Mathematical Institute

Master Thesis

Statistical Science for the Life and Behavioural Sciences

Investigating hospital heterogeneity with a competing risks frailty model

Author:
Anja Juana Rüten-Budde

First Supervisor:
Dr. Marta Fiocco
LUMC & LU

Second Supervisor:
Prof. dr. Hein Putter
LUMC

December 2015

Leiden University Medical Center

Abstract

Survival analysis is used in the medical field to identify the effect of predictive variables and treatment on time to a specific event. Generally not all variation of survival time can be explained by observed covariates. The effect of unobserved covariates on the risk of a patient is called frailty.

In multi-center studies the unobserved center effect induces frailty on its patients. Ignoring the effect of unobserved covariates leads to selection bias over time, since patients with low frailties remain in the risk set longer. It is common practice to account for this by including a random frailty term in the model. Particularly in multi-center studies a random center effect is included in the model.

In competing risks situations more than one type of event is possible. A frailty variable representing the center effect can be incorporated in the analysis independently for each event. However, in the medical context events representing disease progression are likely to be related and this correlation is missed in the independent frailty model.

In this thesis an additive gamma frailty model to allow for correlation between frailties in a model with two competing events is proposed. An instance of the litter frailty model described in Petersen et al. (1996) is used to model correlation between frailties at center level. Correlation between frailties indicates the common center effect on both events and measures how closely the risks are related.

In this work it is illustrated how to estimate the model using the expectation maximization algorithm and a method to estimate the standard error is described.

The model is applied to a data set from a multi-center clinical trial, and results are evaluated. Hospitals are compared by employing empirical Bayes estimates together with the corresponding confidence intervals.

The model seems to be a useful tool to investigate heterogeneity between centers by discriminating common and separate center effects.

Acknowledgement

I like to thank first and foremost my two supervisors Marta Fiocco and Hein Putter. Marta was a mentor for me this year and supported me in giving my very best. She has been my favorite and most influential teacher and is a great role model in work and life. Hein was essential to help me understand the theoretical parts of this thesis and was always very helpful in any questions I had.
I gratefully thank everybody involved in creating the master track *Statistical Science for the Life and Behavioural Sciences*. I benefited immensely from the strong and versatile program.
A special thanks my friends and boyfriend who kept me company during long days in the library. Last but not least, I like to thank my family for supporting me in every aspect of my life.

Contents

Chapter 1

Introduction

1.1 Goals of this thesis

For large studies or when incidence of disease is low data is often collected from multiple treatment centers. Treatment outcome may differ between centers and even after adjusting the analysis for disease specific risk factors possible heterogeneity between centers may remain undetected. If this aspect is ignored in the analysis, results might be biased. Common practice is to adjust for heterogeneity between centers by incorporating a random variable representing the center effect into the model. In situations with multiple events a correlation structure for the random center effect is desired.

The goal of this thesis is to model heterogeneity between centers with correlated frailties for two competing events. This is achieved by incorporating a multivariate frailty component into a cause-specific hazard regression model, as proposed by Petersen et al. (1996). The frailties act on the cause-specific hazards, and frailty correlation indicates whether transitions from the initial state to the competing events are associated. A competing risks frailty model with correlated gamma frailty distribution, where each frailty is associated with a specific transition, is studied.

The method will be illustrated with a data set from a multi-center clinical trial on breast cancer by the European Organisation for Research and Treatment of Cancer. Two competing events are studied here, local recurrence vs distant metastasis or death. Heterogeneity among centers will be presented by calculating the empirical Bayes estimates of the frailties of different causes for each individual center and their associated confidence intervals.

1.2 Structure

In Chapter 2 a short introduction to survival analysis is given, including competing risks and frailty models. In Chapter 3 a motivating data example will be presented and currently available methods to analyze competing risks data will be applied.

The novelty of the thesis is presented in Chapter 4, where the proposed competing risks frailty model with correlated frailties is introduced. Chapter 5 and 6 discuss the

estimation of the model and its standard error respectively. The results of the model applied to a data set from a multi-center clinical trial are presented in Chapter 7. Discussion follows in Chapter 8.

Chapter 2

Background theory

Survival analysis is a branch of statistics that studies time to event data. Such data arises where interest lies on the time it takes for a specific event to occur. The most prominent applications are found in the medical field, where e. g. time from diagnosis of disease until death could be studied. Survival analysis specifically studies the distribution of time between an initial event (e. g. diagnosis of disease) and the occurrence of an event of interest (e. g. death), called *event* or *failure*, even though it might be a success (e. g. recovery).

In this chapter we introduce some basic concepts used in survival analysis in Section 2.1. Competing risks models are introduced in Section 2.2 and the frailty model used in this thesis is outlined in Section 2.3.

2.1 Introduction to survival analysis

2.1.1 Survival time distribution

Different representations of the distribution of time to event are used in survival analysis.

The following equivalent definitions of the distribution of time to event are based on Klein and Moeschberger (2003, chap. 2).

Survival function

Let T be a random variable representing the time from a time origin to the occurrence of the event of interest. In the following T is constricted to be continuous, with probability density function $f(t)$ and cumulative distribution function $F(t) = P(T \leq t)$. The *survival function* at time t is the probability for a random individual to survive until time t and is defined as follows

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^\infty f(s)ds. \tag{2.1}$$

The survival function is monotone, non-increasing, equal to one at time zero and converging to zero as time approaches infinity.

Hazard function

The *hazard function* or *hazard rate* $\lambda(t)$ is the instantaneous rate of failure at time t for a random individual still event-free. In other words, it is the probability of failing in the next instant, conditional on being event-free just before time t and is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}. \tag{2.2}$$

The *cumulative hazard* is defined as

$$\Lambda(t) = \int_0^t \lambda(s) ds = -\log S(t). \tag{2.3}$$

The survival function can further be expressed by

$$S(t) = e^{-\Lambda(t)}. \tag{2.4}$$

Expression (2.4) shows that the survival and the hazard function provide equivalent characterizations for the distribution of T.

Cumulative incidence function

The *cumulative incidence function* is the probability of failing before time t and corresponds to the distribution function $F(t)$. It can be expressed in terms of the hazard and survival function as

$$I(t) = \int_0^t \lambda(s) S(s) ds. \tag{2.5}$$

2.1.2 Censoring and truncation

Survival data is generally incomplete. Since it takes time till event occurrence, it is usually not possible to collect the complete information. This missing information can be classified into two groups called *censoring* and *truncation* (Klein and Moeschberger, 2003, chap. 3).

The survival time of an individual is censored, if the event was not observed. Let T denote the event time. The survival time is *right-censored*, if the event did not occur before the end of the observation period. Instead of T only censoring time C_r, which corresponds to the end of study or end of follow-up is observed. Data can be represented as a pair of random variables (M, δ), where $M = \min(T, C_r)$ and δ is the event indicator, equal to one if the event was observed and zero otherwise.

The survival time of an individual is *left-censored*, if the event had occurred at an unknown time before entering the study. Analogous to right-censored data, data can be represented as a pair of random variables (M, δ), where δ is the event indicator, $M = \max(T, C_l)$ and C_l is the censored time, which corresponds to the start of the observation period. The survival time of an individual is *interval-censored*, if the event occurred within a certain time interval, but it is not further known when.

The survival time of an individual is truncated if he was included in the study conditional on his event time lying within a certain time interval (M_L, M_R). The survival time of an individual is *left-truncated*, if $M_L > 0$ and he was observed because his event time is larger than M_L. The survival time of an individual is *right-truncated*, if $M_R < \infty$ and he was observed because his event time is smaller than M_R. The survival time of an individual is *interval-truncated*, if $M_L > 0$ and $M_R < \infty$ and he was observed because his event time lies within this interval.

2.1.3 Likelihood construction

The *likelihood function* gives the probability of the data conditional on knowing the correct survival function or other function of equivalent information. In the remainder of this thesis only right-censored data will be discussed. The likelihood for right censored observations can be written as

$$L = \prod_{i=1}^{n} [\lambda(t_i)]^{\delta_i} S(t_i), \tag{2.6}$$

where δ_i is the event indicator, $\lambda(t_i)$ the hazard rate and $S(t_i)$ the survival function for individual i (Klein and Moeschberger, 2003, sec. 3.5).

2.1.4 Estimation

There are various approaches to estimate the survival function. While parametric methods assume a particular distribution for the survival time, non-parametric methods can be used without making such assumptions. Two non-parametric approaches are described here (Klein and Moeschberger, 2003, sec. 4.2).

Let $t_1 < t_2 < ... < t_D$ be the ordered event times and d_i the number of events at time t_i. Let Y_i denote the number of individuals at risk at time t_i, which are individuals that have not failed before time t_i. The *product-limit estimator* or *Kaplan-Meier esimator* of the survival function is as follows

$$\hat{S}(t) = \begin{cases} 1, & \text{if } t < t_1 \\ \prod_{t_i \le t} \left(1 - \frac{d_i}{Y_i}\right), & \text{if } t_1 \le t. \end{cases} \tag{2.7}$$

The variance is estimated by Greenwoods's formula

$$\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)}. \tag{2.8}$$

Another estimator can be obtained by first estimating the cumulative hazard with the *Nelson-Aalen estimator*

$$\hat{\Lambda}(t) = \begin{cases} 0, & \text{if } t < t_1 \\ \sum_{t_i \le t} \frac{d_i}{Y_i}, & \text{if } t_1 \le t, \end{cases} \tag{2.9}$$

with variance

$$\hat{V}[\hat{\Lambda}(t)] = \sum_{t_i \leq t} \frac{d_i}{Y_i^2}. \tag{2.10}$$

And then estimating the survival function by $\hat{S}(t) = e^{-\hat{\Lambda}(t)}$.

2.1.5 The proportional hazards model

A popular mathematical model used to incorporate the effect of covariates on survival time is the *Cox proportional hazards model*, which will be defined following Klein and Moeschberger (2003, chap. 8). The hazard at time t of an individual i with covariate vector $\boldsymbol{X}_i$ is assumed to be

$$\lambda(t|\boldsymbol{X}_i) = \lambda_0(t) e^{\boldsymbol{\beta}^T \boldsymbol{X}_i}, \tag{2.11}$$

where $\lambda_0(t)$ is the baseline hazard function, which describes the risk of individuals with covariate vector $\boldsymbol{X}_i = 0$, who serve as reference. The increased risk associated with the covariate vector $\boldsymbol{X}_i$ compared to the reference group is expressed in the *hazard ratio* $e^{\boldsymbol{\beta}^T \boldsymbol{X}_i}$. Note that in this model the increase or decrease in risk is assumed to be the same at all time points t. Model (2.11) is semi-parametric assuming a parametric form for the covariate effects and a non-parametric baseline hazard. The cumulative hazard is

$$\Lambda(t|\boldsymbol{X}_i) = \Lambda_0(t) e^{\boldsymbol{\beta}^T \boldsymbol{X}_i}, \tag{2.12}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ and the relationship to the survival function is

$$S(t|\boldsymbol{X}_i) = S_0(t)^{\exp(\boldsymbol{\beta}^T \boldsymbol{X}_i)}, \tag{2.13}$$

where $S_0(t)$ is the survival function of the reference group.

Under the assumptions of noninformative censoring given covariates, that is the event and censoring time are independent given the covariate vector, and if no tied events occur, based on the hazard function (2.11) a partial likelihood is constructed by

$$L(\boldsymbol{\beta}) = \prod_{i:\delta_i=1} \frac{e^{\boldsymbol{\beta}^T \boldsymbol{X}_i}}{\sum_{R(t_i)} e^{\boldsymbol{\beta}^T \boldsymbol{X}_g}}, \tag{2.14}$$

where δ_i is the event indicator of subject i and $R(t_i) = \{g : t_g \geq t_i\}$ is the set of patients that are still at risk before time t_i. The partial likelihood is the product of probabilities over the event times t_i, that an individual with covariate vector $\boldsymbol{X}_i$ dies at time t_i given one individual in the risk set dies at that time. The product is taken only over the event times, since the baseline hazard is estimated to be zero at nonevent times. Note that the partial likelihood does not depend on the baseline hazard. Conditional on the coefficients $\boldsymbol{\beta}$ the baseline hazard can be estimated by e. g. the Nelson-Aalen estimator.

We conditioned the partial likelihood on continuous time, however in real data sets often ties are present. There are several different methods to address this problem, e. g. by Breslow or Efron (Klein and Moeschberger, 2003, sec. 8.4).

2.2 Competing risks model

Standard survival analysis techniques are used to analyze the distribution of time it takes for a specific type of event to occur from a time origin. In some situations more than one type of end point are possible. Further it could be that other events (e. g. death) will prevent the event of interest to occur. These scenarios can be described by a competing risks model. Definitions and methodology introduced in this section follow Putter et al. (2007).

The competing risks model is illustrated in Figure 2.1. It is described by a starting state (indicated in Figure 2.1 as Alive) and several end states, representing the causes of failure, that can be reached from the starting state.

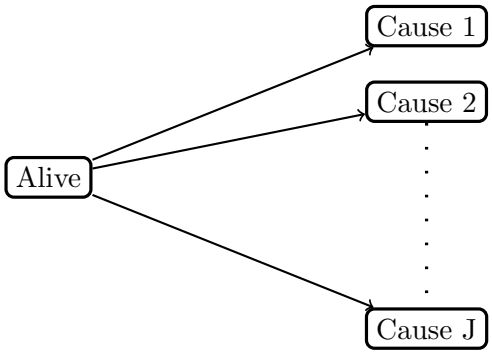

Figure 2.1: Competing risks model with J causes of failure.

An example for a competing risks scenario is the study of different causes of death. Dying from one cause will prevent the occurrence of other events. A different example is the study of time from remission to relapse of disease, where death is a competing event. In other situations only the first event may be of interest. In this case all events will prevent further events to occur.

2.2.1 Approaches to competing risks

In the following only right-censored survival times will be considered.

Let $T_1, T_2, ..., T_J$ be the event times of a random individual for J competing events and n the number of individuals in the study. For each individual only the first event $T = \min(T_1, T_2, ..., T_J)$ and an indicator $\delta = 1, ..., J$ specifying the cause of failure, is observed. The fundamental concept in the competing risks model is the *cause-specific hazards function*, which is the hazard of failing from cause j in the presence of competing events

$$\lambda_j(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(t \leq T < t + \Delta t, \delta = j | T \geq t)}{\Delta t}. \tag{2.15}$$

The cumulative cause-specific hazard and the probability of not failing from any cause

at time t are respectively given as

$$\Lambda_j(t) = \int_0^t \lambda_j(s)ds \tag{2.16}$$

and

$$S(t) = e^{-\sum_{j=1}^{J} \Lambda_j(t)}. \tag{2.17}$$

Expression (2.17) relates the overall survival function to the cause-specific hazards. The cumulative incidence function for cause j is defined as the probability of failing from cause j before time t

$$I_j(t) = \Pr(T \leq t, \delta = j) = \int_0^t \lambda_j(s)S(s)ds. \tag{2.18}$$

Estimation

To estimate the parameters for cause j in a competing risks model, a naive approach would censor observations for which a competing event occurred. In this way the competing risks are ignored and a standard analysis with likelihood function (2.6) can be performed. However, failures from competing risks reduce the number of individuals at risk for the other causes and therefore the number of failures from cause j. The likelihood function for the competing risks model is given as

$$L = \prod_{i=1}^{n} e^{-\sum_j \Lambda_j(t_i)} \prod_{j=1}^{J} \lambda_j(t_i)^{1_{\{\delta_i = j\}}}. \tag{2.19}$$

The cause-specific hazard for cause j can be estimated by counting the number of events of type j divided by the observed number of patients at risk. Let

- d_{ji} : number of patients failing from cause j at t_i
- Y_i : number of patients at risk at t_i
- $R(t) = \{i : t_i \leq t\}$: set of patients with time points before time t

The estimator of the cause-specific hazard for cause j is given as

$$\hat{\lambda}_j(t_i) = \frac{d_{ji}}{Y_i}. \tag{2.20}$$

The estimator for the cause-specific cumulative hazard and cumulative incidence function are respectively given as

$$\hat{\Lambda}_j(t) = \sum_{i \in R(t)} \hat{\lambda}_j(t_i) \tag{2.21}$$

and

$$\hat{I}_j(t) = \sum_{i \in R(t)} \hat{\lambda}_j(t_i)\hat{S}(t_{i-1}), \tag{2.22}$$

where $\hat{S}(t_{i-1})$ is the Kaplan-Meier estimator (defined in (2.7)), the probability of remaining event-free just before time t_i.

2.2.2 Covariate effects

When the effect of covariates on the different causes of failure are of interest, a model analogous to the Cox proportional hazards model is appropriate. The traditional Cox model can be used, to perform separate analysis for each cause of failure, censoring observations which failed from a competing event. The cause-specific hazard of cause j for a subject i with covariate vector $\boldsymbol{X}_i$ is

$$\lambda_j(t|\boldsymbol{X}_i) = \lambda_{j0}(t)e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_i}, \tag{2.23}$$

where λ_{j0} is the cause-specific baseline hazard for cause j and $\boldsymbol{\beta}_j$ assesses the effect of the covariates $\boldsymbol{X}_i$ on the progression rate to cause j.

Estimation

The general form of the likelihood function expressed in terms of cause-specific hazards is

$$L = \prod_{i=1}^{n} e^{-\sum_j \Lambda_{j0}(t_i)\exp(\boldsymbol{\beta}_j^T \boldsymbol{X}_i)} \prod_{j=1}^{J} \left(\lambda_{j0}(t_i)e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_i}\right)^{1_{\{\delta_i=j\}}}, \tag{2.24}$$

where Λ_{j0} is the cause-specific cumulative baseline hazard for cause j and $\delta_i = 1, ..., J$ is the indicator for type of failure for subject i.

The parameters are estimated in the same way as in the classical framework with only one type of end point, by maximizing the partial likelihood (2.14). The difference is in the interpretation: here the effects are quantified on the cause-specific hazard and not on the marginal hazard. Only if the censoring due to the competing risks is noninformative conditionally to the covariates in the model, then the estimates can also be interpreted as effects on the marginal hazard. In the competing risks context it is more appropriate to call the model *proportional cause-specific hazards model* and not a Cox model.

The covariate effects are proportional on the cause-specific hazard. However, the relation with the cumulative incidence function for cause j associated to different values of the same covariate can be unpredictable. The reason is that the effect of a covariate not only influences the hazard of cause j but has also an effect on the other causes of failure, which together contribute to the cumulative incidence function (2.18). As a

consequence, it can happen that the cause-specific hazard is lower for a subgroup over the whole time range, whereas the cause-specific cumulative incidence is higher for part of the time scale. For this reason interpretation of the hazard ratios requires caution.

The covariate effects on the cause-specific hazards are estimated analogous to the single event case for each transition separately. This can be done with readily available software, e. g. in `R` (R Core Team, 2015) with the `survival` package (Therneau, 2015). However, the results must be interpreted in the cause-specific context.

2.3 Frailty model

2.3.1 Introduction

In the medical field the term frailty is often used (Rockwood, 2005). The term comes from the field of gerontology where it is used to indicate that frail people have an increased risk for complication and mortality (Gillick, 2001). This condition can also be referred to as heterogeneity, indicating variation in treatment outcome between patients. Statistical models explain part of this variation by incorporating patient characteristics into the model, however not all variation can be explained by observed covariates.

Unobserved heterogeneity can lead to biased results. Frail individuals will experience the event on average earlier and hence the individuals remaining in the risk set will have lower frailty. This gives a selection bias that reduces the population hazard rate over time.

The concept of frailty provides a proper way to introduce random effects in the model to account for the presence of association and unobserved heterogeneity. The variance of this random component is a measure used to quantify heterogeneity in the data set. Vaupel et al. (1979) discussed univariate frailty models with a gamma distribution and applied this concept to survival. Clayton (1978) used frailties in the multivariate analysis of chronic disease incidence in families.

Frailty or heterogeneity can also be considered at center level. Multi-center studies combine data collected in different centers. A potential problem with this type of data is heterogeneity between centers. The reason for this might be an imbalance of the distribution of patient-specific characteristics (e. g. age) over centers or an effect of center-specific factors (e. g. geography) on the outcome.

Heterogeneity can be addressed by incorporating both patient- and center-specific factors into the model, that explain part of the heterogeneity. Those factors reduce the variance of the random effect, however it is possible that not all heterogeneity can be explained. In spite of treatment protocols there remains some variability due to slightly different interpretation at the treatment centers. A random center effect models this unobserved differences. A survival model with random effect is called a frailty model.

2.3.2 Univariate frailty model

In this subsection we will discuss the multiplicative gamma frailty model following Aalen et al. (2008, chap. 2). In its simplest form, a frailty is an unobserved random factor

varying over the population of individuals, which is assumed to have a multiplicative effect on the hazard of a single individual.

