

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/86285> holds various files of this Leiden University dissertation.

Author: Rüten-Budde, A.J.

Title: Personalised medicine for multiple outcomes : methods and application

Issue Date: 2020-03-10

Investigating hospital heterogeneity with a competing risks frailty model

This chapter has been published in *Statistics in Medicine* 38(2) (2018) 269–288 as A.J. Rueten-Budde , H. Putter and M. Fiocco, "Investigating hospital heterogeneity with a competing risks frailty model" [17].

Abstract

Survival analysis is used in the medical field to identify the effect of predictive variables on time to a specific event. Generally, not all variation of survival time can be explained by observed covariates. The effect of unobserved variables on the risk of a patient is called frailty. In multicenter studies, the unobserved center effect can induce frailty on its patients, which can lead to selection bias over time when ignored. For this reason, it is common practice in multicenter studies to include a random frailty term modeling center effect. In a more complex event structure, more than one type of event is possible. Independent frailty variables representing center effect can be incorporated in the model for each competing event. However, in the medical context, events representing disease progression are likely related and correlation is missed when assuming frailties to be independent. In this work, an additive gamma frailty model to account for correlation between frailties in a competing risks model is proposed, to model frailties at center level. Correlation indicates a common center effect on both events and measures how closely the risks are related. Estimation of the model using the expectation-maximization algorithm is illustrated. The model is applied to a data set from a multicenter clinical trial on breast cancer from the European Organisation for Research and Treatment of Cancer (EORTC trial 10854). Hospitals are compared by employing empirical Bayes estimates methodology together with corresponding confidence intervals.

§2.1 Introduction

Survival data arises where interest lies on the time from a specific time origin until occurrence of an event of interest. Prominent applications are found in the medical field, where e. g. time from diagnosis of disease until death could be studied. What distinguishes survival analysis from other types of statistical analysis is the type of data it deals with: it is generally incomplete. Since it takes time to observe an event, it is usually not possible to collect complete information. A popular method to model the effect of covariates on risk of event occurrence is through the semi-parametric Cox proportional hazards model [44]. In some situations more than one type of end point are possible, when e. g. different causes of death are studied. Analogous to the single endpoint situation the Cox model can be used to model the effect of covariates on the cause-specific transition hazards [119] of each cause of failure. A more complicated event structure with intermediate states can be modeled by a multi-state model [119]. Dependence in survival data can be modeled by a random frailty term, which models heterogeneity between observations or between clusters of observations. The frailty term represents unobserved covariates on the individual or cluster level that act on the risk of event occurrence. The frailty variance can be interpreted as a measure of heterogeneity between clusters or individuals, however it can also be seen as a measure of dependence within a cluster.

Multicenter studies are a common strategy to collect sufficient data for a clinical study. Patients are clustered within treatment centers and possible correlation between patients within a center can be modeled by using a shared frailty model. Shared frailty models are able to model dependence, however these models limit the unobserved covariates modeled by the frailty to have the same effect within a cluster. In the presence of competing events the use of one frailty per center acting on all causes of failure is questionable. Similarly using J independent frailties per center one for each cause of failure does not yield a complete picture of the data structure. Frailties for different competing events within a center are likely to be correlated, since they represent the same unobserved covariates on cluster level. Yashin et al. [166] first introduced a correlated gamma frailty model to analyze twin survival data. They decompose a twin's frailty into a sum of two independent frailties, one of which is shared by both twins. Petersen et al. [116] use this idea of adding frailty components, which act multiplicatively on the individual hazard and describe more complex variance components models for survival data.

Clustered data in the presence of competing risks further complicate possible dependence structures and different approaches are taken. Extensions of Fine and Gray's subdistribution hazard model [57] incorporate a frailty term to model cluster dependence on the cumulative incidence function of the event of interest in the presence of competing events [89, 131, 51, 172]. Wienke et al. [161, 163] analyze correlated frailty models in the presence of competing risks, however assuming independence between risks. The assumption of independence is questionable since related events (e. g. events representing disease progression) might be influenced similarly by the same unobserved covariates. Wienke et al. [162] extend the bivariate correlated gamma frailty model of Yashin et al. [166] to model dependence among competing

risks based on parametric marginal survival functions. Gorfine and Hsu [72] combine frailty components multiplicatively to model dependence between competing risks for clustered survival data. Liquet et al. [99] analyze hospital heterogeneity in multi-state models using independent and joint frailty models to model dependence between transition intensities. Rotolo et al. [123] propose to incorporate correlated frailties in multi-state models acting on the transition-specific hazard functions. They construct frailties by combining a common cluster component and a transition-specific component multiplicatively.

In this paper we propose an additive gamma frailty model which acts multiplicatively on the cause-specific hazard to model dependence within clusters and between two competing events. The method can be used to investigate hospital heterogeneity in a competing risks setting. An elegant estimation procedure using the EM-algorithm is outlined as well as a strategy to calculate the standard error of the estimates. In contrast to Wienke et al. [162] who model dependence among competing risks by using a parametric approach our method is based on the semi-parametric Cox model [44]. Compared to methods suggested by Gorfine and Hsu [72] and Rotolo et al. [123] which combine frailty components multiplicatively in this article a gamma decomposition is proposed to model dependence between risks. The advantage of our method is its simplicity in construction and estimation, which is based on the mathematical properties of the gamma distribution. Additionally, estimation through the EM-algorithm provides empirical Bayes estimates for each center's frailty, which can be used to compare centers.

In Section 2.2 and 2.3 the cause-specific hazards model and frailty model will be reviewed briefly. The proposed competing risks frailty model is presented in Section 2.4. In Section 2.5 the method is applied to a data example and corresponding results are presented. A simulation study to investigate the performance of the correlated frailty model is discussed in Section 2.6. A discussion follows in Section 2.7.

§2.2 Competing risks model

Competing risks models are used when more than one type of failure is possible. An example is the study of different causes of death. A fundamental concept in competing risks is the cause-specific hazard. It is the hazard of failing from a particular cause given still event free at that time.

For right censored survival times the cause-specific hazard of cause j for a subject i with covariate vector \mathbf{X}_i is as follows:

$$\lambda_j(t|\mathbf{X}_i) = \lambda_{j0}(t)e^{\beta_j^T \mathbf{X}_i}, \quad (2.2.1)$$

where λ_{j0} is the cause-specific baseline hazard for cause j and β_j assesses the effect of the covariates \mathbf{X}_i on the progression rate to cause j [119]. Here the effects of covariates are quantified on the cause-specific hazard and not on the marginal hazard. Only if the censoring due to the competing risks is non-informative conditionally on the covariates in the model, the estimates can also be interpreted as effects on the marginal hazard.

§2.3 Frailty model

The concept of frailty introduces random effects in survival models, which represent the presence of unobserved heterogeneity. The variance of this random component is a measure used to quantify heterogeneity in the data. Vaupel et al. [157] discussed univariate frailty models with a gamma distribution and applied this concept to survival. Clayton [40] used frailties in the multivariate analysis of chronic disease incidence in families.

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 or a group or cluster of individuals. In univariate frailty models each individual has its own independent frailty, while in shared frailty models clustered individuals share a common frailty.

For subject i with covariate vector \mathbf{X}_i belonging to cluster k with frailty W_k the hazard is given as

$$\begin{aligned}\lambda(t|\mathbf{X}_i, W_k) &= W_k \lambda_0(t) e^{\beta^T \mathbf{X}_i} \\ &= \lambda_0(t) e^{\beta^T \mathbf{X}_i + \log(W_k)}.\end{aligned}\tag{2.3.1}$$

A convenient choice for the frailty distribution is the gamma distribution, since its posterior distribution given survival data, stays in the gamma family [116].

§2.4 Competing risks frailty model

Heterogeneity between centers in a competing risks setting can be modeled by assigning each center J frailties, one for each cause of failure. The J frailty terms within a center may be chosen to be independent, however the effects within a center are likely to be related which is ignored in such a model. In a more realistic model frailties within a center are correlated. A model for the dependence structure was first proposed by Yashin et al. [166] in a twin study, decomposing the frailty of each twin as a sum of two independent frailties one of which is shared. Petersen et al. [116] use an additive variance components structure on multiplicative gamma frailty models and outline its estimation. The correlated frailty model proposed in this article follows their approach.

