillation and Pulseless Ventricular Tachycardia" (n.d.)).

Let us take a closer look at some patients. We start with patient `3`. Table 6.2 shows all recorded information we have on this patient. This patient's initial state was 4 at time $t = 0$. At the time $t = 126.342$ (a little over two minutes), the patient enters state 5 and stays there until time $t = 559.487$ (around 9 minutes into the study) when the patient moved back to state 4. The patient, however, only stays in state 4 until time $t = 1000.172$ (almost 17 minutes since entry) when the patient eventually survives in state 5, where they maintain a palpable pulse.

We look at another patient, patient `6` in Table 6.3. This patient was initially in state 4 and moved back and forth between states 4 and 5 until time $t = 1150$ (around 19 minutes after entry) when they entered state 2. From here, they enter state 1 at time $t = 1264$ (21 minutes) and unfortunately never leave that state.

Note that in the real-life dataset, no patient's initial state is state

ID	from	to	entry	time	Age	Gen	Mass	res
3	4	5	0	126.342	13.5	1	45	1
3	5	4	126.342	559.487	13.5	1	45	1
3	4	5	559.487	1000.172	13.5	1	45	1

Table 6.2: Information about patient 3

ID	from	to	entry	time	Age	Gen	Mass	res
6	4	5	0	20	16.17	1	39.9	1
6	5	4	20	356	16.17	1	39.9	1
6	4	5	356	421	16.17	1	39.9	1
6	5	4	421	785	16.17	1	39.9	1
6	4	2	785	1150	16.17	1	39.9	1
6	2	1	1150	1264	16.17	1	39.9	1

Table 6.3: Information about patient 6.

5. In addition, all patients enter this study at time $t = 0$. Then as they transition through different states, internal left-truncation is generated.

6.2 Recode to three states

Given the complexity of the original model, we choose to reduce the number of states from five to three. We were advised to combine states 2 and 4 into state 2 and delete state 3 altogether. The original dataset had 74 patients and contained information about each transition that happened. We first recoded all states `4` into state `2`. The problem that arose was that some rows went from `2` to `2` because they previously had gone from `4` to `2` or `2` to `4`. We made a function that searched through each row going from `2` to `2` per distinct person `ID` and only kept the last recorded `time` and discarded the redundant rows. Since we were

recommended to delete state 3 altogether, we deleted all rows where the patient went *from* `3` in this step. We do not delete the rows going to `3` yet, because the next step was to expand the dataset to keep all the censored cases. If, for example, a patient went from `2` to `3` with status 1 (the event happened), we would like to record all the events that did not happen (with status 0). We had to add two rows for this exemplary patient; a row where the patient went from `2` to `1` with status 0 and a row where the patient went from `2` to `5` with status 0. If we had deleted all rows where patients went to state 3, we would have lost valuable censoring information, which would have biased the survival probability. After expanding the dataset, we could delete all rows going to `3`. The new, modified dataset only contains states `1`, `2`, and `5`. The times associated with each transition and the covariates for the patients are still, of course, included in the dataset. Figure 6.3 shows the three states and their respective hazard functions.

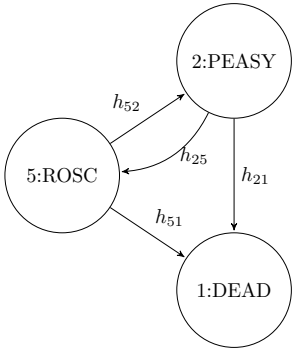

Figure 6.3: Transient model, three states

In theory, transition $5 \to 1$ is possible, but there were no occurrences in the dataset. We keep the transition arrow in Figure 6.3 for the sake of theory. Because we use an extended dataset, there are rows where the `from` column is 5 and the `to` column is 1, but the `status` variable is always 0 for those rows. Since there were no occurrences of $5 \to 1$, we do not include this transition at all

	5	2	1
5	0	29	0
2	120	0	149
1	0	0	0

Table 6.4: Number of observations per transition.

ID	from	to	entry	time	status
1	2	5	93.9	243.985	1
1	2	1	93.9	243.985	0
10	2	5	0	240.051	1
10	2	1	0	240.051	0
10	5	2	240.051	615.273	1
10	5	1	240.051	615.273	0

Table 6.5: Example rows of the new, modified dataset.

in the data frame used in `R-INLA`. That is, we cannot assume a parametric shape for the hazard function h_{51} to enter the `R-INLA` code. Hence, there are only three remaining hazard functions: h_{52}, h_{25} and h_{21}.

A table with the numbers of observations for each transition can be viewed in Table 6.4. We present example rows of the new, modified dataset in Table 6.5. We omit including covariates for ease of presentation. Now there is an additional column `status`, which indicates whether the transition happened or not. Especially, we see that patient `1`'s first original row is deleted, where they originally went from $3 \to 4$. State 4 was also recoded as state 2, so now the only transition for patient `1` is $2 \to 5$ with `status`$= 1$. The time is also left-truncated. An additional row is added since patient `1` was at risk of experiencing a $2 \to 1$ transition, which did not happen. This row's `status` indicator is therefore 0.