In this model the hazard rate corresponding to subject i with frailty W_i is specified as follows

$$\lambda(t|W_i) = W_i\lambda_0(t), \tag{2.25}$$

where $\lambda_0(t)$ is the baseline hazard. If $W_i > 1$ the individual risk increases and if $W_i < 1$ it decreases.

A convenient choice for the frailty distribution is the gamma distribution, since its posterior distribution given survival data stays in the gamma family. The gamma density function is as follows

$$f(w) = \frac{\eta^\nu}{\Gamma(\nu)} w^{\nu-1} e^{-\eta w}, \tag{2.26}$$

where $\nu > 0$ and $\eta > 0$ are the shape and rate parameter respectively. Since part of the frailty can be absorbed in λ_0 it is typically assumed that the expectation of w equals 1, which in the gamma case means that the parameters are equal ($\nu = \eta$). The variance of the gamma frailty with expectation 1 is $1/\eta$ and it is a natural measure of the degree of heterogeneity in the population.

The model can also be represented by its conditional survivor function

$$S(t|W_i) = e^{-W_i\Lambda_0(t)}, \tag{2.27}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$. The survival $S(t|W_i)$ represents the fraction of individuals surviving until time t given W_i.

Covariate effects

A natural choice to incorporate covariate effects in a frailty model is to extent the Cox proportional hazards model. The frailty acts multiplicative on the baseline hazard, it can however also be seen as entering the regression part of the hazard function the same way as the observed covariates. For subject i with covariate vector $\boldsymbol{X}_i$ and frailty W_i the hazard is as follows

$$\begin{aligned} \lambda(t|\boldsymbol{X}_i, W_i) =& W_i\lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{X}_i} \\ =& \lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{X}_i + \log(W_i)}. \end{aligned} \tag{2.28}$$

Univariate frailty models are not identifiable from survival information alone. Elbers and Ridder (1982) have shown the identifiability of model (2.25) for a finite mean frailty distribution, which in case of survival data is given when covariates are included in a proportional hazards model as in model (2.28). Heckman and Singer (1984) have considered alternative conditions of identifiability.

2.3.3 Multivariate frailty model

In multivariate frailty models several individuals share a frailty, inducing a dependence structure. It is applied in situation in which data are divided into groups, like families or treatment centers. The multivariate frailty model with and without covariates is given by (2.28) and (2.25) respectively, with individuals indexed within their group. In a multicenter study a frailty for the center effect models the dependence of patients treated in the same treatment center and its variance measures the heterogeneity between hospitals.

2.3.4 Estimation

In the following the likelihood function for a multivariate frailty model with frailties at center level will be explained. Let $\boldsymbol{W} = (W_1, ..., W_K)$ be the frailty vector associated to K centers. In this context censoring must be independent from time to event, which is the usual assumption and noninformative for the random vector $\boldsymbol{W}$.

Let n_k and d_k denote the number of patients in hospital k and the number of events in hospital k respectively. Let δ_{ki} be the event indicator for subject i in hospital k. The likelihood function conditional on the observed frailties is

$$L(\boldsymbol{\beta}, \lambda_0 | \text{data}\,, \boldsymbol{W}) = \prod_{k=1}^{K} f(W_k) \prod_{i=1}^{n_k} \left(W_k \lambda_0(t_{ki}) e^{\boldsymbol{\beta}^T \boldsymbol{X}_{ki}} \right)^{\delta_{ki}} e^{-W_k \Lambda_0(t_{ki}) \exp(\boldsymbol{\beta}^T \boldsymbol{X}_{ki})}, \quad (2.29)$$

where λ_0 and Λ_0 denote the baseline and cumulative baseline hazard and f the gamma density with parameters (ν, η). If the frailty were observed the estimation of coefficient vector $\boldsymbol{\beta}$ would be reduced to the estimation of the traditional Cox model, with $\log(W_k)$ entering the model as offset (cf. (2.28)). Since the frailty is not observed we need to integrate out the frailty terms in (2.29), resulting in the observed data likelihood

$$L(\boldsymbol{\beta}, \lambda_0 | \text{data}\,) = \prod_{k=1}^{K} \frac{\eta^\nu \Gamma(\nu + d_k + 1)}{\Gamma(\nu)(\eta + \sum_{i=1}^{n_k} \Lambda_0(t_{ki}) e^{\boldsymbol{\beta}^T \boldsymbol{X}_{ki}})^{\nu + d_k + 1}} \prod_{i=1}^{n_k} \left(\lambda_0(t_{ki}) e^{\boldsymbol{\beta}^T \boldsymbol{X}_{ki}} \right)^{\delta_{ki}}. \quad (2.30)$$

This is a more difficult maximization task that can be solved using the frailty option of `coxph()` from the `survival` package (Therneau, 2015) in `R` (R Core Team, 2015), which uses numerical approximation. Another approach is to use the expectation maximization algorithm, which will be further investigated in Chapter 5.

Chapter 3

Motivating example

This chapter will illustrate available methods, to study time to event data in a multi-center study of competing risks, on an example data set. The data set origins from a multi-center clinical trial on breast cancer by the European Organisation for Research and Treatment of Cancer.

Information on the studied population is given in Section 3.1. The data is described in Section 3.2. A competing risks analysis is performed in Section 3.3 and a multivariate frailty analysis in Section 3.4. All analysis are performed using the software `R` (R Core Team, 2015).

3.1 Background information

Breast cancer is one of the most common types of cancer in women. It can be classified by grade, indicating how well the cells are differentiated, as cells progressively lose the features seen in normal breast cells, where low grade corresponds to better differentiated cells.

Breast cancer can also be classified as carcinoma in situ and invasive carcinoma. The former refers to cancer confined within a particular tissue compartment without invasion of the surrounding tissue. Invasive carcinoma, as suggested by the terminology, does not confine itself to the initial tissue compartment.

To define the stage of the cancer the TNM (Tumor, Node, Metastasis) system is used. This technique is based on the tumor size, whether or not cancer has spread to the lymph nodes in the armpits and whether the tumor has metastasized. Stages rank from 0 to IV, where stage 0 indicates non-invasive breast cancer with no evidence of abnormal cells out of their original confined location and stage IV describes invasive cancers that have spread to other body parts (Breastcancer.org, 2015).

The standard treatment for breast cancer is surgery, which may be followed by chemotherapy, radiotherapy or both. Surgeries can be classified in breast conserving surgery (partial breast removal) and mastectomy (complete breast removal).

Disease progression after surgery can be described in terms of events a patient might experience. A patient can develop local recurrence (LR), which means that the tumor

grows back at the site of surgery and/or might develop distant metastasis (DM), which corresponds to a tumor growth not at the site of surgery and/or she might die.

3.2 Data description

The European Organisation for Research and Treatment of Cancer (EORTC) has conducted several large randomized phase III trials. The goal was to optimize the clinical procedure applied to breast cancer in women with stage I or II disease. The data used in this thesis originates from the EORTC trial 10854, which studied the effect of one course of perioperative chemotherapy given directly after surgery on survival, compared to surgery alone (van der Hage et al., 2001). The data set includes 2795 women treated for invasive stage I or II breast cancer, randomized for treatment in 15 different centers. The beginning of follow-up corresponds to the time of randomization, which is close to the date of surgery. The end of follow-up corresponds to the incidence of distant metastasis or the last date of follow-up (due to death, being lost to follow-up, or end of study).

Patients with preoperative chemotherapy, patients not eligible for the study (due to false inclusion or severe protocol violation) and patients with stage III breast cancer were excluded from analysis ($n = 41$). Also patients without full information on all covariates were excluded from analysis ($n = 91$). Further all 5 patients from center 301 were excluded, because of the small amount of patients at this center, leaving a total of 2658 patients from 14 different centers for analysis.

The data was checked and two patients were found to have developed DM after time of death. Those patients were censored for DM at time of death. Two patients were found to have LR at time of death. Those patients were censored for death at time of LR, since LR occurred before death and for the analysis performed in the following only the first event is considered. Further 81 patients experienced LR and DM at the same time and were censored for DM for the same reason.

The prognostic factors considered in this analysis are age, tumor size, whether or not the patient has positive lymph nodes, type of surgery (breast conserving or mastectomy) and treatment (adjuvant chemo- or radiotherapy or perioperative chemotherapy). Patients' characteristics are provided in Table 3.1 and the distribution of patients over centers is given in Figure 3.1.

3.3 Competing risks analysis

The competing risks model studied in this thesis is illustrated in Figure 3.2. Two competing events are considered, local recurrence (LR) and distant metastasis or death (DM/Death). The starting state is the state a patient enters after surgery, alive with no evidence of disease (ANED). The black numbers on the arrows indicate the number of patients moving from one state to the other and the blue numbers correspond to the transition numbers.

Table 3.1: Characteristics of 2658 patients.

Variable	N	(%)
Age		
≥50	1602	(60.3)
40–50	762	(28.7)
<40	294	(11.1)
Tumor size		
<2cm	798	(30.0)
≥2cm	1860	(70.0)
Nodal status		
negative	1407	(52.9)
positive	1251	(47.1)
Surgery		
mastectomy	1164	(43.8)
breast conserving	1494	(56.2)
Perioperative chemotherapy		
yes	1325	(49.8)
no	1333	(50.2)
Adjuvant chemotherapy		
no	2173	(81.8)
yes	485	(18.2)
Adjuvant radiotherapy		
no	54	(2.0)
yes	2604	(98.0)

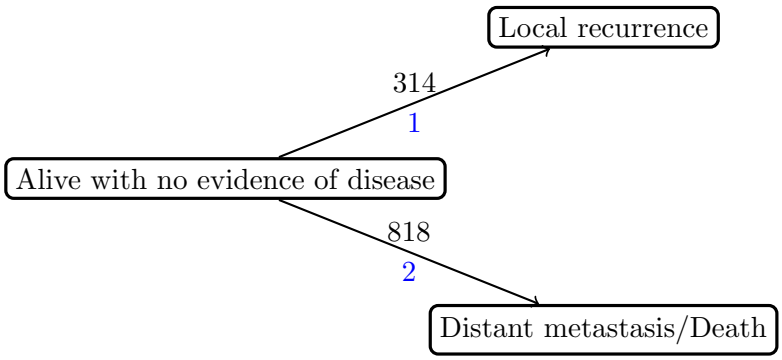

Figure 3.2: Competing risks model, 2658 patients at risk. Numbers denoted in black and blue indicate the number of patients moving from one state to the other and the transition numbers respectively.

Patient Data is collected in a wide format, in which each row represents a single individual. To analyze the data with available software, it is first transformed into a

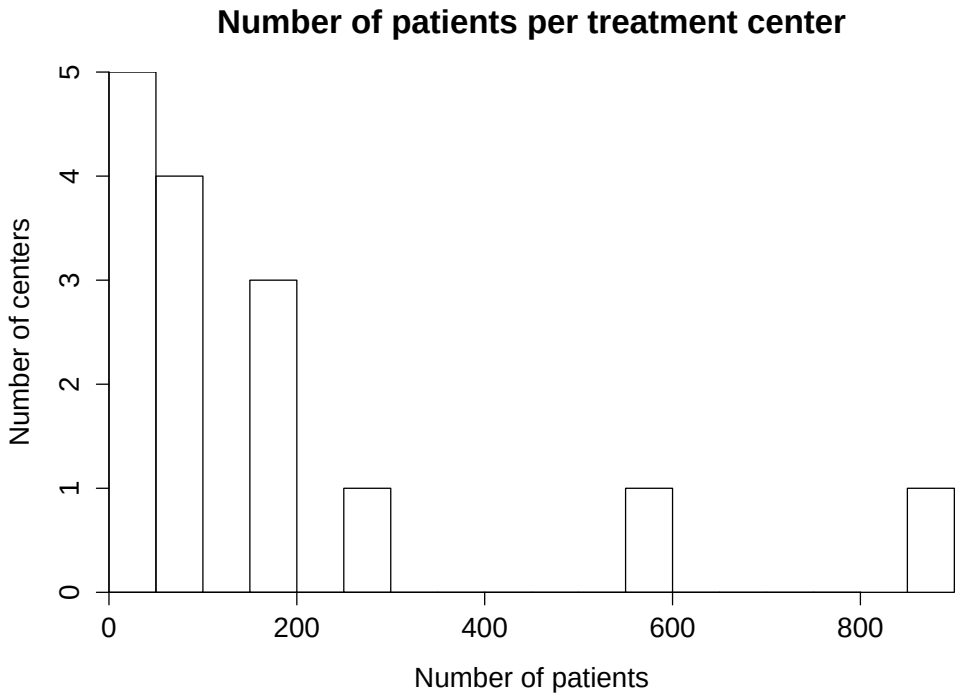

Figure 3.1: Distribution of patients over 14 centers.

specific long format shown for three patients in Output 1. In this format each individual is represented by 2 rows, one for each possible cause of failure. A start and stop variable indicate the time interval a patient is at risk. A status variable indicates whether or not a patient moved from one state to the other. The initial wide format data is transformed in long format using the `msprep()` function from the `R` package `mstate` (de Wreede et al., 2011).

```
An object of class 'msdata'

Data:
  id from to trans Tstart     Tstop status center  age  size
1  1    1  2     1      0  2.006845      0    903 >=50 >=2cm
2  1    1  3     2      0  2.006845      1    903 >=50 >=2cm
3  2    1  2     1      0 13.379877      0    903  <40 >=2cm
4  2    1  3     2      0 13.379877      0    903  <40 >=2cm
5  3    1  2     1      0  1.275838      1    903 >=50 >=2cm
6  3    1  3     2      0  1.275838      0    903 >=50 >=2cm
```

Output 1: Data example in long format.

Separate analysis for the two causes of failure are performed by selecting the data subset corresponding to transition 1 and 2.

Non-parametric cause-specific cumulative hazards for each cause of failure and 95% confidence intervals are shown in Figure 3.3. The hazards are estimated not considering any covariates, as specified in (2.21). The cumulative hazard for LR is lower than for DM/Death, which reflects the number of patients experiencing these events depicted in Figure 3.2.

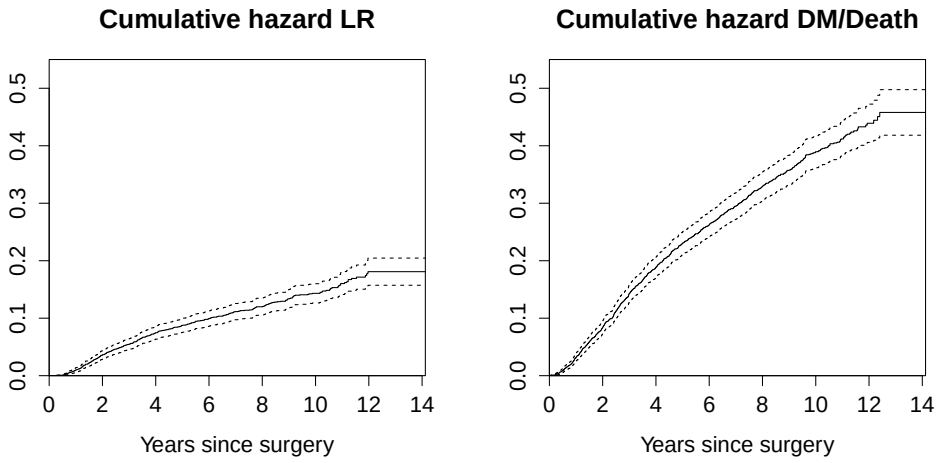

Figure 3.3: Non-parametric cause-specific cumulative hazards for LR and DM/Death.

For the cause-specific regression analysis the choice of covariates is based on a previous study on the same data (de Bock et al., 2009). The following prognostic factors are considered in the analysis: age ($\geq 50, 40-50, <40$), tumor size ($<$ 2cm, $\geq$ 2cm), nodal status (negative, positive), type of surgery (mastectomy, breast conserving), perioper-

ative chemotherapy (yes, no), adjuvant chemotherapy (yes, no), adjuvant radiotherapy (yes, no). Since separate analysis are performed for each cause of failure, we estimate for each covariate a cause-specific hazard ratio (HR). The results of the competing risks analysis are shown in Table 3.2. In the following we consider a p value smaller than 5% to be significant.

For transition 1 (ANED $\rightarrow$ LR) age, nodal status and perioperative chemotherapy have a significant effect on the risk of moving from one state to the other. Both age categories 40–50 and ≥ 50 increase the risk of moving to LR compared to the baseline group with HRs equal to 1.54 (1.16–2.05) and 2.37 (1.7–3.31) respectively. A positive nodal status increases the risk of moving from one state to the other with a HR equal to 1.32 (1–1.73). Perioperative chemotherapy has a protective effect since the cause-specific hazard ratio for patients who did not receive perioperative chemotherapy is equal to 1.46 (1.17–1.83).

For transition 2 (ANED $\rightarrow$ DM/Death) age, tumor size, nodal status and the type of surgery have a significant effect on DM or death. The age group 40–50 has a protective effect on moving to DM/Death with a HR equal to 0.82 (0.67–1) compared to the baseline age group ≥ 50. A larger tumor and a positive nodal status both increase the risk of moving to the state DM/Death with HRs equal to 1.45 (1.22–1.72) and 1.65 (1.41–1.93) respectively. Breast conserving operation has a protective effect with a HR equal to 0.81 (0.7–0.94) compared to mastectomy.

Table 3.2: Competing risks model: effect of covariates on each cause of failure.

	ANED $\rightarrow$ LR			ANED $\rightarrow$ DM/Death		
	HR	0.95 CI	P value	HR	0.95 CI	P value
Age						
$\geq$50	1			1		
40–50	1.54	1.16-2.05	0.003	0.82	0.67-1	0.049
<40	2.37	1.7-3.31	0.000	1.18	0.92-1.51	0.184
Size ($\geq$ 2cm vs <2cm)	1.18	0.92-1.52	0.186	1.45	1.22-1.72	0.000
Node (pos. vs neg.)	1.32	1-1.73	0.048	1.65	1.41-1.93	0.000
Surgery (cons. vs mast.)	1.21	0.95-1.55	0.125	0.81	0.7-0.94	0.005
CTperi (no vs yes)	1.46	1.17-1.83	0.001	1.05	0.92-1.21	0.487
CTadj (yes vs no)	0.73	0.51-1.06	0.098	0.79	0.63-1	0.052
RTadj (yes vs no)	0.66	0.33-1.3	0.227	1.34	0.77-2.34	0.298

Figure 3.4 illustrates stacked cumulative incidence curves of the events LR and DM/Death for two patients with different baseline characteristics, created with the `R` package `mstate` (de Wreede et al., 2011). The width of the colored areas are the patient's probabilities of being in that state at the corresponding time. The prognosis of the patient represented by the left panel seems better than for the patient on the right.

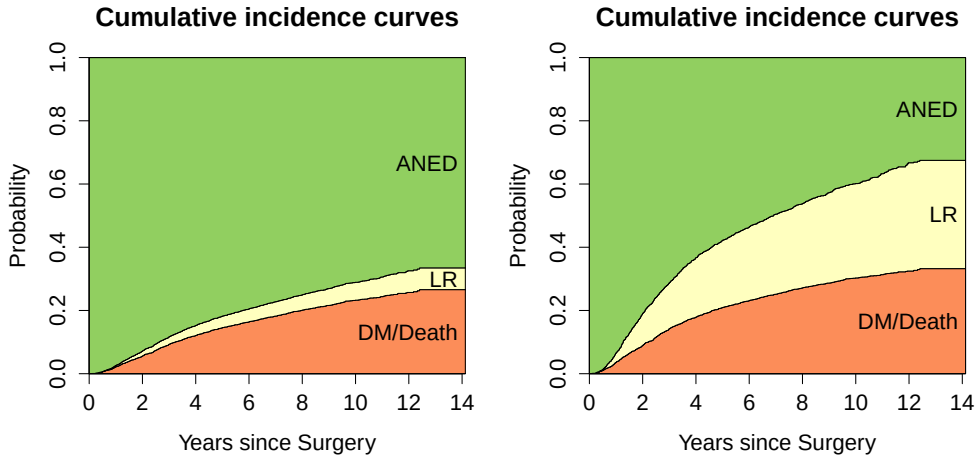

Figure 3.4: Cumulative incidence curves for two patients. Patient's characteristic on the left side: age ≥ 50, small tumor, negative nodal status, mastectomy, perioperative chemotherapy, no adjuvant chemotherapy, radiotherapy. Patient's characteristic on the right side: age < 40, large tumor, positive nodal status, breast conserving therapy, perioperative chemotherapy and no adjuvant chemo- or radiotherapy.

3.4 Frailty model

Heterogeneity between hospitals is indicated by the variation of non-parametric cause-specific cumulative hazards in Figure 3.5, where each line of a different color represents a hospital. The left panel shows the cumulative incidence curves for the cause LR and the right panel for the cause DM/Death.

