UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Joint Models for Time-to-Event and Categorical Longitudinal Outcomes

Auteur·es : Hortense Doms

Promoteur·rices : Prof. Catherine Legrand

Lecteur·rices : Prof. Anouar El Ghouch et Prof. Philippe Lambert

Année académique 2019-2020

Contents

1 Introduction

In medical studies, while the primary interest is often to record the time at which a particular event occurs, information on multiple covariates is also collected longitudinally throughout the follow-up period. Many methods allow the longitudinal and time-to-event responses to be studied separately, but in some situations it is not appropriate to make a separate analysis. Indeed, longitudinal measurements could have a predictive role in the analysis of patients survival. A new objective is therefore to measure the association between repeated measurements and the risk of having an event. However, a longitudinal response may be measured with error and its complete path may be unknown.

A class of statistical models has been developed to account for the particularities of these time-dependent covariates. These are joint models for longitudinal and time-to-event data. These models are needed mainly in two settings. First, when the primary interest is the survival outcome while taking into account the effect of time-dependent covariates measured in error. In the second setting, the focus is on the longitudinal outcome and we want to correct for nonrandom dropout.

Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) introduce the standard joint model in which the longitudinal response is normally distributed and the time-to-event response is described by a proportional hazard model. In the literature on joint modelling framework, we can find both frequentist and Bayesian approaches to estimate the parameters of the model. The main frequentist estimation method consist in maximizing the likelihood (Wulfsohn and Tsiatis, 1997, Wulfsohn and Tsiatis, 1997). Other authors opted for joint modelling under a Bayesian paradigm, as Faucett and Thomas (1996), Wang and Taylor (2001), Brown and Ibrahim (2003) and Ibrahim et al. (2004). They use Markov Chain Monte Carlo techniques to obtain the parameters estimates. Moreover, Faucett and Thomas (1996) proposed non-informative priors on all the parameters in order to obtain similar results to maximum likelihood approaches.

Several extensions have been proposed for the standard joint model. One of them consist on the inclusion of a non-Gaussian longitudinal outcome (Faucett et al., 1998; Li et al., 2009 ; Albert and Follman, 2000; Rizopoulos, 2016). The recent work of Viviani et al. (2012) presents the use of generalized linear mixed effects models to incorporate non-Gaussian longitudinal response in joint models, with particular focus to Poisson and Binomial responses. Others extensions make it possible to introduce multiple longitudinal outcomes in joint models (Brown et al., 2005; Ibrahim, Chen and Sinha, 2001; Rizopoulos and Ghosh, 2011) or to introduce competing risks for time-to-event process (Elashoff et al. 2008).

Complete overviews of joint modelling techniques can be found in Tsiatis and Davidian (2004), Rizopoulos (2012) and Gould et al. (2015). Rizopoulos' book (2012) provides a comprehensive presentation of standard joint models for continuous longitudinal and time-to-event outcomes. It also presents several extensions to the standard joint model. These include several types of parameterizations for the structure of the association between longitudinal and event outcomes and the use of a generalized linear mixed model to account for categorical longitudinal outcomes in the joint model.

Scope and goals of the thesis

We illustrate the joint modelling approach by applying it to a database of intensive care units (ICU). Patients admitted to ICU are likely to develop cerebral dysfunction and, in particular, delirium. Brain dysfunction can be caused by a prolonged state of low blood pressure, called hypotension. However, the potential impact of hypotension on the development of delirium is not well known. It is not known whether the occurrence of multiple brief hypotension episodes can lead to brain damage.

The objective of this work is therefore to use the joint modelling framework to assess whether there is an association between hypotension and the risk of developing delirium. More specifically, we will face this challenge from two different points of view:

1. The main objective is to analyse the impact of hypotension on the brain by considering only the occurrence or non-occurrence of at least one hypotension episode in a day. Does having at least one episode in a day increase the risk of developing delirium?

2. The secondary objective is to take into account the number of episodes observed during a day for a patient instead of only considering whether or not at least one episode occurred that day. We want to assess whether the number of hypotension episodes is related to the risk of having delirium? Does a patient with a higher number of hypotension episodes in a day have a higher risk of delirium on that day?

In order to answer the above research questions, this thesis is structured as follows:

Chapters 2 and 3 present the building blocks of joint models, which are generalized linear mixed-effect models for longitudinal data and proportional hazard models for survival data. In particular, in Chapter 2, we introduce linear mixed-effects models and generalized linear models to describe generalized mixed-effects linear models for Bernoulli and Poisson distributed data.

Chapter 3 first explains the special features of survival data and introduces the basic concepts. We then present the proportional hazards model and in particular the Cox model. Finally, we focus on time-dependent covariates and discuss in which setting it is appropriate to include time-dependent variables in a Cox model.

Chapter 4 begins by motivating joint models from an survival point of view. We introduce the standard joint models for continuous and categorical longitudinal responses. We describe estimation using two different approaches. One uses maximum likelihood estimation and the other is based on a Bayesian paradigm and uses Monte Carlo Markov chain algorithms. We also propose an extension of the standard joint model by including different parameterizations between longitudinal and survival outcomes.