ID	from	to	entry	time	status
28	2	5	207.733	593.58	1
28	2	1	207.733	593.58	0
28	5	1	593.58	611.059	0
28	5	2	593.58	611.059	0
28	2	5	633.882	690.072	0
28	2	1	633.882	690.072	0
28	2	5	980.455	2485.051	1
28	2	1	980.455	2485.051	0
28	5	2	2485.051	2574.85	1
28	5	1	2485.051	2574.85	0

Table 6.6: Patient 28 in the new, modified dataset.

Another thing to notice about this new, modified dataset is that some patients' journey through the states are inconsistent. This is explained best with an example. We follow patient `28` in Table 6.6. We once again omit including covariates for ease of presentation. We observe patient `28` go from $2 \to 5$ at time $t = 593.580$. The next five rows contain censored information, as the patient did not experience those transitions (`status` $= 0$). Then, we observe the patient *again* making a $2 \to 5$ at time $t = 2485.051$. Suddenly, the patient is back at state 2, without having observed any transition back to state 2. This can be explained by looking at the original dataset for patient `28` (the original dataset for `28` is not presented). It can be seen that the patient transitioned back and forth into state 3, eventually from $3 \to 2$, which was deleted altogether. That is how patient `28` came back to state 2 without that information being kept in the new, modified dataset. Recoding the dataset to three states instead of five cost us some consistency, but it was necessary to simplify the model.

6.3 Preparing the dataset for INLA

Using the original scale of the event times and covariates `Age` and `Mass` makes `R-INLA` crash. For numerical stability, it is generally better to have smaller numbers. Therefore, we first divide all entry and event times by 60 to make seconds into minutes. Then, we scale them to take values between 0 and 1. We do this by finding the largest observed event time in minutes and dividing all time entries by that maximum number. Finally, to reduce the size of the covariate values, we standardize them. *Standardization* is a common process of putting covariates on the same scale and is attained by subtracting the mean and dividing by the standard deviance for each observed value of the covariate. For example, for each patient i in the dataset, the standardization of covariate `Age` is attained in the following way:

$$\frac{\text{Age}_i - \text{mean(Age)}}{\text{sd(Age)}} \tag{6.1}$$

Standardizing covariate `Mass` is done with the same procedure. The binary covariates `Gen` and `res` do not have to be standardized. We use the mean age 14.11, and the standard deviation 4.5. The mean mass is 49.52, and the standard deviation is 23.41. The standardized covariates along with the scaled entry and event times now replace the unstandardized and unscaled values in the dataset, now called the *standardized dataset*. The standardized dataset is a 298×12 data frame, consisting of the columns `ID`, `initial`, `last`, `from`, `to`, `entry`, `time`, `Age`, `Gen`, `Mass`, `res` and `status`. Note that all $5 \to 1$ censorings are included in this standardized dataset, but we omit including this transition in the preparation dataset for INLA.

When fitting the `R-INLA` model, a challenge was figuring out which parametric assumptions to make for each of the three transitions $5 \to 2$, $2 \to 5$, and $2 \to 1$. In this thesis, we limit ourselves to the Weibull and the exponential baseline functions. We tried different model combinations and present three of them. A summary table presenting which model uses which parametric

	Model 1	Model 2	Model 3
$h_{52}(t)$	Weibull	Exponential	Exponential
$h_{25}(t)$	Weibull	Weibull	Exponential
$h_{21}(t)$	Weibull	Weibull	Weibull

Table 6.7: Parametric assumptions used in the different models.

assumption is presented in Table 6.7.

6.3.1 Model 1

Choosing only Weibull models for all the baseline hazards (called model 1 from now on) results in the following hazard specifications:

$$
\begin{aligned}
h_{52}(t;Z) &= \alpha_{52} t^{\alpha_{52}-1} \exp(\eta^{52}),\\
h_{25}(t;Z) &= \alpha_{25} t^{\alpha_{25}-1} \exp(\eta^{25}),\\
h_{21}(t;Z) &= \alpha_{21} t^{\alpha_{21}-1} \exp(\eta^{21}),
\end{aligned}
$$

where $\eta^{lj} = \beta_0^{lj} + \beta_{\text{Age}}^{lj} Z_{\text{Age}} + \beta_{\text{Gen}}^{lj} Z_{\text{Gen}} + \beta_{\text{Mass}}^{lj} Z_{\text{Mass}} + \beta_{\text{res}}^{lj} Z_{\text{res}}$. The summaries of the estimated model parameters can be seen in Table 6.8. Figure 6.5 shows a plot of the estimated 95% CIs. The estimated hazard rate for all transitions is presented in Figure 6.6.