§2.4.1 Frailty decomposition

In the following, let W_{k1}, W_{k2} denote the frailty variables corresponding to two causes of failure within hospital k ($k = 1, \dots, K$). Correlation between frailties is constructed by decomposing each frailty as the sum of two independent gamma distributed variables, one of which is common in both frailties [59, 58]. For cause j ($j = 1, 2$), frailties are given as

$$W_{kj} = \frac{Z_{k0} + Z_{kj}}{\nu_0 + \nu_j},\tag{2.4.1}$$

where

$$Z_{k0} \sim \Gamma(\nu_0, 1), \quad Z_{kj} \sim \Gamma(\nu_j, 1). \quad (2.4.2)$$

The random variables Z_{k0} , Z_{k1} , Z_{k2} are independent and from now on referred to as the independent frailty components of hospital k . This results in the following frailty distribution:

$$W_{kj} \sim \Gamma(\nu_0 + \nu_j, \nu_0 + \nu_j). \quad (2.4.3)$$

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

$$\text{Var}(W_{k1}) = \frac{1}{\nu_0 + \nu_1} = \xi_1, \quad \text{Var}(W_{k2}) = \frac{1}{\nu_0 + \nu_2} = \xi_2, \quad (2.4.4)$$

$$\text{Cor}(W_{k1}, W_{k2}) = \nu_0(\xi_1\xi_2)^{1/2} = \rho. \quad (2.4.5)$$

This construction allows for positive correlation only. In many practical situations however it may be justified to disregard negative correlation, e. g. when competing events describe disease progression. A further restriction is that not all variance correlation combinations are possible in this construction. A large correlation does not allow the variances to be too different, or equivalent, different frailty variances do not allow the correlation to be (almost) one:

$$\nu_1 = \frac{1}{\xi_1} - \frac{\rho}{\sqrt{\xi_1\xi_2}} > 0$$

$$\sqrt{\xi_2/\xi_1} > \rho \quad (2.4.6)$$

$$\nu_2 = \frac{1}{\xi_2} - \frac{\rho}{\sqrt{\xi_1\xi_2}} > 0$$

$$\sqrt{\xi_1/\xi_2} > \rho. \quad (2.4.7)$$

From (2.4.6) and (2.4.7) it follows that $\rho < \min(\sqrt{\xi_2/\xi_1}, \sqrt{\xi_1/\xi_2})$.

§2.4.2 Model estimation

Model parameters are obtained by maximizing the log-likelihood function based on the observed data. Since frailties associated to different centers and individuals across hospitals are independent, the likelihood is the product of hospital likelihoods. For simplicity only the log-likelihood and necessary quantities of a single center k are given in the following.

Denote by n_k and d_{kj} the number of patients and the number of patients that fail from cause j ($j = 1, 2$) in hospital k respectively. Let \mathbf{X}_{ki} , t_{ki} and δ_{ki} ($\delta_{ki} = 0, 1, 2$) be

the covariate vector for patient i treated at hospital k , the event or censoring time and the event or censoring indicator respectively. In the following let β_j be the vector of regression coefficients, λ_{j0} the baseline hazard and Λ_{j0} the cumulative baseline hazard for cause j ($j = 1, 2$). If the frailties were observed the complete data yields the following log-likelihood for hospital k

$$\begin{aligned} \ell_k(\beta_1, \beta_2, \lambda_{10}, \lambda_{20}) = & \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=1} \left\{ \log \left(\frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu_1} \right) + \log(\lambda_{10}(t_{ki})e^{\beta_1^T \mathbf{X}_{ki}}) \right\} \\ & + \sum_{i=1}^{n_k} \mathbf{1}_{\delta_{ki}=2} \left\{ \log \left(\frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu_2} \right) + \log(\lambda_{20}(t_{ki})e^{\beta_2^T \mathbf{X}_{ki}}) \right\} \\ & - \frac{Z_{k0} + Z_{k1}}{\nu_0 + \nu_1} \sum_{i=1}^{n_k} \Lambda_{10}(t_{ki})e^{\beta_1^T \mathbf{X}_{ki}} - \frac{Z_{k0} + Z_{k2}}{\nu_0 + \nu_2} \sum_{i=1}^{n_k} \Lambda_{20}(t_{ki})e^{\beta_2^T \mathbf{X}_{ki}} \\ & + \log(f(Z_{k0}, Z_{k1}, Z_{k2})), \end{aligned} \quad (2.4.8)$$

where f is the probability density function of the independent and gamma distributed frailty components.

Integrating out all frailty components specific to each center in the log-likelihood yields the observed data log-likelihood, which is computationally challenging to maximize (see Appendix 2.A for details). Considering the unobserved frailties as missing information yields a typical application of the expectation maximization algorithm (EM-algorithm) [50].

§2.4.3 Implementation

For fixed parameter $\nu = (\nu_0, \nu_1, \nu_2)$, the estimation procedure uses the expectation maximization algorithm (EM-algorithm) to approximate the observed data log-likelihood to find optimal regression coefficients and baseline hazards [50]. The approximated observed data log-likelihood is then employed in a three dimensional search to a find maximum likelihood estimate for ν .

Since ν is fixed throughout the EM iterations, the estimation concerns the regression coefficients and baseline hazards only. The conditional expectations of the terms $\log((Z_{k0} + Z_{kj})/(\nu_0 + \nu_j))$, ($j = 1, 2$) and of $\log(f(Z_{k0}, Z_{k1}, Z_{k2}))$ given observed data are irrelevant to the estimation of the complete data case (2.4.8). Therefore 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. As a result, defining $\Lambda_{kj} = \sum_{i=1}^{n_k} \Lambda_{j0}(t_{ki})e^{\beta_j^T \mathbf{X}_{ki}}$, ($j = 1, 2$), it is sufficient to consider

$$E(Z_{k0}|\text{data}_k) = \int_{z_{k0}} z_{k0} f(z_{k0}|\text{data}_k) dz_{k0} \quad (2.4.9)$$

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

$$E(Z_{k1}|\text{data}_k) = \int_{z_{k1}} z_{k1} f(z_{k1}|\text{data}_k) dz_{k1} \quad (2.4.10)$$

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

$$E(Z_{k2}|\text{data}_k) = \int_{z_{k2}} z_{k2} f(z_{k2}|\text{data}_k) dz_{k2} \quad (2.4.11)$$

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

where f is the conditional probability density function of a frailty component given data, and $c_k(l, m, \nu_0, \nu_1, \nu_2)$ is a function over the number of events of each type of failure for fixed frailty parameters. Details about the computations are outlined in the Appendix 2.A.

Since the conditional distributions of the frailty components Z_{k0}, Z_{k1}, Z_{k2} given observed data are mixtures of gamma distributions (see Appendix 2.A for details), it is straightforward to compute the quantities (2.4.9)-(2.4.11). Notably the factor $c_k(l, m, \nu_0, \nu_1, \nu_2)$ is the same in all three expectations.

The M-step consists of estimating the updated baseline hazards $\Lambda_{10}(t)$, $\Lambda_{20}(t)$ and coefficient vectors β_1 , β_2 , through maximization of the conditional log-likelihood, given frailties estimated in the E-step. This can be done with existing software, e.g. using `coxph()` from the R [122] package `survival` [142], incorporating the logarithm of the expected frailties as `offset` into the cause-specific hazards model. The algorithm iterates over these two steps and stops once the approximation of the observed data log-likelihood converged (e.g. change of smaller than 1e-06).

Until now, the frailty parameter $\nu = (\nu_0, \nu_1, \nu_2)$ was fixed throughout the EM iterations. Profile likelihood is used to obtain maximum likelihood estimates of $(\nu_0, \nu_1, \nu_2, \beta_1, \beta_2, \Lambda_{10}, \Lambda_{20})$; the function `optim()` is used to find the optimal ν , maximizing the observed data log-likelihood approximated with the EM-algorithm (see the supplementary material in this paper).