To account for center effect in a cause-specific regression model each cause of failure within a hospital is assigned its own independent frailty The likelihood functions for competing risks (2.24) and frailty models (2.29) contribute to the likelihood function of a competing risks model with independent frailties

$$L(\beta_1, \beta_2, \lambda_{10}, \lambda_{20} | \text{data}, \boldsymbol{W}) = \prod_{k=1}^{K} f(W_{k1}) f(W_{k2}) \prod_{i=1}^{n_k} e^{-\sum_j W_{kj} \Lambda_{j0}(t_{ki}) \exp(\boldsymbol{\beta}_j^T \boldsymbol{X}_{ki})} \prod_{j=1}^{2} \left(W_{kj} \lambda_{j0}(t_{ki}) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ki}} \right)^{\delta_{ki}}. \tag{3.1}$$

The model can be estimated similarly to the classical competing risks model, by using the `coxph()` function from the `R` package `survival` (Therneau, 2015) and including a frailty term for each transition separately. The results of the estimated model with gamma frailty are shown in Table 3.3.

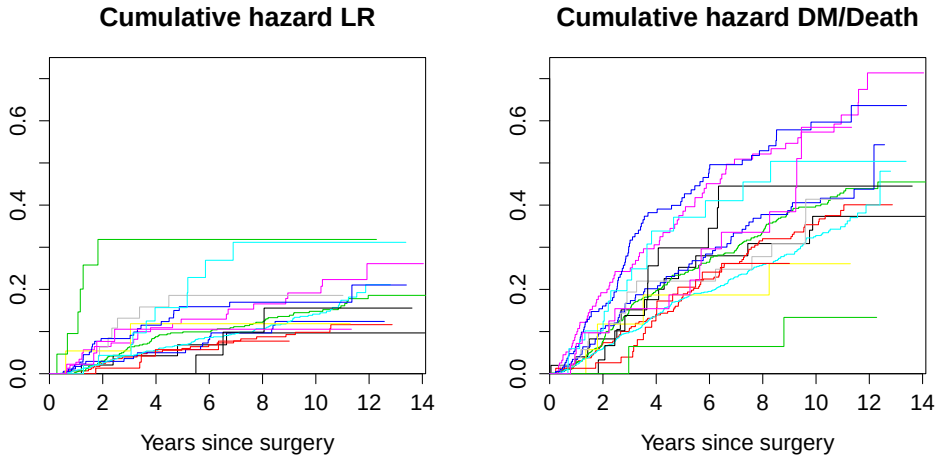

Figure 3.5: Cause-specific cumulative hazards for 14 centers.

For transition 1 (ANED → LR) the hazard ratios are the same as in the traditional competing risks analysis (cf. Table 3.2). This is explained by the estimated frailty variance, which is nearly 0 and not significant. This indicates no center effect on the transition to LR.

For transition 2 (ANED →DM/Death) the hazard ratios are slightly different from the traditional competing risks analysis. However, the set of significant risk factors stay the same. The frailty variance is equal to 0.075 with a significant p value, which indicates a significant center effect on this transition.

A different frailty model assigns to each hospital a shared frailty term for both causes of failure. The likelihood function for this model is similar to (3.1) without the cause index j for the frailty. The model can be estimated using methods from the `R` package `mstate` (de Wreede et al., 2011), which allows to specify covariate effects to be transition specific or not.

Both, the independent and shared frailty model, are not ideal. The former assumes an independent effect of the unobserved covariates on the two events and the latter assumes them to have the same effect on both events. A model allowing for possible correlation between frailties is a more accurate representation of reality.

Table 3.3: Frailty model : effect of covariates on each cause of failure.

	ANED → LR			ANED → DM/Death		
	HR	0.95 CI	P value	HR	0.95 CI	P value
Age						
≥50	1			1		
40–50	1.54	1.16-2.05	0.003	0.76	0.62-0.93	0.007
<40	2.37	1.7-3.31	0.000	1.06	0.82-1.36	0.663
Size (≥ 2cm vs <2cm)	1.18	0.92-1.52	0.186	1.44	1.21-1.71	0.000
Node (pos. vs neg.)	1.32	1-1.73	0.048	1.65	1.41-1.93	0.000
Surgery (cons. vs mast.)	1.21	0.95-1.55	0.125	0.82	0.71-0.96	0.013
CTperi (no vs yes)	1.46	1.17-1.83	0.001	1.05	0.92-1.21	0.483
CTadj (yes vs no)	0.73	0.51-1.06	0.098	0.85	0.66-1.08	0.175
RTadj (yes vs no)	0.66	0.33-1.3	0.227	1.37	0.75-2.48	0.302
	Variance		P value	Variance		P value
Frailty	5e-07		0.901	0.075		0.000

Chapter 4

Competing risks frailty model

The previous chapter described frailty models available in competing risks situations. Independent frailties allow for different center effects on the two events, however a possible correlation of effects remains unnoticed. The unobserved covariates at center level, could have a common effect on both risks.

Heterogeneity between centers in a competing risks setting can be modeled using a correlated frailty model. Such a model may give further insight on the center effect and the relationship of the two events. A model for the dependence structure was first proposed by Yashin et al. (1995) in a twin study, decomposing the frailty of each twin as a sum of two independent frailties one of which is shared by both twins. We follow Petersen et al. (1996) where an additive variance components structure on multiplicative gamma frailty models is proposed and its estimation is outlined.

In this chapter a methodology for a competing risks model with correlated frailties will be introduced. First the correlation structure proposed by Petersen et al. (1996) and the desired frailty properties are presented in Section 4.1. The construction of those frailties will be explained in Section 4.2. Finally the estimation procedure will be discussed in Section 4.3.

4.1 Correlation structure

In this section we describe the approach of Petersen et al. (1996) to construct correlated frailties. A discussion about the desired frailty properties follows.

Let W_{k1}, W_{k2} denote the two frailty variables assigned to each cause of failure within hospital k. Each of those frailties is constructed by adding two independent gamma distributed frailty components, one of which is in common. The frailties' common component models the unobserved covariates that have the same effect on both causes and the independent component allows for unobserved cause-specific effects. The frailties of hospital k are constructed in the following way

$$W_{k1} = Z_{k0} + Z_{k1}, \tag{4.1}$$

$$W_{k2} = Z_{k0} + Z_{k2}, \tag{4.2}$$

where the terms Z_{k0}, Z_{k1}, Z_{k2} are independent gamma distributed random variables with parameters $(\nu_0, \eta), (\nu_1, \eta), (\nu_2, \eta)$ respectively. The common rate parameter η assures that W_{k1} and W_{k2} are again gamma distributed with parameters (ν_a, η_a) and (ν_b, η_b) respectively.

The desired frailty terms have mean equal to one and variances and correlation equal to

$$\text{Var}\,(W_{k1}) = \frac{1}{\nu_a} = \xi_a, \text{Var}\,(W_{k2}) = \frac{1}{\nu_b} = \xi_b, \tag{4.3}$$

$$\text{Corr}\,(W_{k1}, W_{k2}) = \rho. \tag{4.4}$$

The gamma distribution to model dependence between frailties has been criticized for leading to stronger dependence of late events and for being used without biological indication (Hougaard, 1995). However, the mathematical properties justify the large use of this distribution. In the context discussed in this thesis it is convenient that the conditional distribution of the frailties given survival data belongs to the gamma family.

4.2 Frailty decomposition

Correlation of the frailties W_{k1} and W_{k2} can be constructed as illustrated in the previous section. To obtain the frailty properties described in (4.3) and (4.4) the parameters for the frailty components are chosen in a specific way and the sum of frailty components is multiplied by a standardizing constant as follows (Fiocco et al., 2009)

$$W_{k1} = \frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu 1}, \tag{4.5}$$

$$W_{k2} = \frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu 2}, \tag{4.6}$$

where

$$Z_{k0} \sim \Gamma(\nu_0, 1), \qquad Z_{k1} \sim \Gamma(\nu_1, 1), \qquad Z_{k2} \sim \Gamma(\nu_2, 1). \tag{4.7}$$

The random variables Z_{k0}, Z_{k1}, Z_{k2} are independent. This results in the following distributions for the frailties

$$W_{k1} \sim \Gamma(\nu_0 + \nu_1, \nu_0 + \nu_1), \tag{4.8}$$

$$W_{k2} \sim \Gamma(\nu_0 + \nu_2, \nu_0 + \nu_2). \tag{4.9}$$

The expectation of the frailty variables is one, which corresponds to no hospital effect. Their variance and correlation are given by

$$\text{Var}\,(W_{k1}) = \frac{1}{\nu_0+\nu_1} = \xi_a, \text{Var}\,(W_{k2}) = \frac{1}{\nu_0+\nu_2} = \xi_b, \tag{4.10}$$

$$\text{Corr}\,(W_{k1}, W_{k2}) = \nu_0(\xi_a\xi_b)^{1/2} = \rho. \tag{4.11}$$

The construction of this model allows for positive correlation between frailties only, as apparent from (4.11). This is a disadvantage of this approach. However, in many practical situations it may be justified to disregard negative correlation.

4.3 Model estimation

Estimation of the model parameters is obtained by maximization of the likelihood function based on the observed data. As described in Section 3.4 formula (3.1) each center has its own associated frailties acting multiplicative on the baseline hazard. Frailties associated to different centers and individuals across hospitals are independent. Thus the likelihood of the data is the product of hospital likelihoods. For simplicity only the likelihood of a single center is given in the following.

Let Z_{k0}, Z_{k1}, Z_{k2} be the frailty components for hospital k with distribution as specified in (4.7). The general structure of the data has the following elements:

- $k = 1, ..., K$ indicate the hospital
- n_k : number of patients in hospital k
- d_{kj} : number of patients in hospital k failing from cause $(j = 1, 2)$
- $\boldsymbol{X}_{ki}, t_{ki}, (\delta_{ki} = 1, 2)$: the covariate vector, the event or censoring time and the event indicator for patient i in hospital k respectively
- $\boldsymbol{\beta}_j, \Lambda_{j0}$: vector of coefficients and cumulative baseline hazard for the two causes of failure $(j = 1, 2)$
- $\Lambda_{kj} = \sum_{i=1}^{n_k} \Lambda_{j0}(t_{ki}) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ki}}$, $(j = 1, 2)$

The complete data likelihood for hospital k, if the frailties were observed is given as

$$L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20} | \text{data}\,_k, Z_{k0}, Z_{k1}, Z_{k2}) = \prod_{i=1}^{n_k} \left(\frac{Z_{k0}+Z_{k1}}{\nu_0+\nu_1} \lambda_{10}(t_{ki}) e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}} \right)^{1_{\{\delta_{ki}=1\}}} \tag{4.12}$$

$$\left(\frac{Z_{k0}+Z_{k2}}{\nu_0+\nu_2} \lambda_{20}(t_{ki}) e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}} \right)^{1_{\{\delta_{ki}=2\}}}$$

$$\exp\left(- \left(\frac{Z_{k0}+Z_{k1}}{\nu_0+\nu_1} \Lambda_{10}(t_{ki}) e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}} + \frac{Z_{k0}+Z_{k2}}{\nu_0+\nu_2} \Lambda_{20}(t_{ki}) e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}} \right) \right).$$

Each patient contributes to the likelihood function by the product of his cause-specific hazard rate and his survival probability until that point. The first and second factors

in the likelihood represent the probability of failing from cause 1 and 2 respectively. The last factor expresses the probability of remaining event-free until censoring time or failure and depends on all cause-specific hazards.

Integrating out all frailty components specific to each center we obtain terms that depend on the number of failures d_{kj} of cause j in center k. The observed data likelihood for center k is as follows

$$
\begin{aligned}
L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20} | \text{data}_k) = (\nu_0+\nu_1)^{-d_{k1}} (\nu_0+\nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0)\Gamma(\nu_1)\Gamma(\nu_2)} \qquad (4.13)\\
\prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}})^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}})^{1_{\{\delta_{ki}=2\}}} \right] \\
\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} \\
\frac{\Gamma(d_{k1}+d_{k2}+\nu_0-l-m)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{d_{k1}+d_{k2}+\nu_0-l-m}},
\end{aligned}
$$

where Γ is the gamma function.

Since the frailties are not observed the estimation procedure is more complicated than the traditional one. In combination with the unspecified baseline hazard, it is computationally challenging to maximize the likelihood function (4.13). The unobserved data can be seen as missing information and the expectation maximization algorithm (EM-algorithm) can be used to estimate the model parameters (Petersen et al., 1996). The complete data problem is much easier to solve compared to the observed data problem, this is exploited by EM-algorithm. Details about its usage to estimate the model are given in Chapter 5.

Chapter 5

Expectation maximization algorithm

The expectation maximization algorithm (EM-algorithm) is applied to estimate parameters by maximizing the likelihood function in case of incomplete data. It is used when the incomplete data problem is much more difficult to solve than the complete data problem.

The EM-algorithm reduces the problem of optimizing the observed data log-likelihood into sequences of simpler optimization problems, whose maxima are easy to compute. These subproblems are chosen in a way that guarantees their corresponding solution to converge to a local optimum of the observed data log-likelihood.

Its iterative scheme starts from some initial parameters, and determines which values are most likely for the missing data, using the current parameter estimates. Assuming these data completions to be correct, the next set of parameters is calculated through maximization of the complete data log-likelihood. These two steps are repeated until convergence.

The formulation of the algorithm will be illustrated in Section 5.1. The estimation procedure for the competing risks frailty model is described in Section 5.2.

5.1 Algorithm formulation

The following notation and methodology follow Louis (1982). A simple small data example can be found in Do and Batzoglou (2008).

Let x denote the observed data, w the unobserved data and θ the parameter to be estimated. The complete data is given by x, w and the incomplete data by x only. The log-likelihood for the complete and incomplete case are respectively given by

$$\ell(\theta|x,w) = \log f(x,w|\theta), \tag{5.1}$$

$$\ell^*(\theta|x) = \log f^*(x|\theta) = \log \int_R f(x,w|\theta)dw, \tag{5.2}$$

where $R = \{w : x(w) = x\}$ is the set of possible completions for the missing data and f and f^* the probabilities of the complete and incomplete data given θ respectively. Instead of maximizing ℓ^* directly, the EM-algorithm proceeds by taking a starting parameter estimate $\theta^{(0)}$ and solving the pseudo-complete data problem

$$Q(\theta|\theta^{(0)}) = E_{\theta^{(0)}}[\ell(\theta|x, w)|w \in R]. \tag{5.3}$$

The value maximizing (5.3) gives $\theta^{(1)}$, which is used in the next iteration. Summarizing the algorithm:

E-step: Compute

$$Q(\theta|\theta^{(k)}) = E_{\theta^{(k)}}[\ell(\theta|x, w)|w \in R].$$

M-step: Maximize

$$\text{maximize}_{\theta\in\Theta} Q(\theta|\theta^{(k)}).$$

The iteration over these steps is continued until $||\ell^*(\theta^{(v+1)}|x) - \ell^*(\theta^{(v)}|x)||$ is sufficiently small.

Since the algorithm is guaranteed only to converge to a local maximum of the observed data log-likelihood, running the procedure using multiple starting values can be of use (Do and Batzoglou, 2008).

5.2 Implementation

The estimation procedure uses the EM-algorithm to approximate the observed data log-likelihood to find optimal regression coefficients and baseline hazards for fixed $\boldsymbol{\nu} = (\nu_0, \nu_1, \nu_2)$. The in this way approximated observed data log-likelihood is then employed in a three dimensional search to find maximum likelihood estimates for $\boldsymbol{\nu}$.

For the EM procedure the frailties are considered missing data and the complete and observed data likelihood functions are given in (4.12) and (4.13) respectively. The complete data log-likelihood is given as

$$\begin{aligned}
\log L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda_{10}, \lambda_{20}|\text{data}_k, Z_{k0}, Z_{k1}, Z_{k2}) &= \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=1} \log\left(\frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu_1}\right) \\
&+ \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=1} \log(\lambda_{10}(t_k i) e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}}) + \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=2} \log\left(\frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu_2}\right) \\
&+ \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=2} \log(\lambda_{20}(t_k i) e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}}) - \frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu_1} \sum_{i=1}^{n_k} \Lambda_{10}(t_{ki}) e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}} \\
&- \frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu_2} \sum_{i=1}^{n_k} \Lambda_{20}(t_{ki}) e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}}.
\end{aligned} \tag{5.4}$$

For the E-step it is not always necessary to form the probability distribution over completions given in (5.3) explicitly, but rather need only to compute expected sufficient statistics over these completions (Do and Batzoglou, 2008). This can be exploited in our model.

Since $\boldsymbol{\nu}$ is fixed throughout the EM iterations, the estimation concerns the regression coefficients and baseline hazards only. The conditional expectations of the terms $\log\left((Z_{k0}+Z_{kj})/(\nu_0+\nu_j)\right)$, $(j=1,2)$ given observed data are irrelevant to the estimation of the complete data case (5.4), for fixed $\boldsymbol{\nu}$. For this reason the E-step reduces to the calculation of the conditional expectations of the frailties $W_{kj}=(Z_{k0}+Z_{kj})/(\nu_0+\nu_j)$, $(j=1,2)$ given observed data. These are rather easy to calculate, since the conditional distributions of the frailty components Z_{k0}, Z_{k1}, Z_{k2} given observed data are a mixture of gamma distributions.

The expectations of each frailty component conditional on the data for hospital k are given as

$$\begin{aligned}
\mathrm{E}\left(Z_{k0}|\text{data}_{\,k}\right) &= \int_{z_{k0}} z_{k0} f(z_{k0}|\text{data}_{\,k})dz_{k0} \qquad (5.5)\\
&= \sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}} c(l,m,\nu_0,\nu_1,\nu_2)\frac{d_{k1}+d_{k2}+\nu_0-l-m}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)},\\
\mathrm{E}\left(Z_{k1}|\text{data}_{\,k}\right) &= \int_{z_{k1}} z_{k1} f(z_{k1}|\text{data}_{\,k})dz_{k1} \qquad (5.6)\\
&= \sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}} c(l,m,\nu_0,\nu_1,\nu_2)\frac{l+\nu_1}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)},\\
\mathrm{E}\left(Z_{k2}|\text{data}_{\,k}\right) &= \int_{z_{k2}} z_{k2} f(z_{k2}|\text{data}_{\,k})dz_{k2} \qquad (5.7)\\
&= \sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}} c(l,m,\nu_0,\nu_1,\nu_2)\frac{m+\nu_2}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)},
\end{aligned}$$

where

- f: probability of a frailty component given data
- d_{kj} : number of failures in center k by cause $(j=1,2)$
- $\Lambda_{kj}=\sum_{i=1}^{n_k}\Lambda_{j0}(t_{ki})e^{\boldsymbol{\beta}_j^T\boldsymbol{X}_{ki}}$, $(j=1,2)$
-

$$c(l,m,\nu_0,\nu_1,\nu_2)=\frac{\binom{d_{k1}}{l}\binom{d_{k2}}{m}\frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}}\frac{\Gamma(d_{k1}+d_{k2}+v_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+v_0-l-m}}\frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}}}{\sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}}\binom{d_{k1}}{l}\binom{d_{k2}}{m}\frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}}\frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}}\frac{\Gamma(d_{k1}+d_{k2}+\nu_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+\nu_0-l-m}}}$$

Notably the factor $c(l, m, \nu_0, \nu_1, \nu_2)$ is the same in all three expectations (5.5)–(5.7).

The M-step consists of estimating the updated baseline hazards $\Lambda_{10}(t), \Lambda_{20}(t)$ and coefficient vectors $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, which maximize the complete data log-likelihood. This can be done with existing software, e. g. using `coxph()` from the `R` package `survival`, incorporating the logarithm of the expected frailties as offset into the cause-specific hazards model, as in (2.28). The iterations stop once the change in observed data log-likelihood is smaller than 0.000001.

Up until now the frailty parameters $\boldsymbol{\nu} = (\nu_0, \nu_1, \nu_2)$ were fixed throughout the EM iterations. The function `optim()` is used to find the optimal $\boldsymbol{\nu}$, maximizing the observed data log-likelihood approximated with the EM-algorithm. The parameter space is searched on the log-scale to avoid negative or zero values for the parameters, which would give problems in the calculation.

To start the EM-algorithm five different sets of initial parameter estimates are chosen in the following way. Frailty components for each center Z_{k0}, Z_{k1}, Z_{k2} are randomly chosen from a log-normal distribution with means ν_0, ν_1, ν_2 respectively. The variance of the underlying normal distribution is set to 1. The initial regression parameters are calculated, using the `coxph()` function from the `R` package `survival` (Therneau, 2015) with the logarithm of the frailties as an offset. The log-normal distribution was chosen not to allow for negative values.