Immediately, we see from Figure 6.5 that all credible intervals for transition $5 \to 2$ are larger than the others. This is due to the smaller number of transitions from $5 \to 2$, compared to $2 \to 5$ and $2 \to 1$. There are only 29 rows of information in the $5 \to 2$ data frame, and 23 of those are observed events. As a comparison, there are 120 rows of information in the $2 \to 5$ and $2 \to 1$ data frames, where 52 and 41 events occurred, respectively. The CI for β_{Gen}^{52} is also much wider than the others. It can be

due to there being almost twice as many observations for the gender-coded as 0 when `status`=1 than for the gender-coded as 1 when `status`=1. For `status`=1, there are eight observations of `Gen`=1 and 14 observations of `Gen`=0.

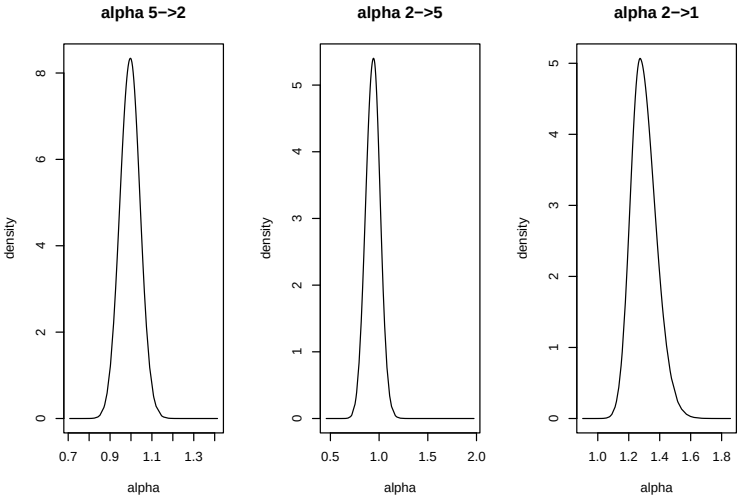

Figure 6.4: Posterior density for the three Weibull hyperparameters.

Note that the α_{52} and α_{25} values are centered around 1, as can be seen in Figure 6.4. This can point to the baseline hazards for h_{52} and h_{25} being exponential models (a Weibull model with $\alpha = 1$ is the exponential model). The CI bands of the estimated hazard for h_{52} are also quite large. In hopes of reducing the CIs for transition $5 \to 2$, we present an analysis using the exponential baseline model for transition $5 \to 2$, which we call model 2. We also present a model where in addition to transition $5 \to 2$, transition $2 \to 5$ is exponential, called model 3.

6.3.2 Model 2

As we have seen, it is fair to assume an exponential model for the baseline hazard of hazard function h_{52}. We keep the Weibull

	mean	0.025quant	0.975quant	mode
β_0^{52}	3.038	1.277	5.026	3.136
β_0^{25}	1.571	1.078	2.024	1.593
β_0^{21}	0.92	0.179	1.595	0.954
β_{Age}^{52}	0.266	-1.82	2.563	0.228
β_{Age}^{25}	-0.072	-0.435	0.292	-0.072
β_{Age}^{21}	0.251	-0.166	0.658	0.257
β_{Gen}^{52}	-0.526	-5.115	3.535	-0.441
β_{Gen}^{25}	-0.382	-0.948	0.167	-0.373
β_{Gen}^{21}	0.07	-0.55	0.692	0.069
β_{Mass}^{52}	-0.257	-1.922	1.25	-0.23
β_{Mass}^{25}	0.448	0.132	0.753	0.454
β_{Mass}^{21}	-0.299	-0.752	0.133	-0.287
β_{res}^{52}	-0.099	-2.061	1.78	-0.109
β_{res}^{25}	-0.43	-0.994	0.132	-0.43
β_{res}^{21}	0.605	-0.049	1.304	0.581
α_{52}	0.995	0.902	1.089	0.997
α_{25}	0.937	0.794	1.08	0.942
α_{21}	1.3	1.158	1.476	1.274

Table 6.8: Estimated parameter values for model 1.

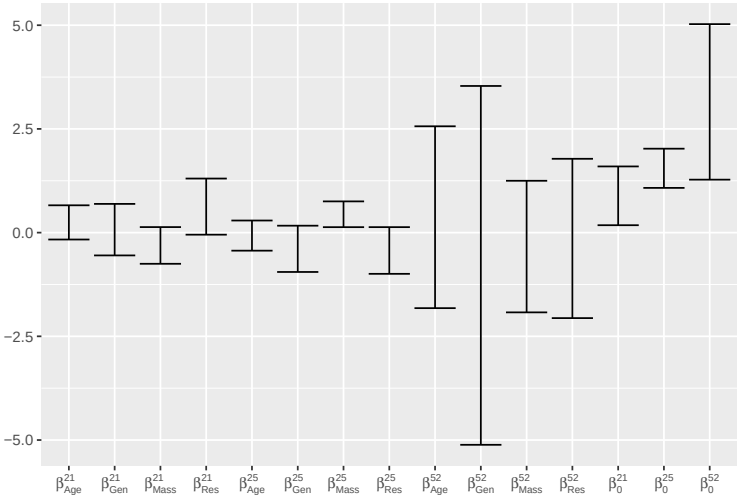

Figure 6.5: Estimated credible intervals for the parameters of model 1.