§2.4.4 Estimation of the standard error

Louis [101] discussed how to obtain the covariance matrix for the regression parameters, that stays within the EM-algorithm framework, using only derivatives of the complete data log-likelihood. This approach does not yet include the uncertainty caused by estimating the frailty parameters $\nu = (\nu_0, \nu_1, \nu_2)$ outside of the EM-algorithm.

Putter and van Houwelingen [120, supplementary material] proposed estimation as described in the following.

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 denote by $\hat{\boldsymbol{\nu}}$ the MLE of $\boldsymbol{\nu}$ maximizing the observed data log-likelihood. The combined covariance matrix of $\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\eta}}$ is given as

$$\begin{pmatrix} \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} & \Sigma_{\boldsymbol{\nu}\boldsymbol{\eta}} \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right)^T \\ \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right) \Sigma_{\boldsymbol{\nu}\boldsymbol{\eta}} & \Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}} + \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right) \Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}} \left(\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \right)^T \end{pmatrix}, \quad (2.4.12)$$

where $\Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}}$ and $\Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}$ are the covariance matrix of $\boldsymbol{\nu}$ and $\hat{\boldsymbol{\eta}}$ respectively and the term $\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ are the partial derivatives of the regression parameters given $\boldsymbol{\nu}$. The term on the bottom right of (2.4.12) represents the covariance of $\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})$ where the term $\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}})$ is obtained using a Taylor expansion of $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu})$ and the score functions of $\hat{\boldsymbol{\eta}}(\boldsymbol{\nu})$ and $\hat{\boldsymbol{\nu}}$ around the MLEs. The off diagonal terms are covariance matrices of $(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\nu}}))$ and can be derived in a similarly way, see Appendix 2.B for details.

The term $\Sigma_{\boldsymbol{\nu}\boldsymbol{\eta}}$ is computed from the Hessian matrix obtained using the `hessian()` function from the `numDeriv` package [69] around the point estimate of $\boldsymbol{\nu}$ found by the `optim()` function in R [122]. 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 $(\boldsymbol{\nu}, \hat{\boldsymbol{\eta}}(\boldsymbol{\nu}))$ [167, sec. 8.6.2].

The term $\frac{\partial \hat{\boldsymbol{\eta}}(\boldsymbol{\nu})}{\partial \boldsymbol{\nu}}$ is approximated numerically. The derivative around the MLE is estimated by calculating the slope between parameters for values of $\boldsymbol{\nu}$ close to the MLE.

The term $\Sigma_{\boldsymbol{\eta}\boldsymbol{\eta}}$ can be computed as described by Louis [101]. 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, see Appendix 2.B for details.

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 (2.4.12). To obtain the standard error of the frailty variances and correlation we apply the multivariate delta method on $\Sigma_{\boldsymbol{\nu}\boldsymbol{\nu}}$ [39, sec. 5.6]. See the supplementary material for implementation in R.

§2.4.5 Empirical Bayes estimates

Heterogeneity between hospitals may raise the question of hospital ranking based on their frailty or relative performance. A popular method to compare institutions is the empirical Bayes approach introduced to this setting by Thomas et al. [143]. If many centers are involved, a crude center effect estimate may explode for small centers due to large variation and not due to a real center effect [153]. The empirical Bayes estimator helps distinguish observations that are "extreme by nature" and those that are "extreme by chance" and is very well suited for the analysis of quality comparison

data [152]. The empirical Bayes approach not only uses information on a particular center to quantify its performance, but uses information on all centers to help improve the estimate.

Following van Houwelingen [152], the empirical Bayes principle will be outlined. Let X_1, \dots, X_K be independent outcomes with densities $f(x_k, \theta_k)$ and $\theta_1, \dots, \theta_K$ iid with distribution G . The optimal estimator under mean squared error loss for each θ_k is given by the Bayes estimator $d(x_k|G) = E(\theta_k|x_k, G)$, when G is known. When G is unknown, one can estimate $E(\theta_k|x_k, G)$ through an estimate of the distribution G . The resulting estimator is shrunk towards the mean, where the amount of shrinkage depends on the variance of the underlying distribution. In the context of center performance X_k represents the outcome and θ_k the true unobserved performance of center k .

The E-step of the EM-algorithm estimates the empirical Bayes estimate of the center frailties given current model parameters and ν . Hence computing a last E-step based on the MLE of regression parameters and ν after convergence of the algorithm will give the empirical Bayes estimate of center frailties.

Even though empirical Bayes estimates are preferred to crude performance estimates when analyzing quality comparison data, interpretation of results should be made with caution as reasons for different outcome may lie outside a center's responsibility. Statistical issues in comparing institutions are discussed in more detail in Goldstein and Spiegelhalter [71].

The conditional distribution of Z_{k0}, Z_{k1} and Z_{k2} given data is the weighted sum of gamma distributions depending on the number of events of each type (see Appendix 2.A for details). To obtain prediction intervals for the empirical Bayes estimates a simplified sampling procedure is applied.

- 1) Sample from set of tuples (l, m) from $\{(0, 0), \dots, (d_{k1}, d_{k2})\}$, where d_{k1} and d_{k2} are the number of events of type 1 and 2 respectively.
- 2) Sample Z_{k0}, Z_{k1} and Z_{k2} from gamma distributions $\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m, 1 + \Lambda_{k1} + \Lambda_{k2})$, $\Gamma(l + \nu_1, 1 + \Lambda_{k1})$ and $\Gamma(m + \nu_2, 1 + \Lambda_{k2})$ respectively.

Repeating this sampling procedure many times lower and upper confidence limits can be found by taking the 2.5% and 97.5% quantile.

§2.5 Data application

§2.5.1 Data description

The data used in this work originates from the European Organization for Research and Treatment of Cancer (EORTC) trial 10854, which studied the effect of one course of perioperative chemotherapy given directly after surgery on survival [147]. The data set includes 2795 women treated for invasive stage I or II breast cancer, randomized for treatment in 15 different centers. Breast cancer is one of the most common types of cancer in women. The standard treatment for breast cancer is surgery, which may

be followed by chemotherapy, radiotherapy or both. 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.

Patients were excluded from this analysis following exclusion criteria of the trial ($n = 41$) and if information on relevant covariates was missing ($n = 91$). Furthermore, all 5 patients from a particular center were excluded, because of the small amount of patients treated at this center, leaving a total of 2658 patients from 14 different centers for analysis.

The competing risks model for this data is illustrated in Figure 2.1. Two competing events are considered, recurrence of disease (LR or DM) and death. The starting state is the state a patient enters after surgery, being alive with no evidence of disease after surgical removal of the primary tumor (ANED).

Table 2.1: Characteristics of 2658 patients.

Variable	N	(%)
Age		
≥ 50	1602	(60.3)
40–50	762	(28.7)
< 40	294	(11.1)
Tumor size		
$< 2\text{cm}$	798	(30.0)
$\geq 2\text{cm}$	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)

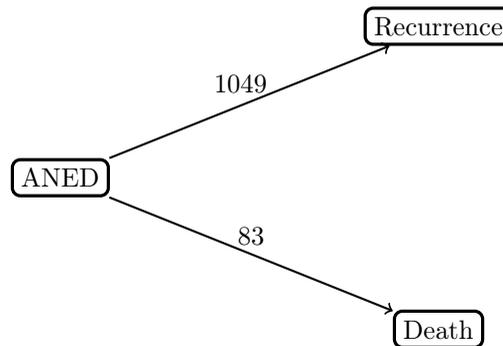


Figure 2.1: Initially, 2658 patients are alive with no evidence of disease (ANED).

The choice of covariates to analyze is based on a previous study on the same data [46]. The following prognostic factors are considered in the analysis: age (≥ 50 , $40 - 50$, < 40), tumor size ($< 2\text{cm}$, $\geq 2\text{cm}$), nodal status (negative, positive), type of surgery (mastectomy, breast conserving), perioperative chemotherapy (yes, no), adjuvant chemotherapy (yes, no), adjuvant radiotherapy (yes, no). Patients' characteristics are provided in Table 2.1.

§2.5.2 Competing risks model with independent frailties

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 model can be estimated similarly to the classical competing risks model, by using `coxph()` together with the `frailty()` function from the R package `survival` [142] or the `emfrail()` function from the `frailtyEM` [27] package. The results of the estimated model with independent gamma frailties are shown in Table 2.2.