In Chapter 5, we apply the appropriate joint models to ICU data to address the research questions cited above. We first fit the generalized linear mixed-effects and Cox submodels. Next, we fit the joint models in the Bayesian framework.

Finally, Chapter 6 concludes this thesis with a discussion of the general conclusions and outlines possible avenues for future work.

2 Longitudinal data analysis

Longitudinal data is data that is collected through a series of repeated observations of the same subjects over a period of time. These analyses are widely used, particularly in the medical field. The outcome of interest is measured repeatedly for each subject but the number of observations and their time points may vary from subject to subject. The time intervals between two successive measurements of a subject may also vary. This leads to unbalanced longitudinal data. For example, during a medical study, patients may leave the study prematurely or may die; this implies unbalance. Another characteristic of longitudinal data is that longitudinal measurements on the same subject are likely to be positively correlated. This implies that the standard statistical tools are no longer appropriate. This section presents a type of model that takes these correlations into account and can adapt to any degree of imbalance in the data; the linear mixed-effects model.
Although this thesis focuses on the analysis of discrete longitudinal data, it is useful to first introduce the analysis of longitudinal data using continuous data (Verbeke and Molenberghs, 2000; West et al. 2006). Then, we introduce the generalized linear model and finally extend the linear mixed model to binary or count data (Molenberghs and Verbeke, 2005; Brown and Prescott, 2006; Fitzmaurice et al. 2008 - Chapter 4, Fitzmaurice et al., 2012).

2.1 Linear mixed model

Measurements made repeatedly for the same subject are typically correlated. We therefore examine the idea that each subject in the population has its own subject-specific mean response profile over time. To achieve this, we introduce the linear mixed model that includes both fixed and random effects. Fixed effects describe the average response for the entire population, while random effects are specific to subjects in the population. Random effects describe the deviation of each subject's profile from the average profile of the population. They account for the correlation between the repeated measurements of each subject.

We let y_{ij} represent the measure of the response variable Y taken on the j-th occasion for the i-th subject. The value of j ($j = 1, \ldots, n_i$), indexes the n_i longitudinal observations for a given subject and i ($i = 1, \ldots, n$) indicates the i-th subject. The general form of the linear mixed model for the analysis of longitudinal responses for the i-th

subject is given by

$$\begin{cases} y_i &= X_i\beta + Z_i b_i + \epsilon_i \\ b_i &\sim N(0, D), \\ \epsilon_i &\sim N(0, \sigma^2 I_{n_i}) \end{cases} \tag{2.1}$$

where $y_i = (y_{i1}, y_{i1}, \ldots, y_{in_i})^t$ is the n_i-dimensional vector of all repeated measurements for the i-th subject. The number of observations, n_i, may vary from one subject to another.
The X_i is an $n_i \times p$ design matrix which represents the n_i known values for each of the p covariates. The β denotes the p-dimensional vector of fixed effects associated with the p covariates used in the X_i matrix.
The $n_i \times q$ design matrix Z_i is associated with the q-dimensional vector of the random effects b_i. Random effects are by definition random variables that follow a multivariate normal distribution. The covariance matrix denoted by D is a $q \times q$ matrix that is symmetric and positive-definite. Elements of this matrix are described as follows:

$$D = \mathrm{Var}(b_i) = \begin{pmatrix} \mathrm{Var}(b_{i0}) & \mathrm{cov}(b_{i0}, b_{i1}) & \ldots & \mathrm{cov}(b_{i0}, b_{iq-1}) \\ \mathrm{cov}(b_{i0}, b_{i1}) & \mathrm{Var}(b_{i1}) & \ldots & \mathrm{cov}(b_{i1}, b_{iq-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(b_{i0}, b_{iq-1}) & \mathrm{cov}(b_{i1}, b_{iq-1}) & \ldots & \mathrm{Var}(b_{iq-1}) \end{pmatrix}.$$

We therefore notice that this linear mixed model implies that some of the regression parameters are the same for all subjects while other parameters are subject-specific.
Then, ϵ_i is an n_i-dimensional vector of error terms. It is assumed that ϵ_i comes from a multivariate normal distribution with a covariance matrix equals to $\sigma^2 I_{n_i}$ where I_{n_i} is the n_i-dimensional identity matrix. The errors terms associated with the observations on the same subject are assumed to be uncorrelated and to have equal variance. Further, we assume that the errors terms ϵ_i and the random effects b_i are independent of each other.

The linear mixed model in (2.1) implies that, conditional on the random effect b_i, y_i is normally distributed with mean vector $X_i\beta + Z_i b_i$ and with covariance matrix $\sigma^2 I_{n_i}$. Let $p(y_i|b_i)$ and $p(b_i)$ be density functions, we obtain the marginal density function of y_i by

$$p(y_i) = \int p(y_i|b_i)\; p(b_i)\; db_i$$

which is the density function of the following normal distribution

$$y_i \sim N(X_i\beta, Z_iDZ_i^t + \sigma^2 I_{n_i})$$

The square matrix $V_i = Z_iDZ_i^t + \sigma^2 I_{n_i}$ represents the marginal covariance structure for the n_i measurements of the ith subject. Marginal covariance terms are generally different from zero, which shows that repeated measurements on the same subject are correlated since they share the same random effect b_i. We introduce the following conditional independence assumption,

$$p(y_i|b_i;\theta_y) = \prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta_y),$$

according to which the n_i longitudinal responses of the i-th subject are independent conditionally on her random effects. The vector θ_y refers the vector of parameters for the longitudinal model.