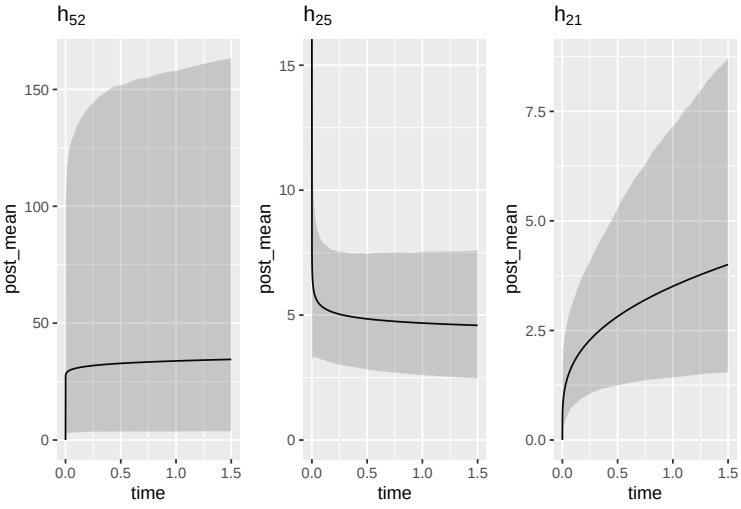

Figure 6.6: Estimated hazard functions for model 1.

assumption for h_{25} and h_{21}. The hazard specifications are:

$$
\begin{aligned}
h_{52}(t;Z) &= \exp(\eta^{52}),\\
h_{25}(t;Z) &= \alpha_{25} t^{\alpha_{25}-1} \exp(\eta^{25}),\\
h_{21}(t;Z) &= \alpha_{21} t^{\alpha_{21}-1} \exp(\eta^{21}),
\end{aligned}
$$

where η^{lj} is as explained in the section above. Note that we now have one less α value to estimate. The estimated model parameters can be seen in Table 6.9. The estimated hazard rate for all transitions is presented in Figure 6.7.

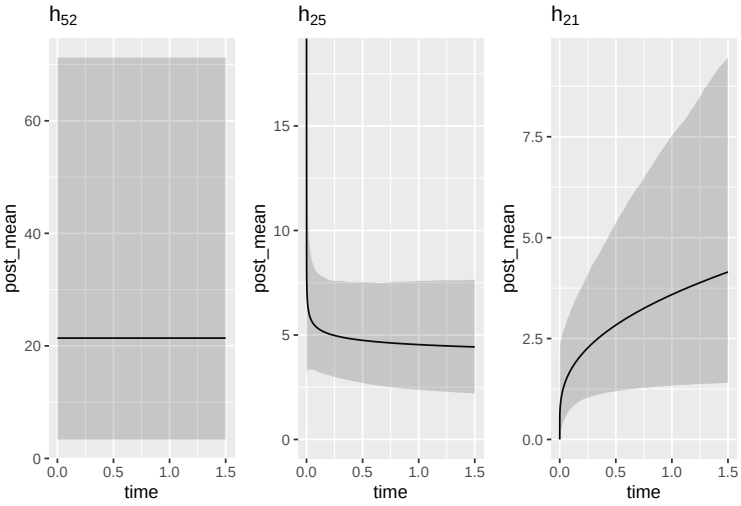

Figure 6.7: Estimated hazard functions for model 2

6.3.3 Model 3

For good measure, we present a third model where transitions $5 \to 2$ and $2 \to 5$ have an exponential baseline model, and $2 \to 1$ is Weibull. The hazard specifications are:

	mean	0.025quant	0.975quant	mode
β_0^{52}	2.791	1.196	4.272	2.951
β_0^{25}	1.553	1.052	2.014	1.574
β_0^{21}	0.924	0.169	1.613	0.959
β_{Age}^{52}	-0.055	-1.918	1.653	-0.15
β_{Age}^{25}	-0.071	-0.434	0.292	-0.071
β_{Age}^{21}	0.253	-0.166	0.662	0.258
β_{Gen}^{52}	0.067	-3.303	3.731	0.341
β_{Gen}^{25}	-0.382	-0.948	0.168	-0.373
β_{Gen}^{21}	0.07	-0.55	0.693	0.069
β_{Mass}^{52}	-0.026	-1.266	1.325	0.041
β_{Mass}^{25}	0.447	0.132	0.752	0.453
β_{Mass}^{21}	-0.3	-0.754	0.133	-0.288
β_{res}^{52}	0.076	-1.491	1.689	0.076
β_{res}^{25}	-0.432	-0.995	0.13	-0.432
β_{res}^{21}	0.606	-0.049	1.306	0.583
α_{25}	0.927	0.758	1.094	0.937
α_{21}	1.306	1	1.658	1.288