During the E-step further quantities are calculated. For estimation of the standard error of model estimates in Chapter 6 and computation of conditional variances of the frailty components for Chapter 7, the following conditional expectations are calculated similarly to (5.5)–(5.7):

$$\begin{aligned}
\mathrm{E}\left(Z_{k0}^2|\mathrm{data}_k\right) &= \int_{z_{k0}} z_{k0}^2 f(z_{k0}|\mathrm{data}_k) dz_{k0} \qquad (5.8)\\
&= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c(l, m, \nu_0, \nu_1, \nu_2) \frac{d_{k1} + d_{k2} + \nu_0 - l - m}{\left(1 + \frac{1}{\nu_0+\nu_1}\Lambda_{k1} + \frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)} \frac{d_{k1} + d_{k2} + \nu_0 - l - m + 1}{\left(1 + \frac{1}{\nu_0+\nu_1}\Lambda_{k1} + \frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)}\\
\mathrm{E}\left(Z_{k1}^2|\mathrm{data}_k\right) &= \int_{z_{k1}} z_{k1}^2 f(z_{k1}|\mathrm{data}_k) dz_{k1} \qquad (5.9)\\
&= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c(l, m, \nu_0, \nu_1, \nu_2) \frac{l + \nu_1}{\left(1 + \frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)} \frac{l + \nu_1 + 1}{\left(1 + \frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)}\\
\mathrm{E}\left(Z_{k2}^2|\mathrm{data}_k\right) &= \int_{z_{k2}} z_{k2}^2 f(z_{k2}|\mathrm{data}_k) dz_{k2} \qquad (5.10)\\
&= \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} c(l, m, \nu_0, \nu_1, \nu_2) \frac{m + \nu_2}{\left(1 + \frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)} \frac{m + \nu_2 + 1}{\left(1 + \frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left(Z_{k0}Z_{k1}|\mathrm{data}_k\right) &= \int_{z_{k1}}\int_{z_{k0}} z_{k0}z_{k1}f(z_{k0},z_{k1}|\mathrm{data}_k)dz_{k0}dz_{k1} \qquad (5.11)\\
&= \sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}} c(l,m,\nu_0,\nu_1,\nu_2)\frac{d_{k1}+d_{k2}+\nu_0-l-m}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)}\frac{l+\nu_1}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)}\\
\mathrm{E}\left(Z_{k0}Z_{k2}|\mathrm{data}_k\right) &= \int_{z_{k2}}\int_{z_{k0}} z_{k0}z_{k2}f(z_{k0},z_{k2}|\mathrm{data}_k)dz_{k0}dz_{k2} \qquad (5.12)\\
&= \sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}} c(l,m,\nu_0,\nu_1,\nu_2)\frac{d_{k1}+d_{k2}+\nu_0-l-m}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)}\frac{m+\nu_2}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)}
\end{aligned}$$

The conditional variances and covariances of the frailty components for cause ($j = 1, 2$) are given as

$$\begin{aligned}
\mathrm{Var}\left(Z_{k0}|\mathrm{data}_k\right) =& \mathrm{E}\left(Z_{k0}^2|\mathrm{data}_k\right) - \mathrm{E}\left(Z_{k0}|\mathrm{data}_k\right)^2, \qquad (5.13)\\
\mathrm{Var}\left(Z_{kj}|\mathrm{data}_k\right) =& \mathrm{E}\left(Z_{kj}^2|\mathrm{data}_k\right) - \mathrm{E}\left(Z_{kj}|\mathrm{data}_k\right)^2, \qquad (5.14)\\
\mathrm{Cov}\left(Z_{k0},Z_{kj}|\mathrm{data}_k\right) =& \mathrm{E}\left(Z_{k0}Z_{kj}|\mathrm{data}_k\right) - \mathrm{E}\left(Z_{k0}|\mathrm{data}_k\right)\mathrm{E}\left(Z_{kj}|\mathrm{data}_k\right). \qquad (5.15)
\end{aligned}$$

Chapter 6

Estimation of the standard error

The previous chapter illustrated how to estimate parameters for the proposed competing risks frailty model. For the interpretation of the results a measure of accuracy is needed.

The standard error measures the variation of population estimates over different samples and gives an idea about the accuracy of the estimates, in our case the regression and frailty parameters. Apart from being a measure by itself it also allows us to construct confidence intervals for the parameters.

In this chapter we illustrate a method to calculate the covariance matrix for the estimated parameters. We first discuss the structure of the covariance matrix in Section 6.1, followed by a detailed explanation of the components in Section 6.2, and the derivation of the standard error from the covariance matrix in Section 6.3.

6.1 Covariance matrix

To obtain the covariance matrix for the regression parameters, two approaches can be used. One is based on the observed data log-likelihood (Korsgaard and Andersen, 1998). The second approach stays within the EM-algorithm framework, using only derivatives of the complete information log-likelihood (Louis, 1982). The latter approach does not yet include the uncertainty that is caused by estimating the frailty parameters $\boldsymbol{\nu} = (\nu_0, \nu_1, \nu_2)$. Putter and Houwelingen (2015, supplementary material) propose the following way of estimation.

Let $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu}) = (\hat{\boldsymbol{\beta}}_1^T(\boldsymbol{\nu}), \hat{\boldsymbol{\beta}}_2^T(\boldsymbol{\nu}), \hat{\boldsymbol{\lambda}}_{10}^T(\boldsymbol{\nu}), \hat{\boldsymbol{\lambda}}_{20}^T(\boldsymbol{\nu}))^T$ denote the maximum likelihood estimates (MLE) of the regression coefficients and baseline hazards given frailty parameters $\boldsymbol{\nu}$, and $\hat{\boldsymbol{\nu}}$ denote the MLE of $\boldsymbol{\nu}$ maximizing the observed data log-likelihood. The combined covariance matrix of $\log(\hat{\boldsymbol{\nu}}), \hat{\boldsymbol{\eta}}$ is given as

$$\begin{pmatrix} \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} & \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right)^T \\ \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right) \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} & \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}} + \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right) \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\right)^T \end{pmatrix}, \tag{6.1}$$

where $\Sigma_{\boldsymbol{\nu\nu}}$ is the covariance matrix of $\log(\boldsymbol{\nu})$, $\Sigma_{\boldsymbol{\eta\eta}}$ is the covariance matrix of $\hat{\boldsymbol{\eta}}$ and $\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ are the partial derivatives of the regression parameters given $\boldsymbol{\nu}$. The term on the bottom right of (6.1) represents the covariance of $\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})$, where $\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})$ is obtained using a Taylor series of $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu})$ and of the score functions of $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu})$ and $\hat{\boldsymbol{\nu}}$ around the MLEs. The off diagonal terms are the covariance matrices of $(\log(\hat{\boldsymbol{\nu}}), \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}}))$ and can be derived in a similarly way.

6.2 Covariance components

We compute the term $\Sigma_{\boldsymbol{\nu\nu}}$ from the Hessian matrix returned by the `optim()` function in `R`, which is used to find the optimal $\log(\boldsymbol{\nu})$. We proceed by inverting the negative of the Hessian matrix, since the inverse of the observed profile information equals the $\boldsymbol{\nu}$ component of the full observed inverse information evaluated at $(\boldsymbol{\nu}, \hat{\boldsymbol{\eta}}(\boldsymbol{\nu}))$ (Young and Smith, 2005, sec. 8.6.2).

The term $\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ is approximated numerically. Since the regression parameters are estimated by approximating the observed data log-likelihood through the EM-algorithm, no closed form function is available to compute the derivative. The derivative around the MLE is estimated by calculating the slope between parameters for values of $\boldsymbol{\nu}$ close to the MLE. Let $\epsilon > 0$ be a small constant, then an estimate of $\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ around $\hat{\boldsymbol{\nu}}$ is computed as follows

$$\frac{\partial \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})}{\partial \boldsymbol{\nu_0}} = \frac{\hat{\boldsymbol{\eta}}((\hat{\nu_0} + \epsilon/2, \hat{\nu_1}, \hat{\nu_2})) - \hat{\boldsymbol{\eta}}((\hat{\nu_0} - \epsilon/2, \hat{\nu_1}, \hat{\nu_2}))}{\epsilon}, \tag{6.2}$$

$$\frac{\partial \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})}{\partial \boldsymbol{\nu_1}} = \frac{\hat{\boldsymbol{\eta}}((\hat{\nu_0}, \hat{\nu_1} + \epsilon/2, \hat{\nu_2})) - \hat{\boldsymbol{\eta}}((\hat{\nu_0}, \hat{\nu_1} - \epsilon/2, \hat{\nu_2}))}{\epsilon}, \tag{6.3}$$

$$\frac{\partial \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})}{\partial \boldsymbol{\nu_2}} = \frac{\hat{\boldsymbol{\eta}}((\hat{\nu_0}, \hat{\nu_1}, \hat{\nu_2} + \epsilon/2)) - \hat{\boldsymbol{\eta}}((\hat{\nu_0}, \hat{\nu_1}, \hat{\nu_2} - \epsilon/2))}{\epsilon}. \tag{6.4}$$

The term $\Sigma_{\boldsymbol{\eta\eta}}$ can be computed as described in Louis (1982). As discussed before this approach works within the EM framework. It requires the gradient vector and second derivative matrix of the complete data log-likelihood, but not the ones associated to the incomplete data case.

The notation in this chapter follow that introduced in Chapter 5. Let x, w be the complete data, x the incomplete data, ℓ and ℓ^* the corresponding log-likelihood functions. Let $R = \{w : x(w) = x\}$ denote the possible completions given the data, and θ the parameters. Louis (1982) show that the gradient of the observed data log-likelihood can be expressed by the gradient of the complete data log-likelihood:

$$\frac{\partial}{\partial \theta} \ell^*(\theta|x) = \mathrm{E}_{\,\theta} \left(\frac{\partial}{\partial \theta} \ell(\theta|x, w) | w \in R \right). \tag{6.5}$$

The second derivative can be expressed as

$$\begin{aligned}\frac{\partial^2}{\partial\theta\partial\theta}\ell^*(\theta|x) =& \mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta\partial\theta}\ell(\theta|x,w)|w\in R\right)\\ &+\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\ell(\theta|x,w)\frac{\partial}{\partial\theta}\ell^T(\theta|x,w)|w\in R\right)\\ &-\frac{\partial}{\partial\theta}\ell^*(\theta|x)\frac{\partial}{\partial\theta}\ell^{*T}(\theta|x),\end{aligned} \tag{6.6}$$

where at the MLE $\hat{\theta}$ the last term is zero, which is the only one depending on the observed data gradient.

We can write the Fisher information for $\hat{\boldsymbol{\eta}}$ of the incomplete data in terms of the complete data quantities as follows

$$\begin{aligned}I_{\boldsymbol{\eta\eta}}(\boldsymbol{\nu}) =& \mathrm{E}_{\boldsymbol{\nu}}\left(-\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}}\ell^*(\boldsymbol{\eta}|\text{data})\right)\\ =&\mathrm{E}_{\boldsymbol{\nu}}\left(-\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}}\ell(\boldsymbol{\eta}|\text{data},\boldsymbol{W})|\boldsymbol{W}\in R\right)\\ &-\mathrm{E}_{\boldsymbol{\nu}}\left(\frac{\partial}{\partial\boldsymbol{\eta}}\ell(\boldsymbol{\eta}|\text{data},\boldsymbol{W})\frac{\partial}{\partial\boldsymbol{\eta}}\ell^T(\boldsymbol{\eta}|\text{data},\boldsymbol{W})|\boldsymbol{W}\in R\right)\\ &+\frac{\partial}{\partial\boldsymbol{\eta}}\ell^*(\boldsymbol{\eta}|\text{data})\frac{\partial}{\partial\boldsymbol{\eta}}\ell^{*T}(\boldsymbol{\eta}|\text{data}),\end{aligned} \tag{6.7}$$

where $\boldsymbol{W}$ are the unobserved frailties. The last term of (6.7) is zero at the MLE $\hat{\boldsymbol{\nu}}$. A simplified notation for the Fisher information at the MLE is

$$I_{\boldsymbol{\eta\eta}}^{(full)} - I_{\boldsymbol{\eta\eta}}^{(loss)}, \tag{6.8}$$

where the first term represents the full information and the second term represents the loss of information due to the missing data. The estimate of the Fisher information is only of interest at the MLE $\hat{\boldsymbol{\nu}}$. Its quantities need to be computed only for the last iteration of the EM-algorithm, where the observed data gradient is zero. The covariance matrix $\Sigma_{\boldsymbol{\eta\eta}}$ is derived by inversion $\Sigma_{\boldsymbol{\eta\eta}} = I_{\boldsymbol{\eta\eta}}^{-1}$.

To compute the second term in (6.7) corresponding to the loss of information, conditional expectations of the product of frailty components are needed. Given data, frailty components between centers are independent, however not anymore within center. The conditional expectations for the product of frailty components of the same center are computed as specified in (5.8)–(5.12).

6.3 Standard error

The standard error of the estimated regression parameters $\boldsymbol{\eta}$ can be calculated by taking the square root of the corresponding diagonal elements of the covariance matrix (6.1).

To obtain the standard error of the frailty variances and correlation we apply the multivariate *delta method* on $\Sigma_{\boldsymbol{\nu\nu}}$. The delta method is a technique that uses a Taylor series approximation of the transformation function g to approximate its moments, e. g. the expectation or variance (Casella and Berger, 2001, sec. 5.6).

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ be the maximum likelihood estimates of the frailty parameters on the log-scale, then

$$\theta \sim N(\theta, \Sigma_{\boldsymbol{\nu\nu}}). \tag{6.9}$$

For the three transformation functions, corresponding to the frailty variances and correlation follows

$$g_1(\boldsymbol{\theta}) = \frac{1}{e^{\theta_1} + e^{\theta_2}} \sim N\left(\boldsymbol{\theta}, \frac{\partial}{\partial \boldsymbol{\theta}} g_1(\boldsymbol{\theta}) \Sigma_{\boldsymbol{\nu\nu}} \frac{\partial}{\partial \boldsymbol{\theta}} g_1^T(\boldsymbol{\theta})\right), \tag{6.10}$$

$$g_2(\boldsymbol{\theta}) = \frac{1}{e^{\theta_1} + e^{\theta_3}} \sim N\left(\boldsymbol{\theta}, \frac{\partial}{\partial \boldsymbol{\theta}} g_2(\boldsymbol{\theta}) \Sigma_{\boldsymbol{\nu\nu}} \frac{\partial}{\partial \boldsymbol{\theta}} g_2^T(\boldsymbol{\theta})\right), \tag{6.11}$$

$$g_3(\boldsymbol{\theta}) = e^{\theta_1} \sqrt{\frac{1}{e^{\theta_1} + e^{\theta_2}} \frac{1}{e^{\theta_1} + e^{\theta_3}}} \sim N\left(\boldsymbol{\theta}, \frac{\partial}{\partial \boldsymbol{\theta}} g_3(\boldsymbol{\theta}) \Sigma_{\boldsymbol{\nu\nu}} \frac{\partial}{\partial \boldsymbol{\theta}} g_3^T(\boldsymbol{\theta})\right). \tag{6.12}$$

Chapter 7

Results

In this chapter we will discuss the results of the model estimation. In Section 7.1 the estimated regression parameters and frailty variances are discussed. The empirical Bayes estimates of the frailties for each hospital and their corresponding confidence intervals are investigated in Section 7.2.

7.1 Competing risks frailty model

In Table 7.1 the results for the competing risks frailty model with correlated frailties are illustrated. The table shows the estimated regression coefficients, the hazard ratios, the standard errors and corresponding 95% confidence intervals. The lower part of the table shows the estimated frailty variances and correlation, as well as their standard errors calculated as described in Section 6.3.

For transition 1 (ANED $\rightarrow$ LR) only age has a significant effect on moving to LR. The age group < 40 has a significantly increased risk of moving to the state LR compared to the baseline group, with a HR equal to 2.22 (1.036–4.776).

For transition 2 (ANED $\rightarrow$ DM/Death) nodal status has a significant and tumor size a marginally significant effect on moving to DM/Death. A positive nodal status significantly increased the risk of moving to the state DM/Death, with a HR equal to 1.65 (1.14–2.4). A larger tumor size has a marginally significant effect, increasing the risk of moving to DM/Death, with a HR equal to 1.45 (0.998–2.102).

The HRs in this model are comparable to the HRs of the traditional competing risks model in Table 3.2. However, the variation added by additionally estimating the frailties, increased the standard errors and only two variables remain significant. Notably for both transitions the confidence intervals for the effect of adjuvant radiotherapy are extremely large, equal to (0.038–11.657) for transition 1 and (0.067–28.234) for transition 2. The reason may be that only few patients did not receive adjuvant radiotherapy (2%, cf. Table 3.1).

The variance of the frailty for transition 1 (ANED $\rightarrow$ LR) is equal to 0.047 with a standard error of 0.024. For transition 2 (ANED $\rightarrow$ DM/Death) the frailty variance is equal to 0.055 with a standard error of 0.031. The correlation of the frailties is estimated

to be equal to 0.92 with a standard error of 0.084. This strong correlation indicates that the unobserved covariates have a common effect on the two events, which implies that the two risks are closely related. Unobserved covariates increasing the risk of one event likely also increase the risk of the other event and covariates decreasing the risk of one event decrease the risk of the other.

7.2 Empirical Bayes estimates

During the EM iterations we repeatedly estimate values for the unobserved frailties in the E-step. The final completions are the *empirical Bayes estimates* for the frailties of each hospital, which can be used to compare the hospital effects on disease progression.

Figures 7.1 and 7.2 show the empirical Bayes estimates of the frailties of each center together with 95% confidence intervals, for event LR and DM/Death respectively. The x-axis represents the hospitals sorted by the number of patients and the y-axis the estimated frailties on the log-scale. A value equal 1 implies that there is no center effect.

The confidence intervals are computed from the conditional variances of the frailties, which depend on the conditional variances and covariances of the frailty components given in (5.13)–(5.15):

$$\begin{aligned}\text{Var}\left(\frac{Z_{k0}+Z_{k1}}{\nu_0+\nu_1}|\text{data}_k\right) =& \frac{1}{(\nu_0+\nu_1)^2}\left(\text{Var}\left(Z_{k0}|\text{data}_k\right)+\text{Var}\left(Z_{k1}|\text{data}_k\right)\right. \\ &\left.+2\text{Cov}\left(Z_{k0},Z_{k1}|\text{data}_k\right)\right)\end{aligned} \tag{7.1}$$

$$\begin{aligned}\text{Var}\left(\frac{Z_{k0}+Z_{k2}}{\nu_0+\nu_2}|\text{data}_k\right) =& \frac{1}{(\nu_0+\nu_2)^2}\left(\text{Var}\left(Z_{k0}|\text{data}_k\right)+\text{Var}\left(Z_{k2}|\text{data}_k\right)\right. \\ &\left.+2\text{Cov}\left(Z_{k0},Z_{k2}|\text{data}_k\right)\right)\end{aligned} \tag{7.2}$$

Figure 7.1 shows the frailties for the event LR. One can see that two hospitals (9 and 11) have a significantly increased risk for their patients to develop LR. One hospital (12) has a significantly decreased risk for its patients to develop DM or die. Further we see that the width of the confidence intervals decreases with a growing number of patients in the hospital.

Figure 7.2 shows that two hospitals (9 and 11) have a significantly increased risk for their patients to move to the state of DM/Death and one hospital (12) has a significantly decreased risk for its patients to develop DM or die. One hospital (8) has a marginally decreased risk for its patients to move to DM/Death. The width of the confidence intervals again decreases as the number of patients per hospital increases.

To visualize the relation of the two frailties within a hospital we plot the empirical Bayes estimates of the two frailties for each center against each other in Figure 7.3. The x-axis represents the frailty of cause LR and the y-axis the frailty of cause DM/Death, both on the log-scale. Each point in the plot represents one hospital. One can recognize the strong correlation between the frailties as seen before in Table 7.1.