It is possible to modify the model defined in (2.1) and suppose another covariance matrix for the subject-specific error components. Indeed, in a linear mixed model, the error terms associated with repeated observations on the same subject can be correlated. We define Σ_i such that $\epsilon_i \sim N(0, \Sigma_i)$. We have supposed above that $\Sigma_i = \sigma^2 I_{n_i}$. We can extend this structure and allow different forms for Σ_i while keeping the matrix symmetric and positive-definite. A frequently used structure is called the compound symmetry and supposes that the all n_i errors terms have constant covariance σ_1^2 and variance $\sigma^2 + \sigma_1^2$. Another structure called the first-order autoregressive supposes that observations that are closer to each other over time have a higher correlation than observations that are further away in time. West et al. (2006) presents these different structures in detail. With these covariance structures, b_i and Σ_i now both try to capture the correlation in the repeated measurements but the conditional independence assumption no longer holds. In this thesis, we only use the simplest covariance matrix $\Sigma_i = \sigma^2 I_{n_i}$.

Random-intercepts model

We introduce a simple and commonly used type of mixed linear model: the random-intercept model given by

$$\begin{cases} y_{ij} = x_{ij}^t\beta + b_{i0} + \epsilon_{ij} \\ b_{i0} \sim N(0, \sigma_b^2) \\ \epsilon_{ij} \sim N(0, \sigma^2) \end{cases} \tag{2.2}$$

where x_{ij} is the p-dimensional vector of fixed-effects associated with the j-th measurement of the i-th subject. In addition to a linear average evolution over time, this model includes a random effect which assumes that each subject differs at baseline. To facilitate interpretation of model parameters, we express the model given by (2.2) as

$$\begin{aligned} y_{ij} &= x_{ij}^t\beta + b_{i0} + \epsilon_{ij} \\ &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_p x_{ijp} + b_{i0} + \epsilon_{ij} \\ &= (\beta_1 + b_{i0}) + \beta_2 x_{ij2} + \cdots + \beta_p x_{ijp} + b_{i0} + \epsilon_{ij} \end{aligned}$$

where $x_{ij1} = 1$ for all i and j and β_1 is the fixed effect intercept term in the model. The intercept for the i-th subject is equal to $\beta_1 + b_{i0}$. We have that b_{i0} represents a random deviation of the i-th subject's intercept from the population mean intercept β_1.

The marginal covariance matrix of the random-intercept model has the form $V_i = \sigma_b^2 1_{n_i} 1_{n_i}^t + \sigma^2 I_{n_i}$ where 1_{n_i} is the n_i-dimensional unit vector. Moreover, the variance of the j-th measure of the i-th is therefore equal to $\mathrm{Var}(y_{ij}) = \mathrm{Var}(b_i) + \mathrm{Var}(\epsilon_{ij}) = \sigma_b^2 + \sigma^2$ and is assumed to be constant over time. We suppose two measurements j and k on the same i-th subject corresponding to any two time points. The marginal covariance between these two measurements is defined by

$$\begin{aligned} \mathrm{Cov}(y_{ij}, y_{ik}) &= \mathrm{Cov}(x_{ij}^t\beta + b_{i0} + \epsilon_{ij}, x_{ik}^t\beta + b_{i0} + \epsilon_{ik}) \\ &= \mathrm{Cov}(b_{i0}, b_{i0}) + \mathrm{Cov}(b_{i0}, \epsilon_{ik}) + \mathrm{Cov}(\epsilon_{ij}, b_{i0}) + \mathrm{Cov}(\epsilon_{ij}, \epsilon_{ik}) \\ &= \mathrm{Var}(b_{i0}) = \sigma_b^2 \end{aligned}$$

and the corresponding correlation is obtained by

$$\begin{aligned} \mathrm{Corr}(y_{ij}, y_{ik}) &= \frac{\mathrm{Cov}(y_{ij}, y_{ik})}{\sqrt{\mathrm{Var}(y_{ij})}\,\sqrt{\mathrm{Var}(y_{ik})}} \\ &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2} \end{aligned}$$

which is a positive correlation. The covariance matrix is known as the compound symmetric matrix. However, we notice that the correlation between the observations

of two different subjects i and l is equal to zero:

$$\begin{aligned}\text{Cov}(y_{ij}, y_{lk}) &= \text{Cov}(x_{ij}^t\beta + b_{i0} + \epsilon_{ij}, x_{lk}^t\beta + b_{l0} + \epsilon_{lk}) \\ &= \text{Cov}(b_{i0}, b_{l0}) + \text{Cov}(b_{i0}, \epsilon_{lk}) + \text{Cov}(\epsilon_{ij}, b_{l0}) + \text{Cov}(\epsilon_{ij}, \epsilon_{lk}) \\ &= 0\end{aligned}$$