Table 6.9: Estimated parameter values for model 2.

	mean	0.025quant	0.975quant	mode
β_0^{52}	3.07	1.8	4.614	2.943
β_0^{25}	1.635	1.162	2.061	1.658
β_0^{21}	0.925	0.167	1.617	0.959
β_{Age}^{52}	0.302	-0.964	2.067	-0.169
β_{Age}^{25}	-0.072	-0.436	0.292	-0.073
β_{Age}^{21}	0.253	-0.166	0.662	0.258
β_{Gen}^{52}	-0.635	-4.11	1.81	0.37
β_{Gen}^{25}	-0.384	-0.951	0.165	-0.376
β_{Gen}^{21}	0.07	-0.55	0.693	0.069
β_{Mass}^{52}	-0.285	-1.565	0.638	0.055
β_{Mass}^{25}	0.45	0.131	0.756	0.456
β_{Mass}^{21}	-0.3	-0.755	0.133	-0.288
β_{res}^{52}	-0.165	-1.778	1.245	-0.004
β_{res}^{25}	-0.423	-0.987	0.141	-0.422
β_{res}^{21}	0.606	-0.05	1.307	0.583
α_{21}	1.299	1.005	1.648	1.272

Table 6.10: Estimated parameter values for model 3.

$$
\begin{aligned}
h_{52}(t;Z) &= \exp(\eta^{52}),\\
h_{25}(t;Z) &= \exp(\eta^{25}),\\
h_{21}(t;Z) &= \alpha_{21} t^{\alpha_{21}-1} \exp(\eta^{21}),
\end{aligned}
$$

We present the estimated hazard functions in Figure 6.8, and the parameter estimates are presented in Table 6.10.

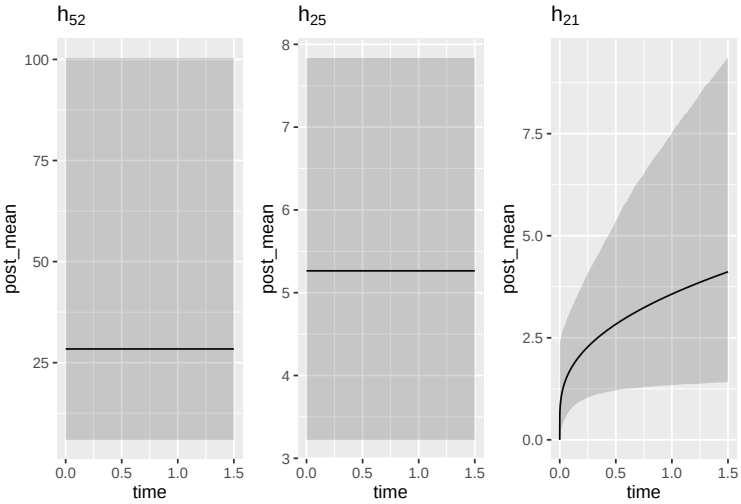

Figure 6.8: Estimated hazard functions for model 3

6.3.4 Comparing models

We could have presented plots similar to Figure 6.5 for models 2 and 3 as well, but we choose to present Figure 6.9 instead. The reason for this is that all three models estimate the same CIs for all covariates for transitions $2 \to 5$ and $2 \to 1$. The only transition where the CIs vary is transition $5 \to 2$. Therefore, we present the CIs each model generates for each parameter estimate. This way, we see that the CIs for model 1 is generally larger than for model 2 and 3. Each model is color-coded, which can be seen in the plot's legend. There does not seem to be a significant decrease between models 2 and 3 in estimating the CIs.

We will use the two measures *deviance information criterion* (DIC) (Spiegelhalter et al. (2002)) and *Watanabe-Akaike information criterion* (WAIC) (Watanabe (2010)) to evaluate and compare our Bayesian models (Gelman, Hwang, and Vehtari (2013)). Both are Bayesian measures for complexity and fit,