A young age (< 40) significantly increases the risks of experiencing recurrence (HR: 1.43; CI: 1.16-1.76), as well as a larger tumor size (HR: 1.41; CI: 1.22-1.64), a positive nodal status (HR: 1.55; CI: 1.34-1.79) and whether or not perioperative chemotherapy and adjuvant chemotherapy was administered (HR: 1.15; CI: 1.02-1.30 and HR: 0.79; CI: 0.64-0.97 respectively). The frailty variance for transition 1 is estimated to be equal to 0.05.

A larger tumor size and a positive nodal status also have a significant effect on death before recurrence with HR: 1.46 (CI: 1.21-1.76) and HR: 2.22 (CI: 1.87-2.63). For death also type of surgery has a significant effect with HR equal to 0.82 (CI: 0.70-0.97) for breast conserving therapy compared to mastectomy. This finding is unexpected and should probably be ascribed to insufficient adjustment for factors relates to choice of primary surgical treatment. The frailty variance for this transition is estimated to be equal to 0.13.

A different frailty model assigns to each hospital a shared frailty term for both causes of failure. Both the independent and shared frailty model are not realistic. 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

Table 2.2: Cause-specific hazards model with independent frailties.

	ANED \rightarrow Recurrence		ANED \rightarrow Death	
	HR	0.95 CI	HR	0.95 CI
Age				
≥ 50	1.00		1.00	
40–50	1.00	0.85-1.19	0.84	0.68-1.04
< 40	1.43	1.16-1.76	1.03	0.79-1.34
Tumor size (≥ 2 vs < 2 cm)	1.41	1.22-1.64	1.46	1.21-1.76
NodST (pos. vs neg.)	1.55	1.34-1.79	2.22	1.87-2.63
Surgery (cons. vs mast.)	0.92	0.80-1.05	0.82	0.70-0.97
PeriCT (no vs yes)	1.15	1.02-1.30	1.11	0.96-1.29
AdjCT (yes vs no)	0.79	0.64-0.97	0.82	0.64-1.05
AdjRT (yes vs no)	1.20	0.73-1.98	1.12	0.62-2.00
	Variance	SE	Variance	SE
Frailty	0.05	0.03	0.13	0.06

Abbreviations: NodST (pos. vs neg.), Nodal status (positive vs negative); Surgery (cons. vs mast.), Surgery (breast conserving vs mastectomy); PeriCT, Perioperative chemotherapy; AdjCT, Adjuvant chemotherapy; AdjRT, Adjuvant radiotherapy; ANED, alive with no evidence of disease; CI, confidence interval; HR, hazard ratio; SE, standard error.

for possible correlation between frailties is probably a more accurate representation of reality.

§2.5.3 Competing risks model with correlated frailties

In Table 2.3 the results for the competing risks frailty model with correlated frailties are shown.

The hazard ratios for recurrence are almost unchanged compared to the independent frailty model. However, in the correlated frailty model nodal status and size are the only significant factors. The hazard ratios for death without recurrence are very different from the independent frailty model. This can be explained by the small number of deaths without recurrence in the data set. The variation added by additionally estimating the frailties, increased the standard errors and fewer variables are significant.

The variance of the frailty for transition 1 (ANED \rightarrow Recurrence) is equal to 0.05 with a standard error of 0.03. For transition 2 (ANED \rightarrow Death) the frailty variance is equal to 0.27 with a standard error of 0.22. The correlation of the frailties is estimated to be equal to 0.37 with a standard error of 0.18. Given these frailty variances the maximum correlation between frailties in this model is 0.43 resulting from inequalities (2.4.6) and (2.4.7).

Table 2.3: Cause-specific hazards model with correlated frailties.

	ANED → Recurrence		ANED → Death	
	HR	0.95 CI	HR	0.95 CI
Age				
≥50	1.00		1.00	
40–50	1.00	0.69-1.44	0.35	0.05-2.76
<40	1.42	0.92-2.18	0.62	0.06-6.48
Tumor size (≥ 2 vs <2 cm)	1.41	1.05-1.89	0.96	0.25-3.73
NodST (pos. vs neg.)	1.55	1.15-2.08	1.72	0.47-6.27
Surgery (cons. vs mast.)	0.92	0.70-1.22	0.65	0.18-2.31
PeriCT (no vs yes)	1.15	0.89-1.48	1.14	0.35-3.70
AdjCT (yes vs no)	0.79	0.50-1.27	0.80	0.06-10.08
AdjRT (yes vs no)	1.18	0.81-1.71	0.66	0.12-3.72
	Variance	SE	Variance	SE
Frailty	0.05	0.03	0.27	0.22
	Correlation		SE	
Correlation	0.37		0.18	

Abbreviations: NodST (pos. vs neg.), Nodal status (positive vs negative); Surgery (cons. vs mast.), Surgery (breast conserving vs mastectomy); PeriCT, Perioperative chemotherapy; AdjCT, Adjuvant chemotherapy; AdjRT, Adjuvant radiotherapy; ANED, alive with no evidence of disease; CI, confidence interval; HR, hazard ratio; SE, standard error.

§2.5.4 Empirical Bayes estimates

Figure 2.2 shows the empirical Bayes estimates of the frailties of each center together with 95% prediction intervals, for event recurrence and death. A value equal to 1 implies that there is no center effect. Centers are ordered by number of patients treated. The prediction intervals are computed by sampling from the gamma mixture distribution of frailties and taking 2.5% and 97.5% quantiles as lower and upper limit.

The left panel of Figure 2.2 shows the frailties for the event recurrence for 14 hospitals ordered by number of patients treated. Two hospitals (9 and 11) have a significantly increased risk for their patients to develop recurrence. One hospital (12) has a significantly decreased risk for its patients to develop recurrence. Further one can see that the width of the prediction intervals decrease with a growing number of patients in the hospital.

The right panel of Figure 2.2 shows that one hospital (11) has an increased risk for its patients to move to the state death. One hospital (14) has a marginally significant decreased risk for its patients to die.

To visualize the relation of the frailties within a hospital the empirical Bayes estimates of the two frailties for each center are plotted against each other in Figure 2.3, together with the joint empirical distribution of the frailties for two centers with index 11 and 12.

The hospital effects on a patient can be investigated by looking at the difference in cumulative hazard and cumulative incidence between the hospitals for a particular

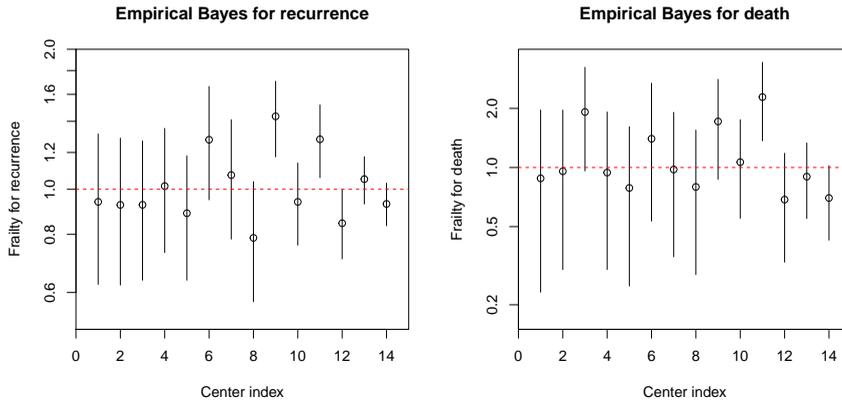


Figure 2.2: Empirical Bayes estimates of frailties and 95% prediction intervals for event recurrence and death of 14 centers, sorted by number of patients.

patient. This is shown in Figure 2.4, for a patient whose covariate values correspond to the mean covariate values in the data.

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 2.5, which shows the stacked cumulative incidence curves for an average patient treated in the two hospitals with the lowest and highest frailty for recurrence. The prognosis shown in the left panel estimates a lower risk for both events, compared to the right panel. This is explained by the estimated correlation between frailties (Table 2.3) and the empirical Bayes estimates of the hospitals (Figure 2.3), which indicate that a hospital with a decreased risk for one cause also has a decreased risk for the other cause. This makes the hospital corresponding to the left panel more appealing.