The hospital effects on a patient can be investigated by looking at the difference in cumulative hazard and cumulative incidence between the hospitals. This is shown in

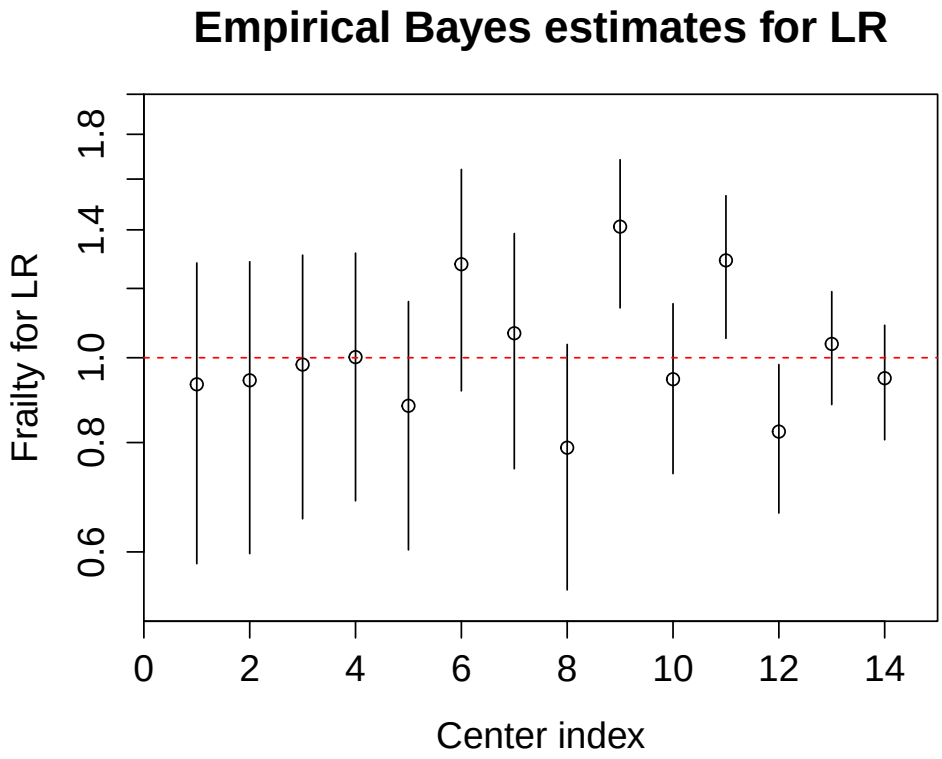

Figure 7.1: Frailties and 95% confidence intervals for event LR of 14 centers, sorted by number of patients.

Figures 7.4 and 7.5, for an average patient.

A pairwise comparison of cumulative incidence curves for an average patient treated in two hospitals further illustrates the difference in effects. This is depicted in Figure 7.6, which shows the stacked cumulative incidence curves for an average patient treated in the two most extreme hospitals. The hospitals are located at the top right and left down of Figure 7.3. The prognosis shown in the left panel estimates a lower risk for both events, compared to the right panel. This is explained by the high estimated correlation between frailties (cf. Table 7.1) and the empirical Bayes estimates of the hospitals (cf. Figure 7.3), which indicate that a hospital with a decreased risks for one cause also has a decreased risk for the other cause. This makes the hospital corresponding to the left panel more attractive.

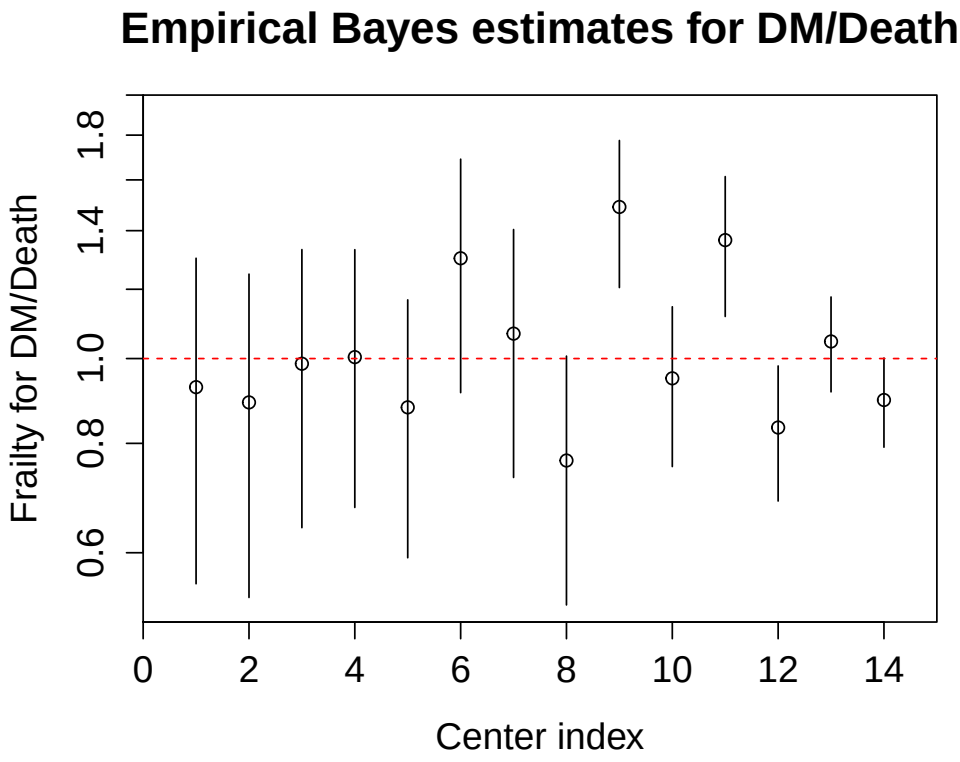

Figure 7.2: Frailties and 95% confidence intervals for event DM/Death of 14 centers, sorted by number of patients.

Table 7.1: Competing risks frailty model, with correlated frailties.

	ANED → LR				ANED → DM/Death			
	coef	se(coef)	HR	0.95 CI	coef	se(coef)	HR	0.95 CI
Age								
≥50			1				1	
40–50	0.38	0.33	1.47	0.769-2.802	-0.26	0.24	0.77	0.482-1.234
<40	0.80	0.39	2.22	1.036-4.776	0.08	0.29	1.08	0.614-1.912
Size (≥ 2cm vs <2cm)	0.18	0.27	1.19	0.703-2.025	0.37	0.19	1.45	0.998-2.102
Node (pos. vs neg.)	0.26	0.32	1.29	0.692-2.425	0.50	0.19	1.65	1.14-2.4
Surgery (cons. vs mast.)	0.19	0.33	1.20	0.631-2.3	-0.20	0.24	0.81	0.509-1.304
CTperi (no vs yes)	0.38	0.24	1.47	0.917-2.349	0.05	0.15	1.05	0.784-1.411
CTadj (yes vs no)	-0.26	0.46	0.77	0.314-1.908	-0.19	0.35	0.83	0.418-1.65
RTadj (yes vs no)	-0.41	1.46	0.67	0.038-11.657	0.32	1.54	1.38	0.067-28.234
	Variance		SE		Variance		SE	
Frailty	0.047		0.024		0.055		0.031	
	Correlation				SE			
Correlation	0.92				0.084			

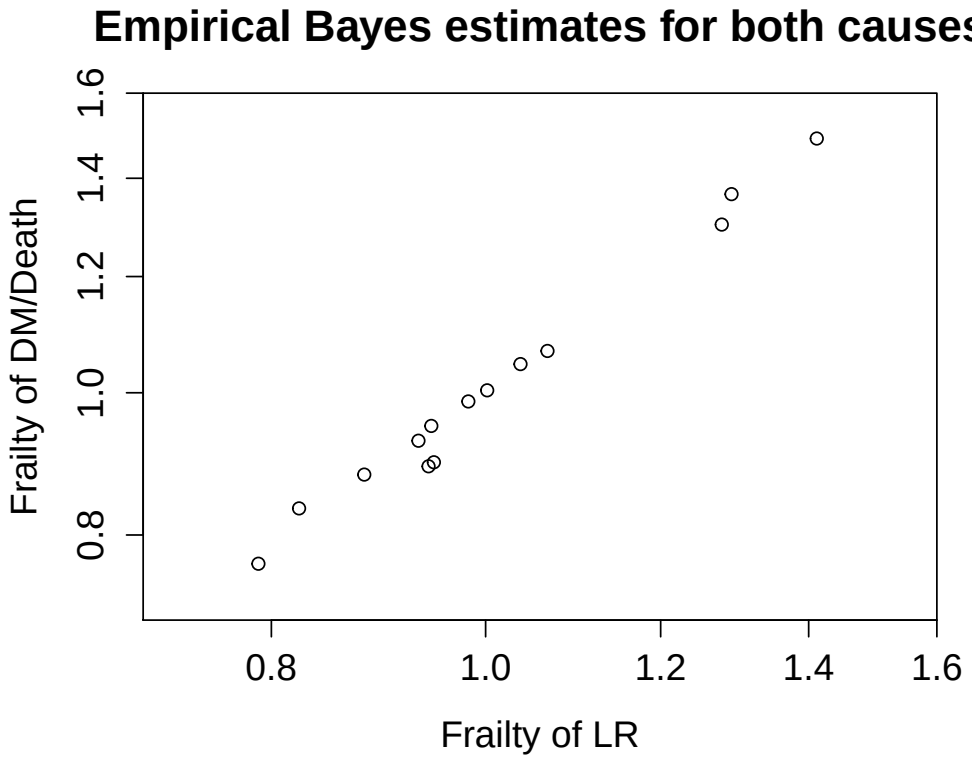

Figure 7.3: Frailties for the two causes of failure for 14 centers.

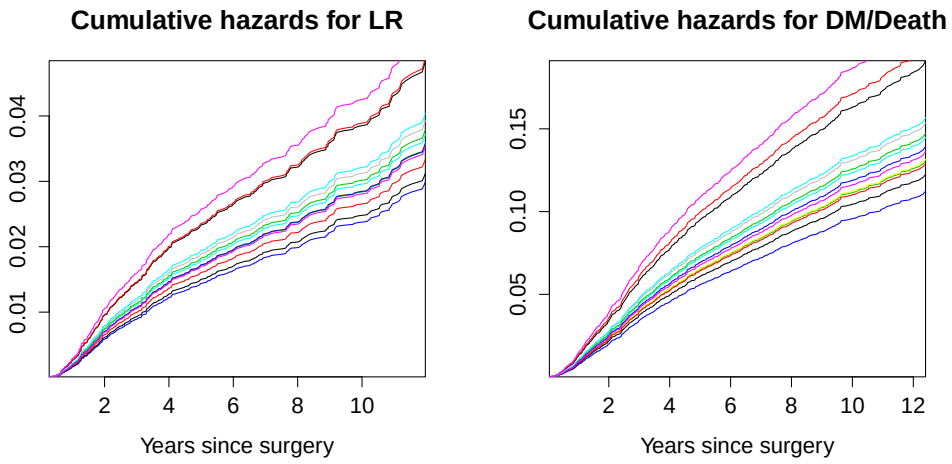

Figure 7.4: Left panel: cumulative hazard for cause LR for an average patient where each line represents a hospital. Right panel: cumulative hazard for cause DM/Death for an average patient where each line represents a hospital.

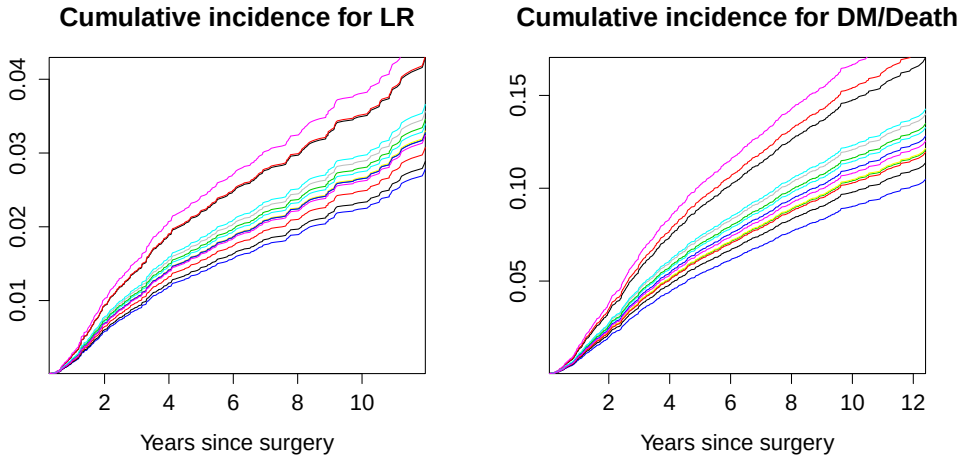

Figure 7.5: Left panel: cumulative incidence for cause LR for an average patient where each line represents a hospital. Right panel: cumulative incidence for cause DM/Death for an average patient where each line represents a hospital.

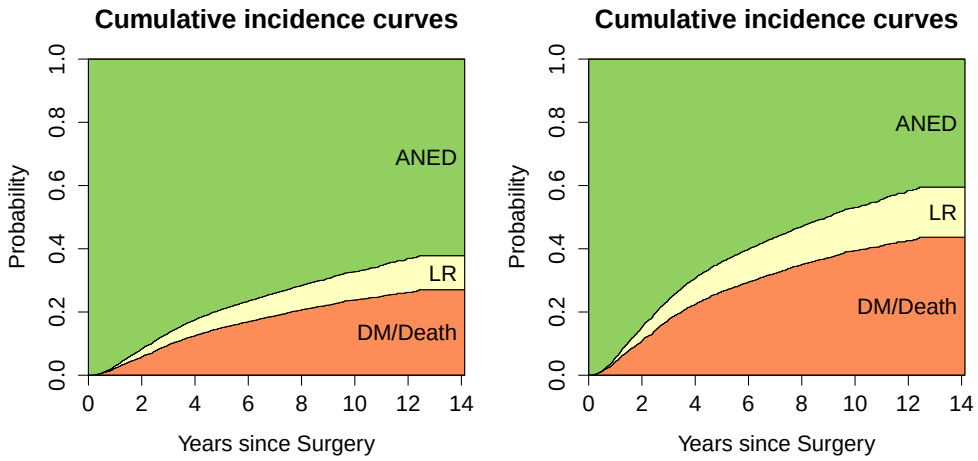

Figure 7.6: Left panel: stacked cumulative incidence curves for an average patient treated in hospital with lowest estimated frailty. Right panel: cumulative incidence curves for an average patient treated in hospital with highest estimated frailty.

Chapter 8

Discussion and Conclusion

It is common practice to account for unobserved heterogeneity by incorporating a frailty term in the analysis. This is particularly important in multi-center studies where data is collected from several treatment centers. In some cases the heterogeneity between centers may be the target of study.

The model presented in this thesis provides a comprehensive analysis of unobserved heterogeneity in competing risks situations. An additive gamma frailty model is used to allow for correlation between frailties. The amount of correlation indicates how closely risks of competing events are related. Estimation of the model is conducted using the EM-algorithm and a method to estimate standard errors for model estimates is illustrated. The estimation procedure simultaneously provides empirical Bayes estimates for hospital frailties, which together with their confidence intervals can be used to compare hospital effects. The method is applied on example data and a strong correlation between the frailties of the competing events is detected.

A limitation of the method are large standard errors for the regression coefficients. After adding the correlated frailties to the estimation for only two variables a significant effect could be shown. However, when the goal is to investigate hospital heterogeneity, the method discussed in this thesis provides a comprehensive analysis of hospital effects.

8.1 Future research

The introduced model covers the simplest competing risks situation, with two competing events. It could be extended to an arbitrary amount of competing events and even be applied to multi-state models, in which transitions to intermediate events are possible. These models describe disease progression in a more accurate form and could give additional inside in heterogeneity between hospitals.

To further investigate the performance of this model a simulation study could be conducted. Unobserved covariates with common and cause-specific effects could be simulated with varying strength of effects inducing different amounts of correlation between frailties. An interesting aspect would be to investigate the relationship between the performance of the model and presence of correlation in the data.

Bibliography

Aalen, O. O., Ø. Borgan, and H. K. Gjessing (2008). *Survival and Event History Analysis*. Springer.

Breastcancer.org (2015). *What is breast cancer?* `http://www.breastcancer.org/symptoms/understand_bc/what_is_bc`. Accessed: 2015-09-19.

Casella, G. and R. L. Berger (2001). *Statistical Inference*. 2nd ed. Duxbury Press.

Clayton, D. G. (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence". *Biometrika* 65.1, pp. 141–151.

de Bock, G. H. et al. (2009). "The impact of loco-regional recurrences on metastatic progression in early-stage breast cancer: a multistate model". *Breast Cancer Research and Treatment* 117.2, pp. 401–408.

de Wreede, L. C., M. Fiocco, and H. Putter (2011). "mstate: An R package for the analysis of competing risks and multi-state models". *Journal of Statistical Software* 38.7, pp. 1–30.

Do, C. B. and S. Batzoglou (2008). "What is the expectation maximization algorithm?" *Nature Biotechnology* 26.8, pp. 897–899.

Elbers, C. and G. Ridder (1982). "True and spurious duration dependence: The identifiability of the proportional hazard model". *The Review of Economic Studies* 49.3, pp. 403–409.

Fiocco, M., H. Putter, and J. C. van Houwelingen (2009). "Meta-analysis of pairs of survival curves under heterogeneity: A Poisson correlated gamma-frailty approach". *Statistics in Medicine* 28.30, pp. 3782–3797.

Gillick, M. (2001). "Pinning down frailty". *J. Gerontol. Ser. A-Biol. Sci. Med. Sci.* 56.3, pp. M134–M135.

Heckman, J. and B. Singer (1984). "The identifiability of the proportional hazard model". *The Review of Economic Studies* 51.2, pp. 231–241.

Hougaard, P. (1995). "Frailty models for survival data". *Lifetime Data Analysis* 1.3, pp. 255–273.

Klein, J. P. and M. L. Moeschberger (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. Springer.

Korsgaard, I. R. and A. H. Andersen (1998). "The additive genetic gamma frailty model". *Scandinavian Journal of Statistics* 25.2, pp. 225–269.

Louis, T. A. (1982). "Finding the observed information matrix when using the em algorithm". *J. R. Statist. Soc.* 44.2, pp. 226–233.

Petersen, J. H., P. K. Andersen, and R. D. Gill (1996). "Variance components models for survival data". *Statistica Neerlandica* 50.1, pp. 193–211.

Putter, H. and H. C. van Houwelingen (2015). "Dynamic frailty models based on compound birth-death processes". *Biostatistics* 16.3, pp. 550–564.

Putter, H., M.Fiocco, and R. B. Geskus (2007). "Tutorial in biostatistics: Competing risks and multi-state models". *Statistics in Medicine* 26.11, pp. 2389–430.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Rockwood, K. (2005). "Frailty and its definition: a worthy challenge". *J. Am. Geriatr. Soc.* 53.6, pp. 1069–1070.

Therneau, T. (2015). *A package for survival analysis in S*. R package Version 2.28. URL: `http://CRAN.R-project.org/package=survival`.

van der Hage, J. A. et al. (2001). "Improved survival after one course of perioperative chemotherapy in early breast cancer patients: long-term results from the European Organization for Research and Treatment of Cancer (EORTC) Trial 10854." *Eur. J. Cancer* 37.17, pp. 2184–2193.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality". *Demography* 16.3, pp. 439–454.

Yashin, A. I., J. W. Vaupel, and I. A. Iachine (1995). "Correlated individual frailty: an advantageous approach to survival analysis of bivariate data". *Mathematical Population Studies* 5.2, pp. 145–159.

Young, G. A. and R. L. Smith (2005). *Essentials of Statistical Inference*. Cambridge University Press.