Random-intercepts and random-slopes model

Then, we can extend the model in (2.2) by including a random slopes term for each subject. This model is called the random-intercepts and random-slopes model:

$$\begin{cases} y_{ij} = x_{ij}^t\beta + b_{i0} + b_{i1}t_{ij} + \epsilon_{ij} \\ b_i \sim N(0, D) \\ \epsilon_{ij} \sim N(0, \sigma^2) \end{cases}$$

with $b_i = (b_{i0}, b_{i1})$ and the covariance matrix is equal to

$$D = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_{01}} \\ \sigma_{b_{01}} & \sigma_{b_1}^2 \end{pmatrix}.$$

This model assumes that the rate of change of the response is different for each subject. The structure of the marginal covariance matrix V_i of this model is different from that of the random-intercepts model. Indeed, the marginal function for any pair of responses of the same individual is given by:

$$\begin{aligned}\text{Cov}(y_{ij}, y_{ik}) &= \begin{bmatrix} 1 & t_{ij} \end{bmatrix} \; D \; \begin{bmatrix} 1 \\ t_{ik} \end{bmatrix} + \sigma^2 \\ &= \sigma_{b_1}^2 t_{ij} t_{ik} + \sigma_{b_{01}}(t_{ij} + t_{ik}) + \sigma_{b_0}^2 + \sigma^2\end{aligned}$$

The variance and the correlation are no longer constant over time. As expected, this model allows to introduce more flexibility in the covariance structure compared to the model (2.2).

2.2 Generalized linear model

We want to analyze other types of repeatedly measured outcomes that are not normally distributed. To achieve this, we use an extension of the linear models which is called the generalized linear models. These models make it possible to model the impact of fixed effects on all types of response, as long as the distribution of the response variable is assumed to belong to the exponential distribution family. A random response variable Y follows a distribution that belongs to the exponential family if the density is of the form

$$p(y|\theta,\phi) = \exp\left\{\frac{y\theta - \psi(\theta)}{\phi} + c(y,\phi)\right\}$$

where θ is a natural parameter and ϕ is a scale parameter. The functions $\psi(.)$ and $c(.,.)$ are known functions which have different forms for each distribution.

Let $y_1, y_2, \ldots, y_n$ be a sample of n independent observations from a distribution in the exponential family. It can be shown that the mean and variance of Y can be written such that

$$\begin{cases} \mathrm{E}(y_i) = \mu_i \\ \mathrm{Var}(y_i) = \phi v(\mu_i) \end{cases} \tag{2.3}$$

The variance of the response can be written as the product of the dispersion parameter, $\phi > 0$, and a variance function denoted $v(.)$ which depends on the mean μ_i.

We let $x_1, x_2, \ldots, x_n$ denote vectors of covariates such that x_i is a p-dimensional vector for the i-th subject. The systematic component of a generalized linear model can be expressed as

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},$$

where η_i is called the linear predictor and is equal to a linear combination of the regression coefficient β and the covariates, x_i.
In order to appropriately relate the mean response to the covariates, we introduce a link function $g(.)$ such that

$$g(\mu_i) = \eta_i = x_i^t \beta.$$

We obtain that the transformed mean response changes linearly with change in the values of the covariates. The link function takes a suitable form according to the distribution of the data. A possible choice can be $g(.) = \psi'^{-1}(.)$ and this type is called the canonical link function.

We now introduce the elements of the generalized linear model for the Normal, Bernoulli and Poisson distributions. These three distributions belong to the family of exponential distributions.

The linear regression model

Suppose that Y is normally distributed with mean μ and variance σ^2. The normal distribution is part of the exponential family with the parameters $\theta = \mu$ and $\phi = \sigma^2$. The variance function is equal to $v(\mu) = 1$ and the canonical link function is equal to the identify function. This leads to the model $y_i \sim N(\mu_i, \sigma^2)$ with $\mu_i = x_i^t\beta$.

The logistic regression model

Let Y be Bernoulli distributed such that $P(Y = 1) = \pi$. The variance of a binary response is equal to $\text{Var}(y_i) = \pi_i(1-\pi_i)$, which depends on the mean π_i. We note that $\phi = 1$. The canonical link function is the logit link defined by $\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$. We obtain the resulting logistic regression model given by $y_i \sim \text{Bernoulli}(\pi_i)$ with,

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = x_i^t\beta.$$

This model assumes a linear relationship between the log odds of success and x_i, where the odds of success are defined by $\pi_i/(1-\pi_i)$.
This model can also be expressed in terms of probability of success,

$$\pi_i = \frac{\exp(x_i^t\beta)}{1+\exp(x_i^t\beta)} = \text{logistic}(x_i^t\beta).$$

The Poisson regression model

Let Y be Poisson distributed with mean and variance λ. Suppose $\theta = \log(\lambda)$, the density can be written as a member of the exponential family with $\phi = 1$. Since $\psi(\theta) = \exp(\theta)$, the canonical link function is the logarithm. We obtain the Poisson regression model described $y_i \sim \text{Poisson}(\lambda_i)$, with $\log(\lambda_i) = x_i'\beta$.