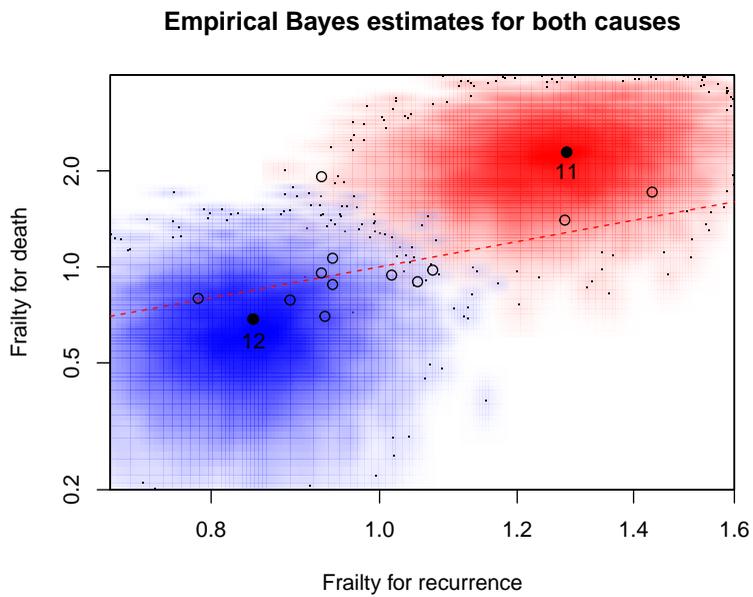


Figure 2.3: Empirical Bayes estimates of frailties for two causes of failure plotted together for 14 centers. For centers with index 11 and 12 the joint empirical distribution of the frailties is shown in red and blue respectively.

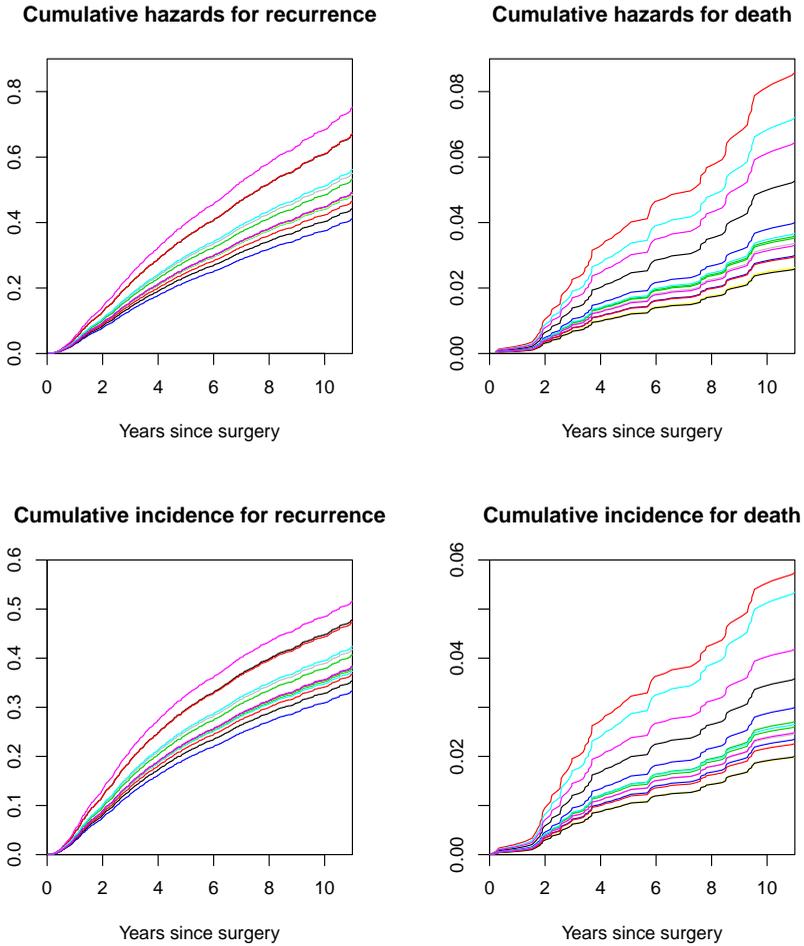


Figure 2.4: Upper panels: cumulative hazards for an average patient for recurrence (on the left) and death (on the right); each line represents a hospital. Lower panels: cumulative incidence of an average patient for recurrence (on the left) and death (on the right); each line represents a hospital.

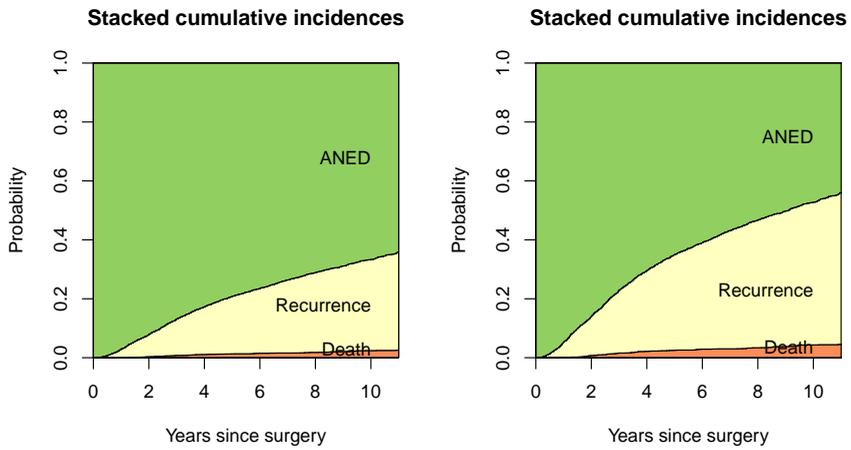


Figure 2.5: Left panel: stacked cumulative incidence curves for an average patient treated in hospital with lowest estimated frailty for recurrence. Right panel: cumulative incidence curves for an average patient treated in hospital with highest estimated frailty for recurrence.

§2.6 Simulation

To investigate the performance of the correlated frailty model a simulation study is conducted. Multiple data scenarios are simulated and the results of the independent and correlated frailty model are compared. Motivated by the data example from Section 2.5 a similar scenario with 2700 patients distributed equally over 15 centers is used for simulation. To study how the number of centers affects the estimation, different scenarios with 5, 30 and 50 centers are considered, while keeping the total number of patients fixed to 2700 (see Table 2.4).

Table 2.4: Scenarios for simulation.

Scenario	n	K	n_k	$\text{Var}(W_{k1})$	$\text{Var}(W_{k2})$	$\text{Cor}(W_{k1}, W_{k2})$	Correlation Bounds
A	2700	5	540	0.25	0.25	0.3	(0, 1)
B	2700	15	180	0.25	0.25	0.3	(0, 1)
C	2700	30	90	0.25	0.25	0.3	(0, 1)
D	2700	50	54	0.25	0.25	0.3	(0, 1)
E	2700	15	180	0.1	0.3	0.8	(0, 0.58)
F	2700	15	180	0.25	0.25	-0.3	(0, 1)

Notation: n , total number of patients; K , number of centers; n_k , number of patients per center; W_{kj} ($j = 1, 2$), center-specific frailty for cause j .

Survival times are generated by using two Weibull baseline hazards with a common shape parameter a and rate parameters b_1 and b_2 for the two causes of failure respectively. Weibull parameters are fixed throughout the data scenarios and are estimated from the data example of Section 2.5 ($a = 1.01, b_1 = 0.05, b_2 = 0.03$).

Different frailty variance structures are simulated in the different scenarios. Using an additive gamma model as presented in Section 2.4 correlated frailties are sampled with variances equal to 0.25 and correlation equal to 0.3 for scenarios A, B, C and D. As discussed in Section 2.4 different frailty variances by construction do not allow too large correlations; in addition correlation is assumed to be positive to use the proposed method. To study the performance of the method proposed in this article data scenarios E and F which violate these assumptions are simulated. Center and patient distribution are set closest to the data example (15 centers with 180 patients each). Frailties for scenarios E and F in Table 2.4 are sampled from a multivariate lognormal distribution. Scenario E considers a situation in which the correlation is too large to be modeled: frailty variances are equal to 0.1 and 0.3 for cause 1 and 2 respectively while correlation is equal to 0.8. Scenario F represents a situation in which negative correlation is present, with frailty variances equal to 0.25 and correlation equal to -0.3.