Appendices

Appendix A

Probabilities for E-step

Let

- $z_{kj} \sim \Gamma(v_j, 1)$, $(j = 0, 1, 2)$
- d_{kj} : number of failures in center k by cause $(j = 1, 2)$
- $\Lambda_{kj} = \sum_{i=1}^{n_k} \Lambda_{j0}(t_{ki}) e^{\boldsymbol{\beta}_j^T \boldsymbol{X}_{ki}}, (j = 1, 2)$

$$
\begin{aligned}
f(\text{data}_k | z_{k0}, z_{k1}, z_{k2}) =& \prod_{i=1}^{n_k} \left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}) \right)^{1_{\{\delta_{ki}=1\}}} \left(\frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}) \right)^{1_{\{\delta_{ki}=2\}}} \\
& \exp\left(- \left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \Lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}) + \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \Lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}) \right) \right) \\
=& (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1} - l} z_{k1}^{l} \right) \left(\sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2} - m} z_{k2}^{m} \right) \\
& \prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right] \\
& \exp\left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp\left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right) \exp\left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)
\end{aligned}
$$

$$
\begin{aligned}
f(\text{data}_k | z_{k0}, z_{k2}) =& \int_{z_{k1}} f(z_{k1}) f(\text{data}_k | z_{k0}, z_{k1}, z_{k2}) dz_{k1} \\
& = (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2} - m} z_{k2}^{m} \right) \\
& \prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right] \\
& \exp\left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp\left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \\
& \frac{1}{\Gamma(\nu_1)} \sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1} - l} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}}
\end{aligned}
$$

$$
\begin{aligned}
f(\text{data}_k|z_{k0}) = &\int_{z_{k2}} f(z_{k2}) f(\text{data}_k|z_{k0}, z_{k2}) dz_{k2} \\
=&(\nu_0+\nu_1)^{-d_{k1}}(\nu_0+\nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_1)\Gamma(\nu_2)} \\
&\prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=1\}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=2\}} \right] \\
&\exp\left(-z_{k0} \left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1} + \frac{1}{\nu_0+\nu_2}\Lambda_{k2} \right) \right) \\
&\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} z_{k0}^{d_{k1}+d_{k2}-l-m}
\end{aligned}
$$

$$
\begin{aligned}
f(\text{data}_k) = &\int_{z_{k0}} f(z_{k0}) f(\text{data}_k|z_{k0}) dz_{k0} \\
=&(\nu_0+\nu_1)^{-d_{k1}}(\nu_0+\nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0)\Gamma(\nu_1)\Gamma(\nu_2)} \\
&\prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=1\}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=2\}} \right] \\
&\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} \\
&\frac{\Gamma(d_{k1}+d_{k2}+\nu_0-l-m)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{d_{k1}+d_{k2}+\nu_0-l-m}}
\end{aligned}
$$

$$
\begin{aligned}
f(\text{data}_k|z_{k0}, z_{k1}) = &(\nu_0+\nu_1)^{-d_{k1}}(\nu_0+\nu_2)^{-d_{k2}} \left(\sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} z_{k1}^{l} \right) \\
&\prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=1\}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=2\}} \right] \\
&\exp\left(-z_{k0} \left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1} + \frac{1}{\nu_0+\nu_2}\Lambda_{k2} \right) \right) \exp\left(-z_{k1} \frac{1}{\nu_0+\nu_1}\Lambda_{k1} \right) \\
&\frac{1}{\Gamma(\nu_2)} \sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}}
\end{aligned}
$$

$$
\begin{aligned}
f(\text{data}_k|z_{k1}) =& \int_{z_{k0}} f(z_{k0}) f(\text{data}_k|z_{k0}, z_{k1}) dz_{k0} \\
=& (\nu_0+\nu_1)^{-d_{k1}} (\nu_0+\nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0)\Gamma(\nu_2)} \\
& \prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=1\}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=2\}} \right] \\
& \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} \frac{\Gamma(d_{k1}+d_{k2}+v_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+v_0-l-m}} \\
& z_{k1}^l \exp\left(-z_{k1}\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(\text{data}_k|z_{k2}) =& \int_{z_{k0}} f(z_{k0}) f(\text{data}_k|z_{k0}, z_{k2}) dz_{k0} \\
=& (\nu_0+\nu_1)^{-d_{k1}} (\nu_0+\nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0)\Gamma(\nu_1)} \\
& \prod_{i=1}^{n_k} \left[(\lambda_{10}(t_{ki}) \exp(\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=1\}} (\lambda_{20}(t_{ki}) \exp(\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}))^{1\{\delta_{ki}=2\}} \right] \\
& \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(d_{k1}+d_{k2}+v_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+v_0-l-m}} \\
& z_{k2}^m \exp\left(-z_{k2}\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)
\end{aligned}
$$

$$
\begin{aligned}
f(z_{k0}|\text{data}_k) =& \frac{f(\text{data}_k|z_{k0}) f(z_{k0})}{f(\text{data}_k)} \\
=& \frac{\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} z_{k0}^{d_{k1}+d_{k2}+\nu_0-l-m-1}}{\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} \frac{\Gamma(d_{k1}+d_{k2}+\nu_0-l-m)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{d_{k1}+d_{k2}+\nu_0-l-m}}} \\
& \exp\left(-z_{k0}\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
f(z_{k1}|\text{data}_k) =& \frac{f(\text{data}_k|z_{k1}) f(z_{k1})}{f(\text{data}_k)} \\
=& \frac{\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} \frac{\Gamma(d_{k1}+d_{k2}+v_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+v_0-l-m}} z_{k1}^{l+\nu_1-1}}{\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}} \frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}} \frac{\Gamma(d_{k1}+d_{k2}+\nu_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+\nu_0-l-m}}} \\
& \exp\left(-z_{k1}\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
f(z_{k2}|\text{data}_k) =& \frac{f(\text{data}_k|z_{k2})f(z_{k2})}{f(\text{data}_k)} \\
=& \frac{\sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}}\binom{d_{k1}}{l}\binom{d_{k2}}{m}\frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}}\frac{\Gamma(d_{k1}+d_{k2}+v_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+v_0-l-m}}z_{k2}^{m+\nu_2-1}}{\sum_{l=0}^{d_{k1}}\sum_{m=0}^{d_{k2}}\binom{d_{k1}}{l}\binom{d_{k2}}{m}\frac{\Gamma(l+\nu_1)}{\left(1+\frac{1}{\nu_0+\nu_1}\Lambda_{k1}\right)^{l+\nu_1}}\frac{\Gamma(m+\nu_2)}{\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)^{m+\nu_2}}\frac{\Gamma(d_{k1}+d_{k2}+\nu_0-l-m)}{\left(1+\left(\frac{1}{\nu_0+\nu_1}\Lambda_{k1}+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)^{d_{k1}+d_{k2}+\nu_0-l-m}} \\
& \exp\left(-z_{k2}\left(1+\frac{1}{\nu_0+\nu_2}\Lambda_{k2}\right)\right)
\end{aligned}
$$

Appendix B

Observed information of regression parameters

In the following the quantities needed for $I_{\boldsymbol{\eta\eta}}$ are described.
Let

- $t_{kl}, l = 1, ..., d_{k1}$: ordered event times for cause 1 in hospital k
- $t_{km}, m = 1, ..., d_{k2}$: ordered event times for cause 2 in hospital k
- $\Lambda_{10}(t_{ki}) = \sum_{t_{kl} \leq t_{ki}} \lambda_{10}(t_{kl})$
- $\Lambda_{20}(t_{ki}) \sum_{t_{km} \leq t_{ki}} \lambda_{20}(t_{km})$
- $\sum_{i=1}^{n_k} e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}} \Lambda_{10}(t_{ki}) = \sum_{l=1}^{d_{k1}} \lambda_{10}(t_{kl}) \sum_{i:t_{ki} \geq t_{kl}} e^{\boldsymbol{\beta}_1^T \boldsymbol{X}_{ki}}$
- $\sum_{i=1}^{n_k} e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}} \Lambda_{20}(t_{ki}) = \sum_{m=1}^{d_{k2}} \lambda_{20}(t_{km}) \sum_{i:t_{ki} \geq t_{km}} e^{\boldsymbol{\beta}_2^T \boldsymbol{X}_{ki}}$
- d_{k1}, d_{k2}: number of failures of cause 1 and 2 in hospital k respectively
- d_1, d_2: number of failures of cause 1 and 2 in total respectively
- $t_{l'}$, $(l' = 1, ..., d_1)$: ordered event times for cause 1
- $t_{m'}$, $(m' = 1, ..., d_2)$: ordered event times for cause 2
- $d_{kl'}$: number of failures of cause 1 at time $t_{l'}$ in hospital k
- $d_{km'}$: number of failures of cause 2 at time $t_{m'}$ in hospital k
- $d_{1l'}$: number of failures of cause 1 at time $t_{l'}$
- $d_{2m'}$: number of failures of cause 2 at time $t_{m'}$
- $R_k(t_{l'}) = \{i : t_{kl} \geq t_{l'}\}$: risk set at time $t_{l'}$ for hospital k

The complete data log-likelihood can expressed as

$$\begin{aligned}\ell = &\sum_k d_{k1}\log(\frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}) + \sum_{l=1}^{d_{k1}}\log(\lambda_{10}(t_{kl})) + \sum_{l=1}^{d_{k1}}\boldsymbol{\beta}_1^T\boldsymbol{X}_{kl} \\ &- \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}\sum_{l=1}^{d_{k1}}\lambda_{10}(t_{kl})\sum_{i:t_{ki}\geq t_{kl}} e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}} \\ &+ d_{k2}\log(\frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}) + \sum_{m=1}^{d_{k2}}\log(\lambda_{20}(t_{km})) + \sum_{m=1}^{d_{k2}}\boldsymbol{\beta}_2^T\boldsymbol{X}_{km} \\ &- \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}\sum_{m=1}^{d_{k2}}\lambda_{20}(t_{km})\sum_{i:t_{ki}\geq t_{km}} e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}.\end{aligned}$$

The elements of the gradient vector of the complete data log-likelihood are:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}_{1j}}\ell =& \sum_k\left[\sum_{l=1}^{d_{k1}}\boldsymbol{X}_{klj} - \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}\sum_{l=1}^{d_{k1}}\lambda_{10}(t_{kl})\sum_{i\in R_k(t_{kl})}\boldsymbol{X}_{kij}e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}}\right] \\ \frac{\partial}{\partial\boldsymbol{\beta}_{2j}}\ell =& \sum_k\left[\sum_{m=1}^{d_{k2}}\boldsymbol{X}_{kmj} - \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}\sum_{m=1}^{d_{k2}}\lambda_{20}(t_{km})\sum_{i\in R_k(t_{km})}\boldsymbol{X}_{kij}e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}\right] \\ \frac{\partial}{\partial\lambda_{10l'}}\ell =& \sum_k\left[\frac{d_{kl'}}{\lambda_{10l'}(t_{l'})} - \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}d_{kl'}\sum_{i\in R_k(t_{l'})}e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}}\right] \\ =& \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})} - \sum_k\frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}d_{kl'}\sum_{i\in R_k(t_{l'})}e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}} \\ \frac{\partial}{\partial\lambda_{20m'}}\ell =& \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \sum_k\frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}d_{km'}\sum_{i\in R_k(t_{m'})}e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}\end{aligned}$$

The second order derivatives to calculate the full information matrix $I^{(full)}$ are:

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{1j}\partial \boldsymbol{\beta}_{1h}}\ell = -\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}\sum_{l=1}^{d_{k1}}\lambda_{10}(t_{kl})\sum_{i\in R_k(t_{kl})}\boldsymbol{X}_{kij}\boldsymbol{X}_{kih}e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}}$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{1j}\partial \boldsymbol{\beta}_{2h}}\ell = 0$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{1j}\partial \lambda_{10l'}}\ell = -\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}d_{kl'}\sum_{i\in R_k(t_{l'})}\boldsymbol{X}_{kij}e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}}$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{1j}\partial \lambda_{20m'}}\ell = 0$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{2j}\partial \boldsymbol{\beta}_{2h}}\ell = -\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}\sum_{m=1}^{d_{k2}}\lambda_{20}(t_{km})\sum_{i\in R_k(t_{km})}\boldsymbol{X}_{kij}\boldsymbol{X}_{kih}e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{2j}\partial \lambda_{10l'}}\ell = 0$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}_{2j}\partial \lambda_{20m'}}\ell = -\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}d_{km'}\sum_{i\in R_k(t_{m'})}\boldsymbol{X}_{kij}e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}$$

$$\frac{\partial^2}{\partial \lambda_{10p'}\partial \lambda_{10l'}}\ell = 0, \qquad \frac{\partial^2}{\partial \lambda_{10l'}\partial \lambda_{10l'}}\ell = -\frac{d_{1l'}}{\lambda_{10l'}(t_{l'})^2}$$

$$\frac{\partial^2}{\partial \lambda_{10p'}\partial \lambda_{20m'}}\ell = 0$$

$$\frac{\partial^2}{\partial \lambda_{20p'}\partial \lambda_{20m'}}\ell = 0, \qquad \frac{\partial^2}{\partial \lambda_{20m'}\partial \lambda_{20m'}}\ell = -\frac{d_{2m'}}{\lambda_{20m'}(t_{m'})^2}$$

Let

- $\boldsymbol{X}_{1j} = \sum_k \sum_{l=1}^{d_{k1}} \boldsymbol{X}_{klj}$
- $\boldsymbol{X}_{2j} = \sum_k \sum_{m=1}^{d_{k2}} \boldsymbol{X}_{klj}$
- $S_{k1j} = \sum_{l=1}^{d_{k1}} \lambda_{10}(t_{kl}) \sum_{i\in R_k(t_{kl})} \boldsymbol{X}_{kij}e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}}$
- $S_{k2j} = \sum_{m=1}^{d_{k2}} \lambda_{20}(t_{km}) \sum_{i\in R_k(t_{km})} \boldsymbol{X}_{kij}e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}$
- $S_{k1l'} = d_{kl'} \sum_{i\in R_k(t_{l'})} e^{\boldsymbol{\beta}_1^T\boldsymbol{X}_{ki}}$
- $S_{k2m'} = d_{km'} \sum_{i\in R_k(t_{m'})} e^{\boldsymbol{\beta}_2^T\boldsymbol{X}_{ki}}$

We re-wright the elements of the gradient vector:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}_{1j}}\ell =& \boldsymbol{X}_{1j} - \sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1j} \\
\frac{\partial}{\partial \boldsymbol{\beta}_{2j}}\ell =& \boldsymbol{X}_{2j} - \sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2} S_{k2j} \\
\frac{\partial}{\partial \lambda_{10l'}}\ell =& \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})} - \sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1l'} \\
\frac{\partial}{\partial \lambda_{20m'}}\ell =& \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2} S_{k2m'}
\end{aligned}$$

The elements of the product of the gradient vector for $I^{(loss)}$ are:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}_{1j}}\ell \frac{\partial}{\partial \boldsymbol{\beta}_{1h}}\ell =& \boldsymbol{X}_{1j}\boldsymbol{X}_{1h} - \boldsymbol{X}_{1j}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1h} - \boldsymbol{X}_{1h}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1j} \\
&+ \sum_k \sum_c (z_{k0}z_{c0} + z_{k0}z_{c1} + z_{k1}z_{c0} + z_{k1}z_{c1}) \left(\frac{1}{\nu_0+\nu_1}\right)^2 S_{k1h}S_{c1j} \\
\frac{\partial}{\partial \boldsymbol{\beta}_{1j}}\ell \frac{\partial}{\partial \boldsymbol{\beta}_{2h}}\ell =& \boldsymbol{X}_{1j}\boldsymbol{X}_{2h} - \boldsymbol{X}_{1j}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2} S_{k2h} - \boldsymbol{X}_{2h}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1j} \\
&+ \sum_k \sum_c (z_{k0}z_{c0} + z_{k0}z_{c2} + z_{k1}z_{c0} + z_{k1}z_{c2}) \frac{1}{\nu_0+\nu_1}\frac{1}{\nu_0+\nu_2} S_{k1j}S_{c2h} \\
\frac{\partial}{\partial \boldsymbol{\beta}_{1j}}\ell \frac{\partial}{\partial \lambda_{10l'}}\ell =& \boldsymbol{X}_{1j}\frac{d_{1l'}}{\lambda_{10l'}(t_{l'})} - \boldsymbol{X}_{1j}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1l'} - \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1j} \\
&+ \sum_k \sum_c (z_{k0}z_{c0} + z_{k0}z_{c1} + z_{k1}z_{c0} + z_{k1}z_{c1}) \left(\frac{1}{\nu_0+\nu_1}\right)^2 S_{k1j}S_{c1l'} \\
\frac{\partial}{\partial \boldsymbol{\beta}_{1j}}\ell \frac{\partial}{\partial \lambda_{20m'}}\ell =& \boldsymbol{X}_{1j}\frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \boldsymbol{X}_{1j}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2} S_{k2m'} - \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1} S_{k1j} \\
&+ \sum_k \sum_c (z_{k0}z_{c0} + z_{k0}z_{c2} + z_{k1}z_{c0} + z_{k1}z_{c2}) \frac{1}{\nu_0+\nu_1}\frac{1}{\nu_0+\nu_2} S_{k1j}S_{c2m'}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}_{2j}}\ell\frac{\partial}{\partial \boldsymbol{\beta}_{2h}}\ell =& \boldsymbol{X}_{2j}\boldsymbol{X}_{2h} - \boldsymbol{X}_{2j}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2h} - \boldsymbol{X}_{2h}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2j} \\
&+ \sum_k\sum_c (z_{k0}z_{c0}+z_{k0}z_{c2}+z_{k2}z_{c0}+z_{k2}z_{c2})\left(\frac{1}{\nu_0+\nu_2}\right)^2 S_{k2h}S_{c2j} \\
\frac{\partial}{\partial \boldsymbol{\beta}_{2j}}\ell\frac{\partial}{\partial \lambda_{10l'}}\ell =& \boldsymbol{X}_{2j}\frac{d_{1l'}}{\lambda_{10l'}(t_{l'})} - \boldsymbol{X}_{2j}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}S_{k1l'} - \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2j} \\
&+ \sum_k\sum_c (z_{k0}z_{c0}+z_{k0}z_{c2}+z_{k1}z_{c0}+z_{k1}z_{c2})\frac{1}{\nu_0+\nu_1}\frac{1}{\nu_0+\nu_2}S_{k1l'}S_{c2j} \\
\frac{\partial}{\partial \boldsymbol{\beta}_{2j}}\ell\frac{\partial}{\partial \lambda_{20m'}}\ell =& \boldsymbol{X}_{2j}\frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \boldsymbol{X}_{2j}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2m'} - \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2j} \\
&+ \sum_k\sum_c (z_{k0}z_{c0}+z_{k0}z_{c2}+z_{k2}z_{c0}+z_{k2}z_{c2})\left(\frac{1}{\nu_0+\nu_2}\right)^2 S_{k2j}S_{c2m'}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \lambda_{10l'}}\ell\frac{\partial}{\partial \lambda_{10p'}}\ell =& \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})}\frac{d_{1p'}}{\lambda_{10p'}(t_{p'})} - \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}S_{k1p'} \\
&- \frac{d_{1p'}}{\lambda_{10p'}(t_{p'})}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}S_{k1l'} \\
&+ \sum_k\sum_c (z_{k0}z_{c0}+z_{k0}z_{c1}+z_{k1}z_{c0}+z_{k1}z_{c1})\left(\frac{1}{\nu_0+\nu_1}\right)^2 S_{k1p'}S_{c1l'} \\
\frac{\partial}{\partial \lambda_{10l'}}\ell\frac{\partial}{\partial \lambda_{20m'}}\ell =& \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})}\frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2m'} \\
&- \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})}\sum_k \frac{z_{k0}+z_{k1}}{\nu_0+\nu_1}S_{k1l'} \\
&+ \sum_k\sum_c (z_{k0}z_{c0}+z_{k0}z_{c2}+z_{k1}z_{c0}+z_{k1}z_{c2})\frac{1}{\nu_0+\nu_1}\frac{1}{\nu_0+\nu_2}S_{k1l'}S_{c2m'}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \lambda_{20m'}}\ell\frac{\partial}{\partial \lambda_{20p'}}\ell =& \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})}\frac{d_{2p'}}{\lambda_{20p'}(t_{p'})} - \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2p'} \\
&- \frac{d_{2p'}}{\lambda_{20p'}(t_{p'})}\sum_k \frac{z_{k0}+z_{k2}}{\nu_0+\nu_2}S_{k2m'} \\
&+ \sum_k\sum_c (z_{k0}z_{c0}+z_{k0}z_{c2}+z_{k2}z_{c0}+z_{k2}z_{c2})\left(\frac{1}{\nu_0+\nu_2}\right)^2 S_{k2p'}S_{c2m'}
\end{aligned}$$