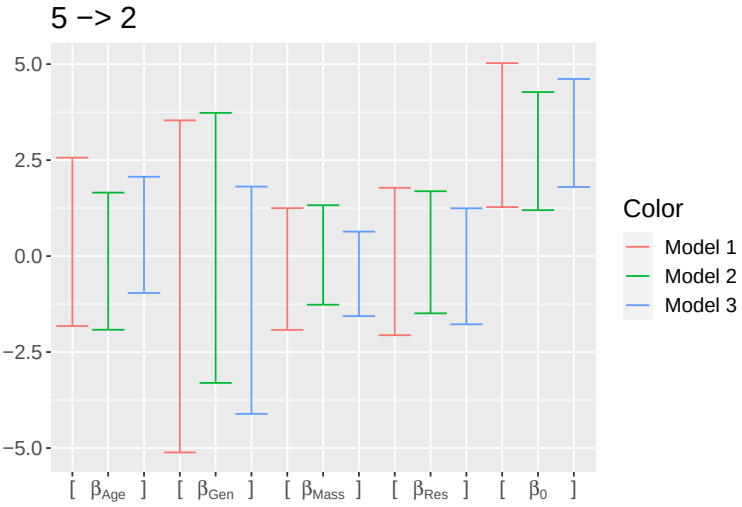

Figure 6.9: CIs for transition 5 to 2, compared for each model.

used for comparing complex hierarchical models. `R-INLA` computes these measures if we tell it to, by including `dic=TRUE` and `waic=TRUE` in `control.compute=list()`. We compare six models. Model 1 is the first model we considered, with all Weibull baseline hazards. Model 2 is the model where we assumed an exponential baseline hazard for transition $5 \to 2$, with the rest of the baseline hazards being Weibull. Model 3 is the model where we assume exponential baseline hazards for both h_{52} and h_{25}, but keep the Weibull baseline for h_{21}. Model 4 does not have any covariates and only includes an intercept term in the linear predictor per transition. Model 5 includes the covariate Z^{25}_{Mass} only for transition $2 \to 5$, as it was the only covariate whose 95% CI did not cover 0. Lastly, model 6 includes the covariate Z_{Mass} for all three transitions. Models 4, 5, and 6 use the same baseline hazards used in model 3. Each model's DIC and WAIC values are presented in Table 6.11. Since the general rule is to choose the model with the smallest DIC and WAIC, we conclude that model 5 is preferred for this data. More information about

	DIC	WAIC
Model 1	-68.37	181.4
Model 2	-84.5	589.97
Model 3	-88.03	-8.55
Model 4	-105.02	-104.67
Model 5	-112.47	-111.67
Model 6	-108.65	-108.94

Table 6.11: DIC and WAIC values for the different models.

the information criteria for Bayesian models can be found in Gelman, Hwang, and Vehtari (2013).

6.4 Discussing the results

The significance of variables can be deduced by examining the overlap of their 2.5% and 97.5% posterior estimates with zero. Figure 6.5 and Table 6.8 can make us question whether the covariates have any effect on the model at all. All CIs for the fixed effects parameters, except for β^{25}_{Mass}, cover 0. If a parameter's CI overlaps zero, there is reason to believe the fixed effect of said parameter to be redundant to the model. Hence, the real-life data analysis results show that the covariates `Gen`, `Age`, `Mass`, and `res` do not significantly influence the hazard rate for children and adolescents experiencing death after a cardiac arrest. We did find a small significance in the covariate `Mass` for the transition $2 \to 5$, which is the recoded transition from PEA/asystole to ROSC. The 95% CI of this covariate did not cover 0, but it was very close. However, when comparing the model without covariates to the model with only Z^{25}_{Mass} included in Figure 6.11, we saw an improvement in the model in terms of decreased DIC and WAIC. Note that the "risk of transition" here is "experiencing the event ROSC," which has a positive connotation. Hence, we find that having a higher mass might

increase the chance of returning to spontaneous circulation, i.e., surviving a cardiac arrest. Note that this is a finding from the recoded dataset from five states to three, and the significance is barely there. To make any further conclusions about the specific interpretations of health data would require professional help. Even though we did not find any significant covariates, one should be careful saying that these covariates are *unimportant* in a larger context, as the results of this analysis are heavily sample-based. The sample size was small, and a larger portion of the patients had a respiratory cause for the cardiac arrest instead of circulatory. We conclude that the small dataset limits the findings' generalizability.

Chapter 7

Discussion

Here we presented INLA as the inferential scheme for complex multistate models. Building on the work done by Niekerk, Bakka, and Rue (2019), we showed how we could present multistate models as LGMs to fulfill the assumptions needed for INLA to be applied. We presented two fully parametric simulation studies; a competing risks and a transient model. The simulation studies showed that the true values were estimated well by using INLA. The motivation behind this thesis was a dataset on children and adolescents suffering from cardiac arrest, on which we used our model. The original five-state multistate model was complex, so we simplified it to a three-state model. Note that the same methods described in this thesis can of course be used on the original five-state model. Due to the amount of data being so small, the results were uncertain. The covariates included did not play a significant role in hazard estimation, but we cannot say these covariates are unimportant in general given the small dataset. More research needs to be done with a larger dataset to get valuable results properly.

A semi-parametric model can be used for the hazards if a fully parametric baseline is unreasonable. The results in Skogvoll et al. (2020) show that they estimated a hazard function with a

different shape than an exponential or Weibull baseline model can produce. If there is strong reason to believe a non-parametric shape for the baseline hazard, we direct the interested reader to the paper by Martino, Akerkar, and Rue (2011), where they show that semi-parametric hazard models also can be fit using INLA.