Table 2.4 summarizes all scenarios simulated. Censoring times are simulated from a uniform distribution between 9 and 14 years, motivated by the data example.

For each scenario, 1000 data sets are simulated for which two models are estimated: a model with independent frailties for the two causes and the proposed correlated frailty model. Results for frailty variance and empirical Bayes estimates are shown in Table 2.5 and Table 2.6, respectively.

Table 2.5 shows that the independent frailty model generally estimates the frailty

variances to be too high with a large bias and large root-mean-square error (RSME). This seems to be more apparent in data sets with fewer centers.

The correlated frailty model estimates on average results that are closer to the true parameter values with bias of less than half the empirical standard error apart from scenarios E and F. Empirical standard errors are smaller compared to the independent model and are comparable to the average standard error, which even though close, is consistently smaller than the empirical standard error. Root-mean-square errors are generally smaller for the correlated frailty model compared to the independent model.

For scenarios with a larger number of centers better estimation results are obtained. Scenario D with 50 centers per data set shows the best estimation results. Average standard errors are close to the empirical standard errors and RSMEs are small. Scenario E showcases a situation in which the correlation is too large to be modeled with the additive gamma construction. Given frailty variances correlation is restricted to $\rho < \sqrt{0.1/0.3} = 0.58$ (see equations (2.4.6-2.4.7)). The method in this case finds a middle ground and underestimates the frailty variance for cause 2 to allow for a larger correlation. Scenario F considers negative correlation. In this case frailty variances are underestimated, however they are still closer to the true values compared to estimates of the independent model and the correlation estimate is very close to 0.

Table 2.6 shows summary measures of empirical Bayes estimates over the different data scenarios. Bias as well as RMSEs are reported together with coverage probabilities of prediction intervals acquired using the sampling method described in Section 2.4 and studied for each scenario. The number of centers has a stronger effect on the empirical Bayes estimates compared to the frailty variance estimates. Scenario A with only 5 centers shows very poor coverage of the 95% prediction intervals with probabilities of 0.394 and 0.492 for empirical Bayes estimates corresponding to cause 1 and 2 respectively. Scenarios with 15 centers (B, E and F) achieved coverage probabilities between 0.749 and 0.864 and scenarios with more centers (C and D) achieved values between 0.877 and 0.930. Bias and RSME of empirical Bayes estimates appear consistent over different scenarios. To quantify the performance of the method on the estimation of the center-specific cumulative incidence, its bias and RMSE are estimated at quartiles of the theoretical overall event time distribution ($t_1 = 3.55$, $t_2 = 8.48$, $t_3 = 16.85$). The estimates appear unbiased but worsen for the later time t_3 . Interestingly, the bias and RSMEs appear not to be influenced much by the amount of centers and it even appears to become slightly worse if more centers are present in the data. An explanation could be that the estimation of the cumulative incidence becomes more challenging due to the data being generated from many different hazard rates.

For some of the simulated data sets the standard error of the frailty variance and correlation estimate could not be obtained because the hessian matrix obtained during optimization was not positive definite. In this case another attempt was made by starting the optimization of the frailty components from another starting value. This procedure was able to compute results in some cases (see Table 2.7). In case the hessian was not positive definite the data set was discarded. The amount of failed estimation was strongly dependent on the amount of centers in the data set.

2. Investigating hospital heterogeneity with a competing risks frailty model

Percentages of second attempts and discarded data sets are given illustrated in Table 2.7.

Table 2.5: Frailty variance results of simulation study for 6 different data scenarios.

Scenario	Parameter	Truth	Correlated frailty model			Independent frailty model				
			Mean	(avSE; empSE)	Bias	RMSE	Mean (empSE)	Bias	RMSE	
A	Var (W_{k1})	0.25	0.20	(0.13; 0.15)	-0.05	0.16	0.72	(0.28)	0.47	0.55
	Var (W_{k2})	0.25	0.20	(0.13; 0.15)	-0.05	0.16	0.59	(0.29)	0.34	0.45
	Cor (W_{k1}, W_{k2})	0.3	0.29	(0.20; 0.29)	-0.01	0.29				
B	Var (W_{k1})	0.25	0.24	(0.10; 0.10)	-0.01	0.10	0.57	(0.18)	0.32	0.37
	Var (W_{k2})	0.25	0.24	(0.10; 0.10)	-0.01	0.10	0.42	(0.18)	0.17	0.24
	Cor (W_{k1}, W_{k2})	0.3	0.3	(0.21; 0.22)	0.00	0.22				
C	Var (W_{k1})	0.25	0.25	(0.07; 0.08)	0.00	0.08	0.39	(0.13)	0.14	0.19
	Var (W_{k2})	0.25	0.25	(0.08; 0.08)	0.00	0.08	0.30	(0.12)	0.05	0.12
	Cor (W_{k1}, W_{k2})	0.30	0.33	(0.19; 0.17)	0.03	0.17				
D	Var (W_{k1})	0.25	0.25	(0.06; 0.06)	0.00	0.06	0.29	(0.09)	0.04	0.10
	Var (W_{k2})	0.25	0.24	(0.07; 0.07)	-0.01	0.07	0.25	(0.08)	0.00	0.08
	Cor (W_{k1}, W_{k2})	0.3	0.33	(0.16; 0.15)	0.03	0.16				
E	Var (W_{k1})	0.1	0.09	(0.04; 0.04)	-0.01	0.04	0.18	(0.12)	0.08	0.15
	Var (W_{k2})	0.3	0.19	(0.08; 0.08)	-0.11	0.13	0.41	(0.18)	0.11	0.22
	Cor (W_{k1}, W_{k2})	0.80	0.67	(0.17; 0.15)	-0.13	0.20				
F	Var (W_{k1})	0.25	0.20	(0.08; 0.08)	-0.05	0.20	0.51	(0.18)	0.26	0.32
	Var (W_{k2})	0.25	0.20	(0.08; 0.09)	-0.05	0.10	0.37	(0.19)	0.12	0.22
	Cor (W_{k1}, W_{k2})	-0.3	0.02	(0.05; 0.06)	0.32	0.32				

Abbreviations and notation: empSE, empirical standard error; avSE, average standard error; RMSE, root-mean-square error; W_{kj} ($j = 1, 2$), center-specific frailty for cause j .

Table 2.6: Empirical Bayes results of simulation study for 6 different data scenarios.

Scenario	Parameter	Bias	RMSE	Coverage	Bias($F(t_1)$)	Bias($F(t_2)$)	Bias($F(t_3)$)	RMSE($F(t_1)$)	RMSE($F(t_2)$)	RMSE($F(t_3)$)
A	W_{k1}	0.10	0.30	0.39	0.00	0.00	-0.04	0.01	0.02	0.05
	W_{k2}	0.10	0.30	0.49	0.00	0.00	-0.02	0.01	0.01	0.03
B	W_{k1}	0.12	0.23	0.75	0.00	0.00	-0.04	0.02	0.03	0.06
	W_{k2}	0.11	0.24	0.82	0.00	0.00	-0.02	0.01	0.03	0.04
C	W_{k1}	0.10	0.23	0.88	0.00	0.00	-0.04	0.02	0.04	0.07
	W_{k2}	0.09	0.25	0.91	0.00	0.00	-0.03	0.02	0.03	0.05
D	W_{k1}	0.09	0.25	0.92	0.00	0.00	-0.04	0.03	0.05	0.07
	W_{k2}	0.08	0.28	0.93	0.00	0.00	-0.02	0.02	0.04	0.06
E	W_{k1}	0.04	0.14	0.86	0.00	0.00	-0.04	0.02	0.03	0.06
	W_{k2}	0.10	0.23	0.80	0.00	0.00	-0.02	0.01	0.02	0.04
F	W_{k1}	0.09	0.20	0.79	0.00	0.00	-0.04	0.02	0.03	0.06
	W_{k2}	0.08	0.22	0.85	0.00	0.00	-0.03	0.01	0.03	0.05

Abbreviations and notation: RMSE, root-mean-square error; Coverage, coverage of 95% prediction intervals; F , cause-specific cumulative incidence; t_1, t_2, t_3 , quartiles of overall event time distribution; W_{kj} ($j = 1, 2$), center-specific frailty for cause j .