Appendix C

R code

C.1 R code for model estimation through EM-algorithm

```
library(survival)
library(mstate)
library(numDeriv)
library(dynpred)
library(plyr)
setwd("/Users/anja/Desktop/Desktop/Uni Leiden/Thesis/material thesis")
load("Data.Rda")
# prepare data for analysis
# censore DM if at the same time as LR (81x)
data$event_DM[data$event_LR & data$event_DM &
                data$time_LR == data$time_DM] <- 0
# combine events DM and death
data$event_DMsurv <- ifelse(data$event_surv | data$event_DM, 1, 0)
data$time_DMsurv <- ifelse(data$event_DM, data$time_DM, data$time_surv)
tmat <- transMat(x=list(c(2, 3), c(), c()),
                     names=c("ANED", "LR", "DM/Death"))
covs=c("center", "age", "size", "node", "surgery",
           "CTperi", "CTadj", "RTadj")
longData <- msprep(data = data, trans = tmat,
                       time = c(NA, "time_LR", "time_DMsurv"),
                       status = c(NA, "event_LR", "event_DMsurv"),
                       keep = covs)
expandedData <- expand.covs(longData, covs, longnames=FALSE)

em <- function(lognu = c(0,0,0), last=FALSE) {
# calculates observed data log-likelihood with EM algorithm,
# if last=True returns final models instead of log-lik
  # set up constant values valid for all hospitals
```

```
print(exp(lognu))
nu0 <- exp(lognu[1])
nu1 <- exp(lognu[2])
nu2 <- exp(lognu[3])
stand1 <- 1/(nu0 + nu1)
stand2 <- 1/(nu0 + nu2)
log.stand1 <- log(stand1)
log.stand2 <- log(stand2)
center <- unique(data$center)
K <- length(center)
X <- as.matrix(expandedData[expandedData$trans==1,
                            c("age1.1", "age2.1", "size.1",
                              "node.1", "surgery.1", "CTperi.1",
                              "CTadj.1", "RTadj.1")])
eventtimes1 <- sort(unique(longData$time[longData$trans == 1 &
                                         longData$status == 1]))
eventtimes2 <- sort(unique(longData$time[longData$trans == 2 &
                                         longData$status == 1]))
# set up constant values for each hospital
datak1 <- datak2 <- dk1  <- dk2 <-
  l <- m <- a1 <- a2 <- a3 <- wh1 <-
  wh2 <- xk <- tk1 <- tk2 <- vector('list', K)
for(i in 1:K){
  datak <- longData[longData$center == center[i],]
  datak1[[i]] <- datak[datak$trans == 1,]
  datak2[[i]] <- datak[datak$trans == 2,]
  dk1[[i]] <- sum(datak1[[i]]$status)
  dk2[[i]] <- sum(datak2[[i]]$status)
  xk[[i]] <- X[data$center == center[i],]
  l[[i]] <- matrix(0:dk1[[i]], nrow = dk1[[i]] + 1, ncol = dk2[[i]] + 1)
  m[[i]] <- matrix(0:dk2[[i]], nrow = dk1[[i]] + 1, ncol = dk2[[i]] + 1,
                   byrow=TRUE)
  a1[[i]] <- l[[i]] + nu1
  a2[[i]] <- m[[i]] + nu2
  a3[[i]] <- dk1[[i]] + dk2[[i]] + nu0 - l[[i]] - m[[i]]
  wh1[[i]] <- which(datak1[[i]]$status == 1)
  wh2[[i]] <- which(datak2[[i]]$status == 1)
  tk1[[i]] <- match(datak1[[i]]$time[wh1[[i]]], eventtimes1)
  tk2[[i]] <- match(datak2[[i]]$time[wh2[[i]]], eventtimes2)
}
obslik <- numeric(5)
models <- vector('list', 5)
for(j in 1:5){
```

```
print(j)
# set up starting parameters for EM (5 times)
Zk0 <- exp(rnorm(K, lognu[1]-0.5,1))
Zk1 <- exp(rnorm(K, lognu[2]-0.5,1))
Zk2 <- exp(rnorm(K, lognu[3]-0.5,1))
longData$zk0 <- Zk0[match(longData$center, center)]
longData$zk1 <- Zk1[match(longData$center, center)]
longData$zk2 <- Zk2[match(longData$center, center)]
m.cause1 <- coxph(Surv(time, status)~ age + size + node +
                    surgery + CTperi + CTadj + RTadj +
                    offset(log((zk0 + zk1)*stand1)),
                  data = longData, subset = (trans == 1))
m.cause2 <- coxph(Surv(time, status)~ age + size + node +
                    surgery + CTperi + CTadj + RTadj +
                    offset(log((zk0 + zk2)*stand2)),
                  data = longData, subset = (trans == 2))
Lambda1 <- basehaz(m.cause1, centered=FALSE)
Lambda1 <- Lambda1[!duplicated(Lambda1$hazard),]
Lambda1 <- Lambda1[Lambda1$hazard != 0,]
Lambda2 <- basehaz(m.cause2, centered = FALSE)
Lambda2 <- Lambda2[!duplicated(Lambda2$hazard),]
Lambda2 <- Lambda2[Lambda2$hazard != 0,]
lambda1 <- diff(c(0, Lambda1$hazard))
lambda2 <- diff(c(0, Lambda2$hazard))
beta1 <- m.cause1$coef
beta2 <- m.cause2$coef
loglik.old <- 1
obs <- numeric(K)
repeat{
  for(i in 1:K){
    # E-step for hospital k
    H10time <- evalstep(Lambda1$time, Lambda1$hazard,
                        newtime = datak1[[i]]$time, subst=0)
    H20time <- evalstep(Lambda2$time, Lambda2$hazard,
                        newtime = datak2[[i]]$time, subst=0)
    hr1 <- exp(xk[[i]]%*%beta1); hr2 <- exp(xk[[i]]%*%beta2)
    st.Lambdak1 <- stand1*sum(H10time*hr1)
    st.Lambdak2 <- stand2*sum(H20time*hr2)
    b1 <- 1 + st.Lambdak1; b2 <- 1 + st.Lambdak2; b3 <- b1 + b2 - 1
    tmp <- lchoose(dk1[[i]], l[[i]]) + lchoose(dk2[[i]], m[[i]]) +
      (lgamma(a1[[i]]) - a1[[i]]*log(b1)) +
      (lgamma(a2[[i]]) - a2[[i]]*log(b2)) +
      (lgamma(a3[[i]]) - a3[[i]]*log(b3))
```

```
    maxtmp <- max(tmp)
    tmp <- tmp - maxtmp
    mat <- exp(tmp)
    clm <- mat/sum(mat)
    Zk0[i] <- sum(clm*(a3[[i]]/b3))
    Zk1[i] <- sum(clm*(a1[[i]]/b1))
    Zk2[i] <- sum(clm*(a2[[i]]/b2))
    # likelihood contribution of hospital k
    obs[i] <- (dk1[[i]]*log.stand1 + dk2[[i]]*log.stand2 -
                lgamma(nu0) - lgamma(nu1) - lgamma(nu2)) +
                sum(log(lambda1[tk1[[i]]]*hr1[wh1[[i]]])) +
                sum(log(lambda2[tk2[[i]]]*hr2[wh2[[i]]])) +
                log(sum(mat)) + maxtmp
  }
  loglik <- sum(obs)
  delta <- loglik.old - loglik
  cat("Log-lik =", loglik, ", delta =", delta, "\n")
  if(abs(delta) < 1e-6) break
  loglik.old <- loglik
  #  M-step
  longData$zk0 <- Zk0[match(longData$center, center)]
  longData$zk1 <- Zk1[match(longData$center, center)]
  longData$zk2 <- Zk2[match(longData$center, center)]
  m.cause1 <- coxph(Surv(time, status)~ age + size + node +
                      surgery + CTperi + CTadj + RTadj +
                      offset(log((zk0 + zk1)*stand1)),
                    data = longData, subset = (trans == 1))
  m.cause2 <- coxph(Surv(time, status)~ age + size + node +
                      surgery + CTperi + CTadj + RTadj +
                      offset(log((zk0 + zk2)*stand2)),
                    data = longData, subset = (trans == 2))
  Lambda1 <- basehaz(m.cause1, centered = FALSE)
  Lambda1 <- Lambda1[!duplicated(Lambda1$hazard),]
  Lambda1 <- Lambda1[Lambda1$hazard != 0,]
  Lambda2 <- basehaz(m.cause2, centered = FALSE)
  Lambda2 <- Lambda2[!duplicated(Lambda2$hazard),]
  Lambda2 <- Lambda2[Lambda2$hazard != 0,]
  lambda1 <- diff(c(0, Lambda1$hazard))
  lambda2 <- diff(c(0, Lambda2$hazard))
  beta1 <- m.cause1$coef
  beta2 <- m.cause2$coef
}
obslik[j] <- loglik
```

```
    models[[j]] <- list(m.cause1 = m.cause1, m.cause2 = m.cause2)
  }
  m.index <- which.max(obslik)
  ifelse(last == FALSE, return(obslik[m.index]),
         return(c(models[[m.index]], longData)))
}

param <- function(lognu){
# returns frailty variances and correlation
  nu <- exp(lognu)
  nu0<- nu[1]
  nu1<- nu[2]
  nu2<- nu[3]
  var1<-1/(nu0+nu1)
  var2<-1/(nu0+nu2)
  corr<-nu0*sqrt(var1*var2)
  return(list(nu=nu,var1=var1,var2=var2,corr=corr))
}
```

C.2 R code for estimation of standard error

```
i.param <- function(lognu, m.cause1, m.cause2) {
# calculate information matrix for regression parameters
  nu0 <- exp(lognu[1])
  nu1 <- exp(lognu[2])
  nu2 <- exp(lognu[3])
  X <- model.matrix(m.cause1)
  Lambda1 <- basehaz(m.cause1, centered=FALSE)
  Lambda1 <- Lambda1[!duplicated(Lambda1$hazard),]
  Lambda1 <- Lambda1[Lambda1$hazard != 0,]
  Lambda2 <- basehaz(m.cause2, centered=FALSE)
  Lambda2 <- Lambda2[!duplicated(Lambda2$hazard),]
  Lambda2 <- Lambda2[Lambda2$hazard != 0,]
  lambda1 <- diff(c(0, Lambda1$hazard))
  lambda2 <- diff(c(0, Lambda2$hazard))
  beta1 <- m.cause1$coef
  beta2 <- m.cause2$coef
  stand1 <- 1/(nu0 + nu1)
  stand2 <- 1/(nu0 + nu2)
  log.stand1 <- log(stand1)
  log.stand2 <- log(stand2)
  center <- unique(data$center)
```

```
K<-length(center)
loglik.old <- 1
Zk0 <- numeric(K)
Zk1 <- numeric(K)
Zk2 <- numeric(K)
obs <- numeric(K)
#---------Quantities for I(full)---------#
n1 <- length(lambda1)
n2 <- length(lambda2)
n <- length(beta1)
eventtimes1 <- longData$time[longData$trans == 1 & longData$status == 1]
eventtimes2 <- longData$time[longData$trans == 2 & longData$status == 1]
d1l <- sapply(Lambda1$time, function(x) sum(eventtimes1 == x))
d2m <- sapply(Lambda2$time, function(x) sum(eventtimes2 == x))
ddbetas1 <- matrix(0, nrow=n, ncol=n)
ddbetas2 <- matrix(0, nrow=n, ncol=n)
ddbetahaz1 <- matrix(0, nrow=n, ncol=n1)
ddbetahaz2 <- matrix(0, nrow=n, ncol=n2)
ddhaz1 <- -diag(d1l/(lambda1^2))
ddhaz2 <- -diag(d2m/(lambda2^2))
#---------Quantities for I(loss)---------#
Zk02 <- numeric(K)
Zk12 <- numeric(K)
Zk22 <- numeric(K)
Zk0Zk1 <- numeric(K)
Zk0Zk2 <- numeric(K)
Zk1Zk2 <- numeric(K)
X <- model.matrix(m.cause1)
d1 <- longData$status[longData$trans == 1] == 1
d2 <- longData$status[longData$trans == 2] == 1
X1 <- apply(X[d1,], 2, sum)
X2 <- apply(X[d2,], 2, sum)
dl <- d1l/lambda1
dm <- d2m/lambda2
S1 <- matrix(0, nrow = K, ncol = n)
S2 <- matrix(0, nrow = K, ncol = n)
Sl1 <- matrix(0, nrow = K, ncol = n1)
Sl2 <- matrix(0, nrow = K, ncol = n2)
for(i in 1:K){
  # Do a last E-step to get calculate all quantities
  datak <- expandedData[expandedData$center == center[i],]
  datak1 <- datak[datak$trans == 1,]
  datak2 <- datak[datak$trans == 2,]
```

```
dk1 <- sum(datak1$status)
dk2 <- sum(datak2$status)
H10time <- evalstep(Lambda1$time, Lambda1$hazard,
                    newtime = datak1$time, subst = 0)
H20time <- evalstep(Lambda2$time, Lambda2$hazard,
                    newtime = datak2$time, subst = 0)
xk <- X[data$center == center[i],]
hr1 <- exp(xk%*%beta1)
hr2 <- exp(xk%*%beta2)
st.Lambdak1 <- stand1*sum(H10time*hr1)
st.Lambdak2 <- stand2*sum(H20time*hr2)
l <- matrix(0:dk1, nrow = dk1 + 1, ncol = dk2 + 1)
m <- matrix(0:dk2, nrow = dk1 + 1, ncol = dk2 + 1, byrow = TRUE)
b1 <- 1 + st.Lambdak1
b2 <- 1 + st.Lambdak2
b3 <- b1 + b2 - 1
a1 <- l + nu1
a2 <- m + nu2
a3 <- dk1 + dk2 + nu0 - l - m
tmp <- lchoose(dk1,l) + lchoose(dk2,m) + (lgamma(a1) - a1*log(b1)) +
  (lgamma(a2) - a2*log(b2)) + (lgamma(a3) - a3*log(b3))
maxtmp <- max(tmp)
tmp <- tmp - maxtmp
mat <- exp(tmp)
clm <- mat/sum(mat)
wh1 <- which(datak1$status == 1)
wh2 <- which(datak2$status == 1)
tk1 <- match(datak1$time[wh1], Lambda1$time)
tk2 <- match(datak2$time[wh2], Lambda2$time)
Zk0[i] <- sum(clm*(a3/b3))
Zk1[i] <- sum(clm*(a1/b1))
Zk2[i] <- sum(clm*(a2/b2))
#---------Contribution to I(full) for hospital i----------#
# create risks sets r1, r2
eventtimesk1 <- datak1$time[wh1]
eventtimesk2 <- datak2$time[wh2]
r1 <- lapply(eventtimesk1, function(x) which(datak1$time >= x))
r2 <- lapply(eventtimesk2, function(x) which(datak2$time >= x))
# calculate ddbetas
haz1 <- lambda1[tk1]  # hazards at event times
haz2 <- lambda2[tk2]
ddbeta <- function(j, h, cause){  # contribution to ddbetas for hospital k
  if(cause == 1){
```

```
    if(dk1 == 0) return(0)
    return(-(Zk0[i] + Zk1[i])*stand1*
             sum(haz1*sapply(r1, function(i) sum(hr1[i]*xk[i, j]*xk[i, h]))))
  }else{
    if(dk2 == 0) return(0)
    return(-(Zk0[i] + Zk2[i])*stand2*
             sum(haz2*sapply(r2, function(i) sum(hr2[i]*xk[i, j]*xk[i, h]))))
  }
}
ddbeta <- Vectorize(ddbeta, vectorize.args = c('j', 'h'))
ddbetas1 <- ddbetas1 + outer(1:n, 1:n, ddbeta, 1)
ddbetas2 <- ddbetas2 + outer(1:n, 1:n, ddbeta, 2)
# calculate ddbetahazs
rl1 <- lapply(Lambda1$time, function(x) which(datak1$time >= x))
rl2 <- lapply(Lambda2$time, function(x) which(datak2$time >= x))
dkl <- sapply(Lambda1$time, function(x) sum(eventtimesk1 == x))
dkm <- sapply(Lambda2$time, function(x) sum(eventtimesk2 == x))
ddbetahaz <- function(j,cause){  # contribution to ddbetahaz for hospital k
  if(cause == 1){
    -(Zk0[i] + Zk1[i])*stand1*dkl*sapply(rl1, function(i) sum(xk[i, j]*hr1[i]))
  }else{
    -(Zk0[i] + Zk2[i])*stand2*dkm*sapply(rl2, function(i) sum(xk[i, j]*hr2[i]))
  }}
ddbetahaz1 <- ddbetahaz1 + t(sapply(1:n, function(x) ddbetahaz(x, 1)))
ddbetahaz2 <- ddbetahaz2 + t(sapply(1:n, function(x) ddbetahaz(x, 2)))
#---------Contribution to I(loss)---------#
Zk02[i] <- sum(clm*(a3/b3)*((a3 + 1)/b3))
Zk12[i] <- sum(clm*(a1/b1)*((a1 + 1)/b1))
Zk22[i] <- sum(clm*(a2/b2)*((a2 + 1)/b2))
Zk0Zk1[i] <- sum(clm*(a3/b3)*(a1/b1))
Zk0Zk2[i] <- sum(clm*(a3/b3)*(a2/b2))
Zk1Zk2[i] <- sum(clm*(a1/b1)*(a2/b2))
sj <- function(j, cause){
  if(cause == 1){
    if(dk1 == 0) return(0)
    return(sum(haz1*sapply(r1, function(i) sum(hr1[i]*xk[i, j]))))
  }
  if(cause == 2){
    if(dk2 == 0) return(0)
    return(sum(haz2*sapply(r2, function(i) sum(hr2[i]*xk[i, j]))))
  }
}
S1[i,] <- sapply(1:n, function(j) sj(j, 1))  # hospital i beta 1
```

```
  S2[i,] <- sapply(1:n, function(j) sj(j, 2))  # hospital i beta 2
  Sl1[i,] <- dkl*sapply(rl1, function(i) sum(hr1[i]))  # hospital i lambda1 l
  Sl2[i,] <- dkm*sapply(rl2, function(i) sum(hr2[i]))  # hospital i lambda2 m
}
#---------I(full)---------#
m <- matrix(0, nrow=n, ncol=n)
m1 <- matrix(0, nrow=n, ncol=n1)
m2 <- matrix(0, nrow=n, ncol=n2)
m12 <- matrix(0, nrow=n1, ncol=n2)
I.full <- rbind(cbind(ddbetas1, m, ddbetahaz1, m2),
                cbind(m, ddbetas2, m1, ddbetahaz2),
                cbind(t(ddbetahaz1), t(m1), ddhaz1, m12),
                cbind(t(m2), t(ddbetahaz2), t(m12), ddhaz2))
#---------I(loss)---------#
# calculate the gradient
db1 <-sapply(1:n, function(j) X1[j] - sum((Zk0 + Zk1)*stand1*S1[, j]))
db2 <-sapply(1:n, function(j) X2[j] - sum((Zk0 + Zk2)*stand2*S2[, j]))
dh1 <- sapply(1:n1, function(l) dl[l] - sum((Zk0 + Zk1)*stand1*Sl1[, l]))
dh2 <- sapply(1:n2, function(m) dm[m] - sum((Zk0 + Zk2)*stand2*Sl2[, m]))

S1ind <- (Zk0 + Zk1)*stand1*S1
S1corr <- (Zk02 + 2*Zk0Zk1 + Zk12)*stand1^2*S1
S2ind <- (Zk0 + Zk2)*stand2*S2
S2corr <- (Zk02 + 2*Zk0Zk2 + Zk22)*stand2^2*S2
S12corr <- (Zk02 + Zk0Zk2 + Zk0Zk1 + Zk1Zk2)*stand1*stand2*S1
S21corr <- (Zk02 + Zk0Zk2 + Zk0Zk1 + Zk1Zk2)*stand1*stand2*S2
Sl1ind <- (Zk0 + Zk1)*stand1*Sl1
Sl1corr <- (Zk02 + 2*Zk0Zk1 + Zk12)*stand1^2*Sl1
Sl12corr <- (Zk02 + Zk0Zk1 + Zk0Zk2 + Zk1Zk2)*stand1*stand2*Sl1
Sl2ind <- (Zk0 + Zk2)*stand2*Sl2
Sl2corr <- (Zk02 + 2*Zk0Zk2 + Zk22)*stand2^2*Sl2

db1db1 <- db1%*%t(db1) - t(S1ind)%*%S1ind + t(S1corr)%*%S1
db1db2 <- db1%*%t(db2) - t(S1ind)%*%S2ind + t(S12corr)%*%S2
db1dh1 <- db1%*%t(dh1) - t(S1ind)%*%Sl1ind + t(S1corr)%*%Sl1
db1dh2 <- db1%*%t(dh2) - t(S1ind)%*%Sl2ind + t(S12corr)%*%Sl2
db2db1 <- t(db1db2)
db2db2 <- db2%*%t(db2) - t(S2ind)%*%S2ind + t(S2corr)%*%S2
db2dh1 <- db2%*%t(dh1) - t(S2ind)%*%Sl1ind + t(S21corr)%*%Sl1
db2dh2 <- db2%*%t(dh2) - t(S2ind)%*%Sl2ind + t(S2corr)%*%Sl2
dh1db1 <- t(db1dh1)
dh1db2 <- t(db2dh1)
dh1dh1 <- dh1%*%t(dh1) - t(Sl1ind)%*%Sl1ind + t(Sl1corr)%*%Sl1
```

```
  dh1dh2 <- dh1%*%t(dh2) - t(Sl1ind)%*%Sl2ind + t(Sl12corr)%*%Sl2
  dh2db1 <- t(db1dh2)
  dh2db2 <- t(db2dh2)
  dh2dh1 <- t(dh1dh2)
  dh2dh2 <- dh2%*%t(dh2) - t(Sl2ind)%*%Sl2ind + t(Sl2corr)%*%Sl2

  I.loss <- rbind(cbind(db1db1, db1db2, db1dh1, db1dh2),
                  cbind(db2db1, db2db2, db2dh1, db2dh2),
                  cbind(dh1db1, dh1db2, dh1dh1, dh1dh2),
                  cbind(dh2db1, dh2db2, dh2dh1, dh2dh2))
  I <- -I.full - I.loss
  #---------Empirical bayes and variances---------#
  Wk1 <- (Zk0 + Zk1)*stand1
  Wk2 <- (Zk0 + Zk2)*stand2
  var.zk0 <- Zk02 - Zk0^2
  var.zk1 <- Zk12 - Zk1^2
  var.zk2 <- Zk22 - Zk2^2
  cov.zk0zk1 <- Zk0Zk1 - Zk0*Zk1
  cov.zk0zk2 <- Zk0Zk2 - Zk0*Zk2
  sdWk1 <- sqrt((var.zk0 + var.zk1 + 2*cov.zk0zk1)*stand1^2)
  sdWk2 <- sqrt((var.zk0 + var.zk2 + 2*cov.zk0zk2)*stand2^2)
  empbayes <- data.frame(center = center, Wk1 = Wk1, Wk2 = Wk2,
                         sdWk1 = sdWk1, sdWk2 = sdWk2)
  return(list(I = I, empbayes = empbayes))
}

#---------Find optimal nu and return final model---------#
set.seed(38317)
DMdeath <- optim(par = c(0, 0, 0), fn = em, control = list(fnscale = -1),
                 lower = c(-10, -10, -10), method = "L-BFGS-B", hessian = TRUE)

set.seed(06100205)
lognu <- DMdeath$par #lognu <- c(2.897764, 1.164795, -8.283077)
final <- em(lognu, last = TRUE)
m.cause1 <- final$m.cause1
m.cause2 <- final$m.cause2

#---------Caulculate covariance matrix of regression parameters---------#
info <- i.param(lognu, m.cause1, m.cause2)
info.param <- info$I
empbayes <- info$empbayes
confbayes <- info$confbayes
covparam <- solve(info.param)
```

```
#---------Get dEta for covaraince matrix---------#
eps <- 0.00001
lognu01 <- lognu - c(eps/2, 0, 0)
lognu02 <- lognu + c(eps/2, 0, 0)
lognu11 <- lognu - c(0, eps/2, 0)
lognu12 <- lognu + c(0, eps/2, 0)
lognu21 <- lognu - c(0, 0, eps/2)
lognu22 <- lognu + c(0, 0, eps/2)

set.seed(2819849)
m01 <- em(lognu01, last = TRUE)
m02 <- em(lognu02, last = TRUE)
m11 <- em(lognu11, last = TRUE)
m12 <- em(lognu12, last = TRUE)
m21 <- em(lognu21, last = TRUE)
m22 <- em(lognu22, last = TRUE)

eta <- function(m){
# returns regression parameters from final models
  Lambda1 <- basehaz(m$m.cause1, centered=FALSE)
  Lambda1 <- Lambda1[!duplicated(Lambda1$hazard),]
  Lambda1 <- Lambda1[Lambda1$hazard != 0,]
  Lambda2 <- basehaz(m$m.cause2, centered = FALSE)
  Lambda2 <- Lambda2[!duplicated(Lambda2$hazard),]
  Lambda2 <- Lambda2[Lambda2$hazard != 0,]
  lambda1 <- diff(c(0, Lambda1$hazard))
  lambda2 <- diff(c(0, Lambda2$hazard))
  beta1 <- m$m.cause1$coef
  beta2 <- m$m.cause2$coef
  return(c(beta1, beta2, lambda1, lambda2))
}

eta01 <- eta(m01)
eta02 <- eta(m02)
eta11 <- eta(m11)
eta12 <- eta(m12)
eta21 <- eta(m21)
eta22 <- eta(m22)

dEta <- cbind((eta01 - eta02)/eps,
              (eta11 - eta12)/eps,
              (eta21 - eta22)/eps)
```

```
#---------Covariance for log(nu)---------#
set.seed(16072100)
covnu <- solve(-DMdeath$hessian)

#---------Combine to final covariance matrix---------#
final.cov <- rbind(cbind(covnu,covnu %*% t(dEta)),
                   cbind(dEta %*% covnu, covparam + dEta %*% covnu %*% t(dEta)))

#---------Calculate standard error of frailties---------#
nu <- exp(lognu)
nu0 <- nu[1]
nu1 <- nu[2]
nu2 <- nu[3]
dg1 <- c(-nu0/(nu0 + nu1)^2, -nu1/(nu0 + nu1)^2, 0)
dg2 <- c(-nu0/(nu0 + nu2)^2, 0, -nu2/(nu0 + nu2)^2)
sd1 <- sqrt(dg1%*%covnu%*%dg1)
sd2 <- sqrt(dg2%*%covnu%*%dg2)
dg3 <- c(0.5*nu0*((nu0 + nu1)*(nu0 + nu2))^(-3/2)*
           (nu0*nu1+nu0*nu2+2*nu1*nu2),
         0.5*nu0*nu1*((nu0 + nu1)^2*(nu0 + nu2)*
                        sqrt(1/((nu0 + nu1)*(nu0 + nu2))))^(-1),
         0.5*nu0*nu2*((nu0 + nu1)*(nu0 + nu2)^2*
                        sqrt(1/((nu0 + nu1)*(nu0 + nu2))))^(-1))
sd3 <- sqrt(dg3%*%covnu%*%dg3)

# Confidence intervals
p <- param(lognu)
lv1 <- p$var1 - 1.96*sd1
uv1 <- p$var1 + 1.96*sd1
lv2 <- p$var2 - 1.96*sd2
uv2 <- p$var2 + 1.96*sd2
lcor <- p$corr - 1.96*sd3
ucor <- p$corr + 1.96*sd3
```