Throughout the simulations and handling of the real-life data, we only used fixed effects $\mathbf{Z}_i$ in the linear predictor η in Equation (2.15). Extending the linear predictor to include frailty effects is possible using `R-INLA`. Including a frailty term means we can consider dependent failure times generated as conditionally independent times given the frailty. We investigated the possibilities of including such a term in our simulated models and the real-life data, but due to lack of time, we decided against it. Since we do not have a frailty term, each row in the dataset is treated as a new individual. However, since there are repeated events for the same individual in our datasets, we recognize that including a frailty term would result in a better model. We highly recommend that this is done in future works and extensions for this particular application.

We used a Bayesian approach to analyze multistate models and have mentioned prior distributions. All the Weibull hyperparameters used unspecified PC priors, the default priors in `R-INLA`. Researching other prior specifications was not a major focus of this project, and as the chosen priors gave reasonable results, we did not find it necessary to change them. However, if more research is done on Bayesian inference for multistate data regarding cardiac arrest in children and adolescents, different prior specifications can help improve the models.

A great source of confusion throughout the research of this thesis has been independence between competing risks and random censoring. I started my research by reading the book Applied Survival Analysis using R (Moore (2016)) and The Statistical Analysis of Failure Time Data (Kalbfleisch and Prentice (2002)), as those were the books provided in my first survival analysis course. These books were great as a way to refresh key concepts

in survival analysis but appear confusing once presenting the topic of competing risks and transient models. In Chapter 9.2 of (Moore (2016)), it is argued that a key assumption about censoring is that it is *independent* of the event in question, which may be questionable in most competing risks applications. They further show how selecting each event as the primary event and treating the other as a censoring event leads to biased estimates of survival curves. This raised an important issue because this was exactly what I was doing in my simulation of competing events. At that time, I was following the algorithm of the book by Beyersmann, Allignol, and Schumacher (2012) and following the R code from Niekerk, Bakka, and Rue (2019). It seemed like the entire assumption on which my model was built was false. After weeks of reading different sections of all three books and several meetings with my supervisors, we finally seemed to figure out that there are primarily two different modeling approaches, namely the cause-specific hazards approach and the subdistribution hazards approach. Moore (2016) and Kalbfleisch and Prentice (2002) seem to favor the subdistribution approach, while Beyersmann, Allignol, and Schumacher (2012) present both. Beyersmann, Allignol, and Schumacher (2012) argue that the cause-specific approach is a valid analytic approach, which is the approach that allows treating each event as the primary event, in turn, censoring the other competing events. They note that it is vital to analyze all cause-specific hazards, or the analysis will be incomplete. This is also the approach Niekerk, Bakka, and Rue (2019) uses, on which our work is based. Debates surrounding the choice between the two approaches are still ongoing. The consensus seems to be that the cause-specific hazards approach should be chosen for causal inference of covariate effects, which was one of our main goals.

Beyersmann, Allignol, and Schumacher (2012) also addresses the interesting question about competing risks being independent themselves. Let us consider two competing risks only, for ease of presentation, where at an event time T, either $X_T = 1$ or $X_T = 2$ occurs. They argue that the concept of statistical independence does not apply since the different failure causes

are simply different values of exactly *one* random variable, X_T. Any further discussion introduces the topic of *latent failure times* which we have not touched on in this thesis. The interested reader is directed to Beyersmann, Allignol, and Schumacher (2012) (Ch. 5.2.2 and Ch. 7.2) and Moore (2016) (Ch. 9.2.1).

Another theme of discussion throughout the thesis was the Markov assumption in the illness-death model with recovery. We use the words "healthy" for state 0, "ill" for state 1, and "dead" for state 2. The baseline hazard for getting "ill" is assumed to be the same regardless of how many times one has been ill. One might think this is an over-simplified model, and the hazard of dying should be greater given that an individual has been ill several times. We can achieve this by including time-dependent covariates that carry the information on the number of times one has been "ill" before or how long one has been "ill" in the past. However, time-dependent covariates pose some interpretational challenges addressed in Beyersmann, Allignol, and Schumacher (2012) (Chapter 11), and is still an ongoing research field.

References

Altshuler, Bernard. 1970. "Theory for the Measurement of Competing Risks in Animal Experiments." *Mathematical Biosciences* 6: 1–11. https://doi.org/https://doi.org/10.1016/0025-5564(70)90052-0.

Andersen, P. K., and R. D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *The Annals of Statistics* 10 (4): 1100–1120. http://www.jstor.org/stable/2240714.

Balan, and Putter. 2020. "A Tutorial on Frailty Models" 29. https://doi.org/10.1177/0962280220921889.

Beyersmann, Allignol, and Schumacher. 2012. *Competing Risks and Multistate Models with r*. 1st ed. Springer-Verlag New York.