Table 2.7: Failed estimation of standard error.

Scenario	Nbr. of data sets	Successful runs	Success attempt	second	Failed estimation	Evaluated
A	1200	1062	80		138	1000
B	1200	1165	17		35	1000
C	1200	1117	1		83	1000
D	1200	1142	0		58	1000
E	1200	1046	232		154	1000
F	1200	1196	0		4	1000

§2.7 Discussion and Conclusion

Using shared frailty models to account for unobserved covariates in multicenter studies is common practice to avoid bias and to measure the amount of heterogeneity between centers. Correlated frailty models extend the shared frailty model by incorporating dependence structures between related individuals. Dependence among transition intensities of competing risks have come of interest [162, 99, 123].

The model presented uses correlated gamma frailties to model dependence within hospitals and between two competing risks. The mathematical properties of the gamma distribution are exploited to construct and estimate correlated frailties. An estimation procedure using the EM-algorithm is outlined and estimation of the standard error is illustrated. The estimation procedure provides empirical Bayes estimates for hospital frailties, which together with their prediction intervals can be used to compare hospital effects. The model is applied to breast cancer data and a moderate correlation between the frailties of the competing events recurrence and death is estimated. A simulation study is conducted to investigate performance of the method in different situations. Data scenarios with differing number of centers and correlation structures are considered and estimates of a model with independent frailties are compared to the proposed correlated frailty model. The performance of the empirical Bayes estimates obtained by the method was studied under different conditions.

The independent frailty model showed that it is not capable of accounting for center frailty in case of correlation between frailties. The correlated frailty model outperformed it in all data scenarios, concerning estimates as well as size of empirical standard errors. Its estimation benefits from a larger number of centers in the data. In data scenarios with unattainable correlation structures it still performed reasonably well and behaved in an expectable way.

The method is well suited to investigate hospital heterogeneity in the presence of competing risks. It distinguishes between common and separate effects of a hospital on two competing events and performed well in a simulation study. The proposed model can be extended to the case of more than two competing events. Dependence between risks can be modeled by adding frailty components, where shared components induce dependence between risks. However, the model is limited to positive correlation between frailties.

Wienke et al.[162] pointed out that in the case of cause-specific mortality the presence of risk factors might increase the risk of death with respect to all disease, making the case for positive dependence between risks. At the same time he argues that everyone dies eventually, so if the risk of death from one cause is decreased the risk from another cause must be increased, which suggests negative correlation between risks. Further study should be dedicated to the nature of dependencies among competing risks.

Putter and van Houwelingen [121] compare a two-point frailty distribution to a gamma distribution to model association between transition times in multi-state models. An advantage of the two-point frailty model is that it allows the two frailty terms to operate on different scale and that, in contrast to the gamma distribution, it allows negative association. In their simulation study the two-point frailty outperforms the

gamma distribution. A similar model could be used in the competing risks setting modelling dependence between risks, possibly with three or four points.

Acknowledgement

Research leading to this article was supported by the KWF Kankerbestrijding grant UL2015-8028. The European Organisation for Research and Treatment of Cancer (EORTC) is gratefully acknowledged for making data from EORTC trial 10854 available for this analysis.

Appendix

§2.A Probabilities for E-step

Let z_{k0} , z_{k1} , z_{k2} be the independent gamma distributed frailty components and let d_{kj} ($j = 1, 2$) be the number of failures of type j in hospital k ($k = 1, \dots, K$). Defining $\Lambda_{kj} = \sum_{i=1}^{n_k} \Lambda_{j0}(t_{ki}) e^{\beta_j^T \mathbf{X}_{ki}}$ ($j = 1, 2$) the conditional probability of the data given frailty components is given as

$$\begin{aligned} f(\text{data}_k | z_{k0}, z_{k1}, z_{k2}) &= \prod_{i=1}^{n_k} \left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}) \right)^{1_{\{\delta_{ki}=1\}}} \left(\frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}) \right)^{1_{\{\delta_{ki}=2\}}} \\ &\quad \exp \left(- \left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \Lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}) + \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \Lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}) \right) \right) \\ &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} z_{k1}^l \right) \left(\sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} z_{k2}^m \right) \\ &\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\ &\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp \left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right) \exp \left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right). \end{aligned}$$

Integrating over the frailty components yields the following conditional probabilities

$$\begin{aligned} f(\text{data}_k | z_{k0}, z_{k2}) &= \int_{z_{k1}} f(z_{k1}) f(\text{data}_k | z_{k0}, z_{k1}, z_{k2}) dz_{k1} \\ &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} z_{k2}^m \right) \\ &\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\ &\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp \left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \\ &\quad \frac{1}{\Gamma(\nu_1)} \sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \end{aligned}$$

$$\begin{aligned} f(\text{data}_k | z_{k0}) &= \int_{z_{k2}} f(z_{k2}) f(\text{data}_k | z_{k0}, z_{k2}) dz_{k2} \\ &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_1) \Gamma(\nu_2)} \\ &\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\ &\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \\ &\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}} z_{k0}^{d_{k1} + d_{k2} - l - m} \end{aligned}$$

$$\begin{aligned}
f(\text{data}_k | z_{k0}, z_{k1}) &= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \left(\sum_{l=0}^{d_{k1}} \binom{d_{k1}}{l} z_{k0}^{d_{k1}-l} z_{k1}^l \right) \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \exp \left(-z_{k0} \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right) \right) \exp \left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right) \\
&\quad \frac{1}{\Gamma(\nu_2)} \sum_{m=0}^{d_{k2}} \binom{d_{k2}}{m} z_{k0}^{d_{k2}-m} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}}
\end{aligned}$$

$$\begin{aligned}
f(\text{data}_k | z_{k1}) &= \int_{z_{k0}} f(z_{k0}) f(\text{data}_k | z_{k0}, z_{k1}) dz_{k0} \\
&= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0) \Gamma(\nu_2)} \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}} \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{d_{k1} + d_{k2} + \nu_0 - l - m} \right)} \\
&\quad z_{k1}^l \exp \left(-z_{k1} \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)
\end{aligned}$$

$$\begin{aligned}
f(\text{data}_k | z_{k2}) &= \int_{z_{k0}} f(z_{k0}) f(\text{data}_k | z_{k0}, z_{k2}) dz_{k0} \\
&= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0) \Gamma(\nu_1)} \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \left(\frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{d_{k1} + d_{k2} + \nu_0 - l - m} \right)} \\
&\quad z_{k2}^m \exp \left(-z_{k2} \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)
\end{aligned}$$

The observed data likelihood is given as

$$\begin{aligned}
f(\text{data}_k) &= \int_{z_{k0}} f(z_{k0}) f(\text{data}_k | z_{k0}) dz_{k0} \\
&= (\nu_0 + \nu_1)^{-d_{k1}} (\nu_0 + \nu_2)^{-d_{k2}} \frac{1}{\Gamma(\nu_0) \Gamma(\nu_1) \Gamma(\nu_2)} \\
&\quad \prod_{i=1}^{n_k} \left\{ (\lambda_{10}(t_{ki}) \exp(\beta_1^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=1\}}} (\lambda_{20}(t_{ki}) \exp(\beta_2^T \mathbf{X}_{ki}))^{1_{\{\delta_{ki}=2\}}} \right\} \\
&\quad \sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} \right)^{l + \nu_1}} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{m + \nu_2}} \\
&\quad \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2} \right)^{d_{k1} + d_{k2} + \nu_0 - l - m}}
\end{aligned}$$

The conditional probabilities of the frailty components given the data necessary for the E-step are given as

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

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

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

where

$$\begin{aligned} \tilde{c}_k(l, m, \nu_0, \nu_1, \nu_2) &= \\ &\binom{d_{k1}}{l} \binom{d_{k2}}{m} \frac{\Gamma(l + \nu_1)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1}\right)^{l + \nu_1}} \frac{\Gamma(m + \nu_2)}{\left(1 + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)^{m + \nu_2}} \frac{\Gamma(d_{k1} + d_{k2} + \nu_0 - l - m)}{\left(1 + \frac{1}{\nu_0 + \nu_1} \Lambda_{k1} + \frac{1}{\nu_0 + \nu_2} \Lambda_{k2}\right)^{d_{k1} + d_{k2} + \nu_0 - l - m}} \\ c_k(l, m, \nu_0, \nu_1, \nu_2) &= \frac{\tilde{c}_k(l, m, \nu_0, \nu_1, \nu_2)}{\sum_{l=0}^{d_{k1}} \sum_{m=0}^{d_{k2}} \tilde{c}_k(l, m, \nu_0, \nu_1, \nu_2)}. \end{aligned}$$

§2.B Observed information of regression parameters

The term $\Sigma_{\eta\eta} = I_{\eta\eta}^{-1}$ can be computed as described by Louis [101].