C.3 R code for tables and figures

```
#---------Competing risks model chapter 3---------#
m.comp <- coxph(Surv(time, status)~age1.1 + age2.1 + size.1 +
                  node.1 + surgery.1 + CTperi.1 + CTadj.1 + RTadj.1 +
                  #transition 2
                  age1.2 + age2.2 + size.2 + node.2 + surgery.2 +
```

```
                  CTperi.2 + CTadj.2 + RTadj.2 +
                  strata(trans),
                data = expandedData)
newData <- function(center = 0, age = 0, size = 0, node = 0, surgery = 0,
                    CTperi = 0, CTadj = 0, RTadj = 0){
  data <- data.frame(trans=1:2,age1.1=0,age1.2=0,age2.1=0,age2.2=0,
                     size.1=0,size.2=0,
                     node.1=0,node.2=0,
                     surgery.1=0,surgery.2=0,
                     CTperi.1=0,CTperi.2=0,
                     CTadj.1=0, CTadj.2=0,
                     RTadj.1=0,RTadj.2=0,
                     center1.1=0,center1.2=0,center2.1=0,center2.2=0,
                     center3.1=0,center3.2=0,center4.1=0,center4.2=0,
                     center5.1=0,center5.2=0,center6.1=0,center6.2=0,
                     center7.1=0,center7.2=0,center8.1=0,center8.2=0,
                     center9.1=0,center9.2=0,center10.1=0,center10.2=0,
                     center11.1=0,center11.2=0,center12.1=0,center12.2=0,
                     center13.1=0,center13.2=0,
                     strata=1:2)
  p <- as.matrix(data)
  if(age==1) diag(p[,2:3]) <- 1
  if(age==2) diag(p[,4:5]) <- 1
  if(size==1) diag(p[,6:7]) <- 1
  if(node==1) diag(p[,8:9]) <- 1
  if(surgery==1) diag(p[,10:11]) <- 1
  if(CTperi==1) diag(p[,12:13]) <- 1
  if(CTadj==1) diag(p[,14:15]) <- 1
  if(RTadj==1) diag(p[,16:17]) <- 1
  if(center==1) diag(p[,18:19]) <- 1
  if(center==2) diag(p[,20:21]) <- 1
  if(center==3) diag(p[,22:23]) <- 1
  if(center==4) diag(p[,24:25]) <- 1
  if(center==5) diag(p[,26:27]) <- 1
  if(center==6) diag(p[,28:29]) <- 1
  if(center==7) diag(p[,30:31]) <- 1
  if(center==8) diag(p[,32:33]) <- 1
  if(center==9) diag(p[,34:35]) <- 1
  if(center==10) diag(p[,36:37]) <- 1
  if(center==11) diag(p[,38:39]) <- 1
  if(center==12) diag(p[,40:41]) <- 1
  if(center==13) diag(p[,42:43]) <- 1
  data <- as.data.frame(p)
```

```
  return(data)
}
p1 <- newData(age=0,size=0,node=0,surgery=0,
              CTperi=0,CTadj=0,RTadj=1)
p2 <- newData(age=2,size=1,node=1,surgery=1,
              CTperi=0,CTadj=0,RTadj=0)
ord <- c(3:1)
par(mfrow=c(1,2))
msfitcox1<-msfit(object=m.comp,newdata=p1,vartype="aalen",trans=tmat)
ptcox1 <- probtrans(msfitcox1,predt = 0, method = "aalen")
plot(ptcox1, main="Cumulative incidence curves",type="filled", ord=ord,
xlab="Years since Surgery")

msfitcox2<-msfit(object=m.comp,newdata=p2,vartype="aalen",trans=tmat )
ptcox2 <- probtrans(msfitcox2,predt = 0, method = "aalen")
plot(ptcox2, main="Cumulative incidence curves",type="filled", ord=ord,
xlab="Years since Surgery")


#---------Nonparam cumulative hazards for each center---------#
CH1 <- survfit(Surv(time, status) ~ center,
               data=longData, subset=(trans==1))
CH2 <- survfit(Surv(time, status) ~ center,
               data=longData, subset=(trans==2))
par(mfrow=c(1,2))
plot(CH1, mark.time=FALSE, xlab="Years since surgery",ylim=c(0,0.75),
     main="Cumulative hazard LR",fun="cumhaz",col=1:14)
plot(CH2, mark.time=FALSE, xlab="Years since surgery",ylim=c(0,0.75),
     main="Cumulative hazard DM/Death",fun="cumhaz",col=1:14)

# combined for all centers
CH1 <- survfit(Surv(time, status) ~ 1,
               data=longData, subset=(trans==1))
CH2 <- survfit(Surv(time, status) ~ 1,
               data=longData, subset=(trans==2))
par(mfrow=c(1,2))
plot(CH1, mark.time=FALSE, xlab="Years since surgery",ylim=c(0,0.55),
     main="Cumulative hazard LR",fun="cumhaz")
plot(CH2, mark.time=FALSE, xlab="Years since surgery",ylim=c(0,0.55),
     main="Cumulative hazard DM/Death",fun="cumhaz")

#---------frailty model chapter 3---------#
m.LRDM <- coxph(Surv(time, status)~age+size+node+
```

```
                  surgery+CTperi+CTadj+RTadj+
                  frailty(center, distribution="gamma"),
                data=longData,subset=(trans==1))
m.surv <- coxph(Surv(time, status)~age+size+node+
                  surgery+CTperi+CTadj+RTadj+
                  frailty(center, distribution="gamma"),
                data=longData,subset=(trans==2))
#common frailty:
m.commonf <- coxph(Surv(time, status)~age1.1+age2.1+size.1+
                   node.1+surgery.1+CTperi.1+CTadj.1+RTadj.1+
                   #transition 2
                   age1.2+age2.2+size.2+node.2+surgery.2+
                   CTperi.2+CTadj.2+RTadj.2+
                   +frailty(center, distribution="gamma")+strata(trans),
                 data=expandedFrailtyData)


#---------Competing risks frailty model confidence intervals chapter 7---------#
m.cause1 <- final$m.cause1
m.cause2 <- final$m.cause2
b <- round(sqrt(diag(final.cov)[4:19]),2)

l1 <- round(exp(m.cause1$coef - 1.96*b[1:8]), 3)
u1 <- round(exp(m.cause1$coef + 1.96*b[1:8]), 3)
l2 <- round(exp(m.cause2$coef - 1.96*b[9:16]), 3)
u2 <- round(exp(m.cause2$coef + 1.96*b[9:16]), 3)
ci1 <- paste(l1, u1, sep="-")
ci2 <- paste(l2, u2, sep="-")

p <- param(lognu)
lv1 <- round(p$var1 - 1.96*sd1, 2)
uv1 <- round(p$var1 + 1.96*sd1, 2)
lv2 <- round(p$var2 - 1.96*sd2, 2)
uv2 <- round(p$var2 + 1.96*sd2, 2)
lcor <- round(p$corr - 1.96*sd3, 2)
ucor <- round(p$corr + 1.96*sd3, 2)
civ1 <- paste(lv1, uv1, sep="-")
civ2 <- paste(lv2, uv2, sep="-")
cicor <- paste(lcor, ucor, sep="-")

#---------Empirical bayes chapter 7---------#
lognu <- c(2.897764, 1.164795, -8.283077)
nu <- exp(lognu)
```

```
nu0<-nu[1]
nu1<-nu[2]
nu2<-nu[3]
Wk1 <- empbayes$Wk1
Wk2 <- empbayes$Wk2
sd1 <- empbayes$sdWk1
sd2 <- empbayes$sdWk2
center <- empbayes$center
K <- length(center)

#CIs for frailties
lower1 <- Wk1 - 1.96*sd1
upper1 <- Wk1 + 1.96*sd1
lower2 <- Wk2 - 1.96*sd2
upper2 <- Wk2 + 1.96*sd2

# sort by patient number
par(mfrow=c(1, 1))
ord <- order(table(data$center))
center.ord <- labels(table(data$center)[ord])[[1]][-1]
index <- match(center.ord,center,)

# cause 1
plot(Wk1[index], ylim=c(0.5, 2), xlim=c(0, 15),
     log="y",
     main="Empirical Bayes estimates for LR",
     xlab="Center index",ylab="Frailty for LR")
for (i in 1:K){
  lines(x=c(i,i), y=c(lower1[index][i],upper1[index][i]))
}
abline(h = 1, col = "red", lty = 2)

# cause 2
plot(Wk2[index], ylim=c(0.5,2), xlim=c(0, 15),
     log="y",
     main="Empirical Bayes estimates for DM/Death",
     xlab="Center index",ylab="Frailty for DM/Death")
for (i in 1:K){
  lines(x=c(i,i), y=c(lower2[index][i],upper2[index][i]))
}
abline(h = 1, col = "red", lty = 2)

# both causes
```

```
plot(Wk1, Wk2, log="xy", ylim=c(0.7,1.6),xlim=c(0.7,1.6),
     main="Empirical Bayes estimates for both causes",
     xlab="Frailty of LR",ylab="Frailty of DM/Death")

#---------create average patient---------#
X <- model.matrix(m.cause1)
x <- apply(X,2,mean)

average.p <- function(hosp=0){
  p.data <- data.frame(trans=1:2,age1.1=0,age1.2=0,age2.1=0,age2.2=0,
                       size.1=0,size.2=0,
                       node.1=0,node.2=0,
                       surgery.1=0,surgery.2=0,
                       CTperi.1=0,CTperi.2=0,
                       CTadj.1=0, CTadj.2=0,
                       RTadj.1=0,RTadj.2=0,
                       wk1.1=0,wk1.2=0,wk2.1=0,wk2.2=0,
                       strata=1:2)
  p <- as.matrix(p.data)
  diag(p[,2:3]) <- x[1]
  diag(p[,4:5]) <- x[2]
  diag(p[,6:7]) <- x[3]
  diag(p[,8:9]) <- x[4]
  diag(p[,10:11]) <- x[5]
  diag(p[,12:13]) <- x[6]
  diag(p[,14:15]) <- x[7]
  diag(p[,16:17]) <- x[8]
  diag(p[,18:19]) <- Wk1[hosp]
  diag(p[,20:21]) <- Wk2[hosp]
  p.data <- as.data.frame(p)
  return(p.data)
}

#---------Stacked cumulative incidences for avarage patient chapter 7---------#
longFrailtyData <- data.frame(final[3:21])
covs=c("center","age","size","node","surgery",
       "CTperi","CTadj","RTadj","lwk1","lwk2")
tmat <- transMat(x=list(c(2,3),c(),c()),
                 names=c("ANED","LR","DM/Death"))
longData$lwk1 <-  log((longFrailtyData$zk0 + longFrailtyData$zk1)/(nu0 + nu1))
longData$lwk2 <-  log((longFrailtyData$zk0 + longFrailtyData$zk2)/(nu0 + nu2))
expandedFrailtyData <-  expand.covs(longData, covs, longnames=FALSE)
```

```
m.final <- coxph(Surv(time, status)~age1.1+age2.1+size.1+
                   node.1+surgery.1+CTperi.1+CTadj.1+RTadj.1+
                   #transition 2
                   age1.2+age2.2+size.2+node.2+surgery.2+
                   CTperi.2+CTadj.2+RTadj.2+
                   offset(lwk1.1+lwk2.2)+strata(trans),
                 data=expandedFrailtyData)

par(mfrow=c(2,2))
p1 <- average.p(which.min(empbayes$Wk1))
p2 <- average.p(which.max(empbayes$Wk1))

ord <- c(3:1)
par(mfrow=c(1,2))
msfitcox1<-msfit(object=m.final,newdata=p1,vartype="aalen",trans=tmat)
msfitcox1$Haz$Haz[(msfitcox1$Haz)$trans==1] <- p1$wk1.1[1]*
  msfitcox1$Haz$Haz[(msfitcox1$Haz)$trans==1]
msfitcox1$Haz$Haz[(msfitcox1$Haz)$trans==2] <- p1$wk2.2[2]*
  msfitcox1$Haz$Haz[(msfitcox1$Haz)$trans==2]
ptcox1 <- probtrans(msfitcox1,predt = 0, method = "aalen")
plot(ptcox1, main="Cumulative incidence curves",type="filled",
     ord=ord, xlab="Years since Surgery")

msfitcox2<-msfit(object=m.final,newdata=p2,vartype="aalen",trans=tmat )
msfitcox2$Haz$Haz[(msfitcox2$Haz)$trans==1] <- p2$wk1.1[1]*
  msfitcox2$Haz$Haz[(msfitcox2$Haz)$trans==1]
msfitcox2$Haz$Haz[(msfitcox2$Haz)$trans==2] <- p2$wk2.2[2]*
  msfitcox2$Haz$Haz[(msfitcox2$Haz)$trans==2]
ptcox2 <- probtrans(msfitcox2,predt = 0, method = "aalen")
plot(ptcox2, main="Cumulative incidence curves",type="filled",
     ord=ord, xlab="Years since Surgery")

#---------Cumulative hazards for average patient in all centers---------#
m.cause1 <- final$m.cause1
m.cause2 <- final$m.cause2
X <- model.matrix(m.cause1)
x <- apply(X,2,mean)
Lambda1 <- basehaz(m.cause1, centered=FALSE)
Lambda1 <- Lambda1[!duplicated(Lambda1$hazard),] # remove duplicate hazard values
Lambda1 <- Lambda1[Lambda1$hazard!=0,]
Lambda2 <- basehaz(m.cause2, centered=FALSE)
Lambda2 <- Lambda2[!duplicated(Lambda2$hazard),]
Lambda2 <- Lambda2[Lambda2$hazard!=0,]
```

```
lambda1 <- diff(c(0,Lambda1$hazard))
lambda2 <- diff(c(0,Lambda2$hazard))
beta1 <- m.cause1$coef
beta2 <- m.cause2$coef
hr1 <- beta1%*%x
hr2 <- beta2%*%x
cumhaz1 <- Lambda1$hazard*hr1
cumhaz2 <- Lambda2$hazard*hr2
par(mfrow=c(1,2))
plot(Lambda1$time,cumhaz1*Wk1[1],col=1,type="l",
     main="Cumulative hazards for LR",
     xlab="Years since surgery",ylab="")
for(i in 2:length(center)){
  lines(Lambda1$time,cumhaz1*Wk1[i],col=i)
}
plot(Lambda2$time,cumhaz2*Wk2[1],col=1,type="l",
     main="Cumulative hazards for DM/Death",
     xlab="Years since surgery",ylab="")
for(i in 2:length(center)){
  lines(Lambda2$time,cumhaz2*Wk2[i],col=i)
}

#---------Cumulative incidences for average patient in all centers---------#
haz1 <- lambda1*hr1
haz2 <- lambda2*hr2
cumhaz1 <- Lambda1$hazard*hr1
cumhaz2 <- Lambda2$hazard*hr2
#create risk sets
l1<-vapply(Lambda2$time, function(x)
            ifelse(any(which.max(which(Lambda1$time<=x))),
            which.max(which(Lambda1$time<=x)),0),numeric(1))
l2<-vapply(Lambda1$time, function(x)
          which.max(which(Lambda2$time<=x)),numeric(1))

par(mfrow=c(1,2))
plot(Lambda1$time,cumsum(haz1*Wk1[1]*
                  exp(-Wk1[1]*cumhaz1-Wk2[1]*cumhaz2[l2])),
     col=1,type="l",
     main="Cumulative incidence for LR",
     xlab="Years since surgery",ylab="")
for(i in 2:length(center)){
  lines(Lambda1$time,cumsum(haz1*Wk1[i]*
                      exp(-Wk1[i]*cumhaz1-Wk2[i]*cumhaz2[l2])),
```

```
                      col=i)
}
plot(Lambda2$time, cumsum(haz2*Wk2[1]*
                          exp(-Wk1[1]*
                                ifelse(l1==0, 0, cumhaz1[l1])-Wk2[1]*cumhaz2)),
     col=1,type="l", main="Cumulative incidence for DM/Death",
     xlab="Years since surgery",ylab="")
for(i in 2:length(center)){
  lines(Lambda2$time,cumsum(haz2*Wk2[i]*
                            exp(-Wk1[i]*
                            ifelse(l1==0,0,cumhaz1[l1])-Wk2[i]*cumhaz2)),
                            col=i)
}
```