2.3 Generalized linear mixed model

In the same way that we extended the linear models to the generalized linear models, it is possible to extend the linear mixed models to the generalized linear mixed models (GLMMs). As in linear mixed models, GLMMs take into account the correlation between non-Gaussian repeated measurements by introducing random effects.

As before, y_i is the n_i-dimensional vector of all measurements available for the subject i. It is assumed that, conditionally on q-dimensional random effects b_i, the responses y_i are independent with densities belonging to the exponential family of distributions.

The general form of the generalized linear mixed model for longitudinal responses is given by

$$\begin{cases} g\{E(y_i \mid b_i, X_i, Z_i)\} = g\{\mu_i\} = X_i\beta + Z_i b_i \\ b_i \sim \quad N(0, D), \end{cases} \tag{2.4}$$

where $g(.)$ is a link function, X_i is the matrix of fixed-effects covariates and Z_i is the matrix of random-effects covariates.

Let y_{ij} be the j-th binary response measured for the i-th subject, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Conditional on random-effects, we have that y_{ij} follows a Bernoulli distribution with mean $\mathrm{E}(y_{ij}|b_i) = \pi_{ij}$ and variance $\mathrm{Var}(y_{ij}|b_i) = \pi_{ij}(1 - \pi_{ij})$. The mixed logistic regression model has the form

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x_{ij}^t\beta + z_{ij}^t b_i$$

where x_{ij} and z_{ij} are vectors of fixed-effects covariates and random-effects covariates, respectively. The random effects still follow a multivariate normal distribution with mean zero and covariance matrix D.

Suppose now that y_{ij} represents count independent responses. Conditional on random effects, y_{ij} has a Poisson distribution with mean and variance $\mathrm{E}(y_{ij}|b_i) = \mathrm{Var}(y_{ij}|b_i)) = \lambda_{ij}$. The mixed Poisson regression model has the form

$$\log(\lambda_{ij}) = x'_{ij}\beta + z'_{ij}b_i.$$

This model can also be called a mixed-effects log-linear model.

Estimation

We want to estimate the parameters of the GLMMs using a maximum likelihood method since we have a complete specification of the distribution of the data. We define that log-likelihood expression by

$$l(\theta) = \sum_{i=1}^{n} \log \int p(y_i \mid b_i; \theta) p(b_i; \theta) db_i$$

where θ are the parameters of the model. However, the two terms of the integrand represent two different distributions. The density of the random effects follows a normal multivariate distribution while $y_i|b_i$ follows a non-normal distribution. In general, the integral in the log-likelihood expression does not have a closed-form solution. Since

no analytic expressions are available for the integral, a numerical approximation is needed. A frequently used one is based on an approximation of the data and called the penalized quasi-likelihood estimation. (Molenberghs and Verbeke 2005, Tuerlinckx et al. 2006)

To introduce the approximation, we first decompose the observations into the means and an error terms,

$$y_{ij} = \mu_{ij} + \epsilon_{ij} = h(x_{ij}^t \beta + z_{ij}^t b_i) + \epsilon_{ij} \tag{2.5}$$

where $h(.) = g^{-1}(.)$ equals the inverse link function. The errors terms are distributed with a mean of zero and a variance equal to $\text{Var}(Y_{ij}|b_i) = \phi v(\mu_{ij})$. The $v(.)$ function was introduced in Section 2.2. Moreover, by using the canonical link function and by the definition of the variance function, we have that $v(\mu_{ij}) = h'(x_{ij}^t \beta + z_{ij}^t b_i)$.

The PQL method performs a first-order Taylor approximation of the model given by (2.5) about the current estimates of the fixed effects $\hat{\beta}$ and of the random effects $\hat{b_i}$. The observation y_{ij} can be then written as:

$$\begin{aligned} y_{ij} \approx \; & h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{b}_i) \\ & + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{b}_i) x_{ij}^t (\beta - \hat{\beta}) \\ & + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{b}_i) z_{ij}^t (b_i - \hat{b}_i) + \epsilon_{ij} \end{aligned}$$

In the vector notation,

$$y_i \approx \hat{\mu}_i + \hat{V}_i X_i (\beta - \hat{\beta}) + \hat{V}_i Z_i (b_i - \hat{b}_i) + \epsilon_i$$

with $\hat{V}_i$ is a diagonal matrix with the diagonal elements $v(\hat{\mu}_{ij})$. We reorganize this equation by bringing $\hat{\mu}_i$ to the left-hand side, pre-multiplying by $\hat{V}_i^{-1}$ and then bringing $X_i \hat{\beta}$ and $Z_i \hat{b}_i$ to the left-hand side. This yields

$$y_i^* \equiv \hat{V}_i^{-1}(Y_i - \hat{\mu}_i) + X_i \hat{\beta} + Z_i \hat{b}_i \approx X_i \beta + Z_i b_i + \epsilon_i^*,$$

where $\epsilon_i^* = \hat{V}_i^{-1} \epsilon_i$. We recognize that the pseudo responses y_i^* follow a linear mixed model as introduced in (2.1) in Section 2.1.