Let ℓ^* and ℓ be the log-likelihood and the conditional log-likelihood given frailties. The Fisher information for $\hat{\eta}$ can be rewritten in terms of the conditional log-likelihood given as

$$\begin{aligned} I_{\eta\eta}(\nu) &= \mathbb{E}_{\nu} \left(-\frac{\partial^2}{\partial \eta \partial \eta} \ell^*(\eta) \right) \tag{2.B.1} \\ &= \mathbb{E}_{\nu} \left(-\frac{\partial^2}{\partial \eta \partial \eta} \ell(\eta|\mathbf{W}) | \mathbf{W} \in R \right) \\ &\quad - \mathbb{E}_{\nu} \left(\frac{\partial}{\partial \eta} \ell(\eta|\mathbf{W}) \frac{\partial}{\partial \eta} \ell^T(\eta|\mathbf{W}) | \mathbf{W} \in R \right) \\ &\quad + \frac{\partial}{\partial \eta} \ell^*(\eta) \frac{\partial}{\partial \eta} \ell^{*T}(\eta), \end{aligned}$$

where \mathbf{W} are the unobserved frailties and R is the set of possible frailties given the data. Notably the last term is zero at the MLE and thus a simplified notation for the Fisher information at the MLE is given as

$$I_{\eta\eta} = I_{\eta\eta}^{(full)} - I_{\eta\eta}^{(loss)}$$

where the first term represents the full information and the second term represents the loss of information due to the unobserved frailties.

Let

- 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
- $d_{kl'}$: 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 $t_{m'}$
- $t_{kl}, l = 1, \dots, d_{k1}$: ordered event times for cause 1 in hospital k
- $t_{km}, m = 1, \dots, d_{k2}$: ordered event times for cause 2 in hospital k
- $t_{l'}, (l' = 1, \dots, d_1)$: ordered event times for cause 1
- $t_{m'}, (m' = 1, \dots, d_2)$: ordered event times for cause 2
- $\Lambda_{10}(t) = \sum_{t_{l'} \leq t} \lambda_{10}(t_{l'})$
- $\Lambda_{20}(t) = \sum_{t_{m'} \leq t} \lambda_{20}(t_{m'})$
- $\sum_{i=1}^{n_k} e^{\beta_1^T \mathbf{X}_{ki}} \Lambda_{10}(t_{ki}) = \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i:t_{ki} \geq t_{l'}} e^{\beta_1^T \mathbf{X}_{ki}}$
- $\sum_{i=1}^{n_k} e^{\beta_2^T \mathbf{X}_{ki}} \Lambda_{20}(t_{ki}) = \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i:t_{ki} \geq t_{m'}} e^{\beta_2^T \mathbf{X}_{ki}}$
- $R_k(t) = \{i : t_{ki} \geq t\}$: risk set at time t for hospital k

The conditional log-likelihood given frailties can be expressed as

$$\begin{aligned} \ell = & \sum_k d_{k1} \log\left(\frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1}\right) + \sum_{l=1}^{d_{k1}} \log(\lambda_{10}(t_{kl})) + \sum_{l=1}^{d_{k1}} \beta_1^T \mathbf{X}_{kl} \\ & - \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i \in R_k(t_{l'})} e^{\beta_1^T \mathbf{X}_{ki}} \\ & + d_{k2} \log\left(\frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2}\right) + \sum_{m=1}^{d_{k2}} \log(\lambda_{20}(t_{km})) + \sum_{m=1}^{d_{k2}} \beta_2^T \mathbf{X}_{km} \\ & - \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i \in R_k(t_{m'})} e^{\beta_2^T \mathbf{X}_{ki}}. \end{aligned}$$

The term $I_{\eta\eta}^{(loss)}$ is the product of the gradient vector of the conditional log-likelihood with itself. The elements of the gradient vector are:

$$\begin{aligned} \frac{\partial}{\partial \beta_{1j}} \ell &= \sum_k \left\{ \sum_{l=1}^{d_{k1}} \mathbf{X}_{klj} - \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i \in R_k(t_{l'})} \mathbf{X}_{kij} e^{\beta_1^T \mathbf{X}_{ki}} \right\} \\ \frac{\partial}{\partial \beta_{2j}} \ell &= \sum_k \left\{ \sum_{m=1}^{d_{k2}} \mathbf{X}_{kmj} - \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i \in R_k(t_{m'})} \mathbf{X}_{kij} e^{\beta_2^T \mathbf{X}_{ki}} \right\} \\ \frac{\partial}{\partial \lambda_{10l'}} \ell &= \sum_k \left\{ \frac{d_{kl'}}{\lambda_{10l'}(t_{l'})} - \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{i \in R_k(t_{l'})} e^{\beta_1^T \mathbf{X}_{ki}} \right\} \\ &= \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})} - \sum_k \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{i \in R_k(t_{l'})} e^{\beta_1^T \mathbf{X}_{ki}} \\ \frac{\partial}{\partial \lambda_{20m'}} \ell &= \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})} - \sum_k \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{i \in R_k(t_{m'})} e^{\beta_2^T \mathbf{X}_{ki}} \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2}{\partial \beta_{1j} \partial \beta_{1h}} \ell &= - \sum_k \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{l'=1}^{d_1} \lambda_{10}(t_{l'}) \sum_{i \in R_k(t_{l'})} \mathbf{X}_{kij} \mathbf{X}_{kih} e^{\beta_1^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \beta_{1j} \partial \beta_{2h}} \ell &= 0 \\ \frac{\partial^2}{\partial \beta_{1j} \partial \lambda_{10l'}} \ell &= - \sum_k \frac{z_{k0} + z_{k1}}{\nu_0 + \nu_1} \sum_{i \in R_k(t_{l'})} \mathbf{X}_{kij} e^{\beta_1^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \beta_{1j} \partial \lambda_{20m'}} \ell &= 0 \\ \frac{\partial^2}{\partial \beta_{2j} \partial \beta_{2h}} \ell &= - \sum_k \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{m'=1}^{d_2} \lambda_{20}(t_{m'}) \sum_{i \in R_k(t_{m'})} \mathbf{X}_{kij} \mathbf{X}_{kih} e^{\beta_2^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \beta_{2j} \partial \lambda_{10l'}} \ell &= 0 \\ \frac{\partial^2}{\partial \beta_{2j} \partial \lambda_{20m'}} \ell &= - \sum_k \frac{z_{k0} + z_{k2}}{\nu_0 + \nu_2} \sum_{i \in R_k(t_{m'})} \mathbf{X}_{kij} e^{\beta_2^T \mathbf{X}_{ki}} \\ \frac{\partial^2}{\partial \lambda_{10p'} \partial \lambda_{10l'}} \ell &= 0, \quad \frac{\partial^2}{\partial \lambda_{10l'} \partial \lambda_{10l'}} \ell = - \frac{d_{1l'}}{\lambda_{10l'}(t_{l'})^2} \\ \frac{\partial^2}{\partial \lambda_{10p'} \partial \lambda_{20m'}} \ell &= 0 \\ \frac{\partial^2}{\partial \lambda_{20p'} \partial \lambda_{20m'}} \ell &= 0, \quad \frac{\partial^2}{\partial \lambda_{20m'} \partial \lambda_{20m'}} \ell = - \frac{d_{2m'}}{\lambda_{20m'}(t_{m'})^2} \end{aligned}$$