Cox. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society. Series B (Methodological)* 34 (2): 187–220. http://www.jstor.org/stable/2985181.

Fine, and Gray. 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association* 94 (446): 496–509. http://www.jstor.org/stable/2670170.

Gamerman, and Lopes. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*.

Gelman, Andrew, Jessica Hwang, and Aki Vehtari. 2013. "Understanding Predictive Information Criteria for Bayesian Models." https://arxiv.org/abs/1307.5928.

He, Peng, Frank Eriksson, Thomas H. Scheike, and Mei-Jie

Zhang. 2016. "A Proportional Hazards Regression Model for the Subdistribution with Covariates-Adjusted Censoring Weight for Competing Risks Data." *Scandinavian Journal of Statistics* 43 (1): 103–22. http://www.jstor.org/stable/24886939.

Hoel, David G. 1972. "A Representation of Mortality Data by Competing Risks." *Biometrics* 28 (2): 475–88. http://www.jstor.org/stable/2556161.

Kalbfleisch, John D., and Ross L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. 2nd ed. Wiley, New York.

Kay, Richard. 1982. "The Analysis of Transition Times in Multistate Stochastic Processes Using Proportional Hazard Regression Models." *Communications in Statistics - Theory and Methods* 11 (15): 1743–56. https://doi.org/10.1080/03610928208828346.

Martino, Akerkar, and Rue. 2011. "Approximate Bayesian Inference for Survival Models." *Scandinavian Journal of Statistics, Vol. 38: 514-528*.

Martino, and Riebler. 2019. "Integrated Nested Laplace Approximations (INLA)." https://arxiv.org/abs/1907.01248.

Moeschberger, M. L., and H. A. David. 1971. "Life Tests Under Competing Causes of Failure and the Theory of Competing Risks." *Biometrics* 27 (4): 909–33. http://www.jstor.org/stable/2528828.

Moore, Dirk. 2016. *Applied Survival Analysis Using r*. https://doi.org/10.1007/978-3-319-31245-3.

Niekerk, Janet van, Haakon Bakka, and Haavard Rue. 2019. "Competing Risks Joint Models Using r-INLA." https://arxiv.org/abs/1909.01637.

Nordseth, Niles, Eftestøl, Sutton, Irusta, Abella, Berg, Nadkarni, and Skogvoll. 2019. "Rhythm Characteristics and Patterns of Change During Cardiopulmonary Resuscitation for in-Hospital Paediatric Cardiac Arrest." *Resuscitation* 135: 45–50. https://doi.org/https://doi.org/10.1016/j.resuscitation.2019.01.006.

Prentice, Kalbfleisch, Peterson, Flournoy, Farewell, and Breslow. 1978. "The Analysis of Failure Times in the Presence of

Competing Risks." *Biometrics* 34 (4): 541–54. http://www.jstor.org/stable/2530374.

"Pulseless Electrical Activity Asystole." n.d. https://nhcps.com/lesson/acls-cases-pulseless-electrical-activity-asystole/.

Rue, and Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. 1st ed. https://doi.org/10.1201/9780203492024.

Rue, Martino, and Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society Series B* 71 (April): 319–92. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Scheike, and Zhang. 2007. "Direct Modelling of Regression Effects for Transition Probabilities in Multistate Models." *Scandinavian Journal of Statistics* 34 (1): 17–32. https://doi.org/https://doi.org/10.1111/j.1467-9469.2006.00544.x.

Simpson, Daniel, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. 2017. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32 (1): 1–28. https://doi.org/10.1214/16-STS576.

Skogvoll, Nordseth, Sutton, Eftestøl, Irusta, Aramendi, Niles, et al. 2020. "Factors Affecting the Course of Resuscitation from Cardiac Arrest with Pulseless Electrical Activity in Children and Adolescents." *Resuscitation* 152: 116–22. https://doi.org/https://doi.org/10.1016/j.resuscitation.2020.05.013.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4): 583–639. https://doi.org/https://doi.org/10.1111/1467-9868.00353.

van Niekerk, J., H. Bakka, and H. Rue. 2021. "A Principled Distance-Based Prior for the Shape of the Weibull Model." *Statistics & Probability Letters* 174: 109098. https://doi.org/https://doi.org/10.1016/j.spl.2021.109098.

"Ventricular Fibrillation and Pulseless Ventricular Tachycardia." n.d. https://nhcps.com/lesson/acls-cases-ventricular-fibrillation-pulseless-ventricular-tachycardia/.

Watanabe. 2010. "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *J. Mach. Learn. Res.* 11 (December): 3571–94.

Wulfsohn, Michael S., and Anastasios A. Tsiatis. 1997. "A Joint Model for Survival and Longitudinal Data Measured with Error." *Biometrics* 53 (1): 330–39. http://www.jstor.org/stable/2533118.

NTNU

Norwegian University of
Science and Technology