To estimate the parameters of this linear mixed model, we iterate between this two-step algorithm.

1. We fit the model to the responses y_i^* to obtain estimates of β and D. Then, we compute the predictions of b_i.

2. We use these estimates to update the pseudo responses y_i^*.

This algorithm is iterated until convergence is reached.

Note that in the linear mixed model the estimator of β is the generalized least squares estimator and the estimator of D is the standard maximum likelihood or restricted maximum likelihood estimators. Note that in the linear mixed model, the estimates of β are obtained with the generalized least squares (GLS) estimator and the estimates of D are obtained with the maximum likelihood (ML) estimator or the restricted maximum likelihood (REML) estimator. Moreover, predictions of b_i are obtained with the best linear unbiased predictor (BLUP) which corresponds to the conditional mean of b_i, given the response vector y_i. More specifically, this algorithm uses the empirical BLUP, which means that the unknown covariance parameters of the mixed linear model are replaced by their ML or REML estimates. The resulting predictor is given by

$$\hat{b_i} = \hat{D} Z_i^t {V_i^*}^{-1} (y_i^* - X_i \hat{\beta})$$

where $V_i^* = Z_i D Z_i^t + W_i^{-1}$ and W_i is a diagonal weight matrix with components w_{ij} on the diagonal. These weights are inversely proportional to the variance of the ϵ_{ij}^*.

Finally, this method is called a penalized quasi-likelihood estimation because it optimizes a quasi-likelihood function which only requires the mean and variance for the observations and there is a penalty term on the random effects to force them to stay close to the zero vector.

3 Survival data analysis

The main objective of survival analysis is to analyse the time until a prespecified event of interest occurs.. In medical studies this event is usually the death of the patient or the recurrence of an illness. The most important characteristic of survival data is the presence of incomplete data linked to a censoring mechanism. Censoring signifies that for certain subject the time to event of interest is not completely determined and is only known to be in a time interval. Depending on the structure of this interval, there are three types of censoring: right, left and by interval. For a left-censored observation the event of interest is only know to occur before a certain time point. In the case of interval-censoring, the event is know to occur between two known times. Right censoring is the most often used. It is common that at the end of a study, some subjects have not yet encountered the event of interest. These observations are called right-censored and we only know that the event occurs after a certain time point. In addition, censoring is generally assumed to be independent and non-informative. This means that the fact that a subject is censored does not affect its actual event time. In this thesis we will only consider non-informative right censoring.
This section is mainly based on the following books: Klein and Moeschberger (2003), Moore (2016), Therneau and Grambsch (2000) and Kalbfleisch and Prentice (2002).

3.1 Basic concepts

Let T^* denote the non-negative continuous random variable which represents the time to an event of interest. There are many ways to represent and describe the distribution of T^*. The most useful in survival analysis is the survival function defined as

$$S(t) = \Pr(T^* > t)$$

It represents the probability that the event of interest has not yet occurred at time t. This is a continuous decreasing function which is equal to one at time $t = 0$ and equal to zero at time $t = \infty$.

Another useful function in survival analysis is the hazard function. This describes the instantaneous conditional risk for an event during the time interval $[t, t + dt)$, given that the event has not occured until t. It is a nonnegative function. The hazard function is given by

$$h(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T^* < t + dt \,|\, T^* \geq t)}{dt}, \;\; t > 0$$

The hazard can be interpretable as the expected number of events per subject per unit

of time. It is not a probability, $h(t)$ can be greater than one. The hazard function can also be called the risk function.

The survival and the hazard functions are linked by the following relationship:

$$S(t) = \exp\Big(- \int_0^t h(u)du \Big) = \exp(-H(t))$$

where $H(.)$ is the cumulative hazard function. This function represents the cumulative risk up to time t and can be interpreted as the expected number of events that have occurred by time t. It is an increasing function of t. Indeed, over time we expect more events to have occurred.

We have defined that T_i^* is a random variable that represents the true event time for the subject $i, i = 1, \ldots, n$. Due to censoring, T_i^* is not always observed for all subjects. We then introduce the observed event time defined by $T_i = \min(T_i^*, C_i)$ and this variable is observed for each of the n subjects. The non-negative random variable C_i represents the censoring time. We also introduce the censoring indicator $\delta_i = I(T_i^* \leq C_i)$ which is equal to 1 if the observed event time corresponds to the true event time. The survival analysis allows to make valid inferences for T_i^* using only the observed information $\{T_i, \delta_i\}$.

Parametric estimate of the survival function can be obtained by assuming a parametric distribution for T^* and calculating the maximum likelihood estimators of the parameters of the given distribution. Often times, it may be difficult to select an appropriate distribution for T^*. In this case, we can use non-parametric estimators which take into account the censoring. One widely used possibility is the Kaplan-Meier estimator which is based on the order in which events and censored observations occur. To introduce this estimator, we suppose that $t_{(1)}, \ldots, t_{(r)}$ denote r ordered event times and $d_{(1)}, \ldots, d_{(r)}$ are the numbers of event at the corresponding event times. The number of subjects still at risk at time t_i is given by r_i. A subject is said being still at risk at time $t_{(i)}$ when he is not censored and he has not yet had the event at a time just prior to $t_{(i)}$. The Kaplan-Meier estimator is expressed as

$$\hat{S}_{KM}(t) = \prod_{i\,:\,t_{(i)} \leq t} \frac{r_{(i)} - d_{(i)}}{r_{(i)}}$$

In some cases, one may be interested in comparing two population groups with respect to survival in order to test whether the distribution of survival times is the same for both groups. This analyse can be performed by the non-parametric Log-Rank test. The idea is that for each event times, the Log-Rank test calculates the observed

number of deaths in each group and the expected number if there was really no difference between the two groups. Since this is a non-parametric test, no distributional assumption is made for the survival times of the two groups.

3.2 Regression models for time-to-event

A major goal of survival analysis is to examine the relationship between the survival time of subjects and one or more predictor variables. To reach this goal, we introduce a popular model called the proportional hazards (PH) model (also known as relative risk models). It describes the instantaneous hazard of a subject according to a set of covariates. This model assumes that there is a multiplicative effect of the covariates on the hazard of an event. It is formulated as,

$$h_i(t|w_i) = h_0(t)\exp(\gamma^t w_i)$$

where $w_i^T = (w_{i1}, \ldots, w_{ip})^t$ is the vector of covariates for the subject i and γ is the corresponding vector of coefficients that measure the impact of covariates. This model allows for both quantitative and categorical covariates. The term $h_0(t)$ is called the baseline hazard function and corresponds to the value of the hazard of an event when all the $w_i = 0$ are equal to zero. That is, $h_0(t)$ represents the conditional instantaneous risk of having an event at time t without the influence of any covariates.

There are different ways of specifying this baseline hazard function, leading to different types of model. We distinguish the parametric and semi-parametric proportional hazard models. In the parametric PH model, the time to an event is supposed to follow a parametric distribution. The baseline hazard has therefore a specific parametric form. Common distribution choices include Weibull, exponential, log-normal, but also more flexible distributions such as piecewise constant functions or splines.
In the semi-parametric PH model, no distribution assumption is made for the baseline hazard function. This model is also called the Cox PH model or simply the Cox model. One of the advantages of this model is that it avoids the impact of a possibly wrong assumption for the distribution of event times.

The main assumption of the Cox model is the proportional hazards assumption. It assumes that the effect of a covariate in the hazard of an event is constant over time. If we look at two subjects i and j with respective covariate values w_i and w_j, the ratio of their hazards is

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)\ \exp(\gamma^t w_i)}{h_0(t)\ \exp(\gamma^t w_j)} \tag{3.1}$$

which is independent of time. To illustrate the quantity (3.1), we suppose that w_{i1} indicates the gender of a subject ($w_{i1} = 1$ if male and $w_{i1} = 0$ if female) and all other co-

variates have the same value between the two subjects. Then, $\exp(\gamma^t w_i)/\exp(\gamma^t w_j) = \exp(\gamma_1)$ is the risk of having the event for males relative to the risk of having the event for females. The quantities $\exp(\gamma_i)$ are called hazard ratios and a value of γ_i greater than zero indicates that as the value of the i-th covariate increases, the hazard of an event increases.

Estimation

The method used to estimated the model parameter of a semi-parametric Cox model is based on likelihood maximization. Since the baseline hazard function is not specified, one did not work on the standard full likelihood. We use an appropriate modification of this likelihood, called partial likelihood.
Suppose we have r observed event times indicated by $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ and $w_{(j)}$ is the vector of covariate associated with a subject whose event time is $t_{(i)}$. We assume that there is no ties between the event times. The partial likelihood is expressed as

$$L_p(\gamma) = \prod_{j=1}^{r} \frac{\exp(\gamma^t w_{(j)})}{\sum_{k \in R(t_{(j)})} \exp(\gamma^t w_k)} \tag{3.2}$$

where $R(t_{(j)})$ is the risk set at time $t_{(j)}$. The risk set represents the set of all subjects who are still at risk (alive and uncensored) at a given time. We notice that this partial likelihood does not depend on the underlying hazard function $h_0(t)$. The partial likelihood numerator depends only on the information from the subject who is experiencing the event at time t_j, while the denominator uses information of all subjects in the risk set at the same time. We also notice that unlike a likelihood function, this partial likelihood is not a probability, since the factors for the censored times are not included.

The maximum partial likelihood estimators are found by maximizing the following partial log-likelihood,

$$l_p(\gamma) = \sum_{j=1}^{r} \gamma^t w_{(j)} - \sum_{j=1}^{r} \log\Big(\sum_{k \in R(t_{(j)})} \exp(\gamma^t w_k) \Big).$$

The obtained estimates of γ are asymptomatically normally distributed.

When events are observed on a discrete time scale, censoring and event times are likely to be tied. This means that more than one event occurs at a given time. In this situation, we need to adjust the definition of partial likelihood. An approximation has been proposed by Beslow (1974). Let $d_{(j)}$ be the number of events at time $t_{(j)}$, the

approximation of the partial likelihood is expressed as

$$L_p(\gamma) = \prod_{j=1}^{r} \frac{\prod_{l:t_l=t_{(j)},\delta_l=1} \exp(\gamma^t w_l)}{\left[\sum_{k:t_k \geq t_{(j)}} \exp(\gamma^t w_k)\right]^{d_{(j)}}}$$

and we recall that $\delta_l = I(T_i \leq C_i)$ is equal to 1 if the subject i is not censored.

3.3 Time-dependent covariates

So far, we have assumed that the hazard depends only on the covariates whose value remains constant over time. However, it may also be interesting to investigate whether the risk for an event depends on time-dependent covariates. The Cox model is extended to handle time-dependent covariates using the counting process formulation (Anderson and Gill, 1982; Therneau and Grambsch, 2000). The event process which was described by $\{T_i, \delta_i\}$ in Section 3.1 is now replaced by $(N_i(t)), R_i(t))$ where

$$N_i(t) = \text{the number of oberved events in } [0,t] \text{ for subject } i$$
$$R_i(t) = \begin{cases} 1 & \text{subject } i \text{ is at risk at time t} \\ 0 & \text{otherwise} \end{cases} .$$

In the case of right-censored survival we have $N_i(t) = I(\{T_i \leq t, \delta_i = 1\})$ and $R_i(t) = I(T_i \geq t)$. The counting processes $N_i(t)$ is equal to zero until subject i has the event and then jumps to one. We notice that $N_i(t)$ cannot jump when $Y_i(t) = 0$. The process $N(t) = \sum_{i=1}^{n} N_i(t)$ is a stochastic process of which the sample paths are right-continuous and piecewise constant with jumps of height 1. The left-continuous process $R(t)$ is predictable, the value of the process at time t is known infinitesimally before t.

We suppose for simplicity that we only want to introduce a single time-dependent covariate denoted by $y_i(t)$ for the subject i. The extended Cox model is expressed as (Rizopoulos, 2012) :

$$h_i(t|\mathcal{Y}_i(t), wi) = h_0(t) R_i(t) \exp\{\gamma^t w_i + \alpha y_i(t)\} \tag{3.3}$$

where $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$ is the covariate history up to t. The regression coefficient parameter α accounts for the effect of the time-dependent covariate on the hazard at a given point in time. As the value of $y_i(t)$ increases by one unit at time t, the relative increase in the risk for an event at the same time point is measured by $\exp(\alpha)$. In this approach, the hazard ratio at time t between two subjects i and j, with baseline covariates w_i and w_j, and time-dependent covariates $y_i(t)$ and $y_j(t)$ is no

longer constant in time.

An important assumption of the model (3.3) is that the time-dependent covariates must be predictable processes, measured without error, and have their complete path fully specified. More specifically, the values of the time-dependent covariates may correspond to measurements made at different times such as follow-up visits. The extended Cox model assumes that these covariate values change at follow-up visits but remain constant over the time between visits.

We distinguish the time-dependent covariates into two categories namely endogenous and exogenous covariates. First, an exogenous covariate describes a predictable process, its value at time t is known infinitesimally before t. Moreover, its future path up to any time $t > u$ is not affected by the occurrence of an event at time u. This means that the value of exogenous covariates does not depend on the event process at any time. An exogenous covariate could be a covariate for which his complete path is determined in advance of the study for each subject. This type of covariate could correspond to the age of the subject or his treatment. Another sort of exogenous covariate could be the output of a stochastic process that is external to the subject under study. An example of such covariate would be a measure of the air pollution.
In contrast, endogenous covariates usually result from time-dependent measurements taken on individuals under study and in many instances require the survival of the subject in order to be observed. A common example arises in medical studies where biomarkers or clinical parameters of patients are measured at regular time intervals. If a patient dies, these measurements can no longer be observed for that patient. Furthermore, since such measurements are only known for specific occasions, the complete path up to any time t is not fully observed for endogenous covariates. Finally, it is often reasonable to assume that these covariates are typically measured with error.

We therefore obtain that exogenous covariates are appropriate to be used in the extended Cox model in (3.3). On the other hand, endogenous covariates do not respect at all the assumption of model (3.3) introduced above and it is not reasonable to assume that endogenous covariates, such as biomarkers have constant values between two follow-up visits. Since measurements are made only at a set of specific time points, biomarker values may be missing at some observed event time, and partial likelihood discussed in Section 3.2 cannot be defined at these time points. Some imputation methods such as "last observation carried forward" can be used, but may lead to a biased estimation of the model. We therefore need to introduce a new model that takes into account time-dependent endogenous covariates and their specificities.

4 Joint modelling

In medical studies, it is common that time-to-event data and repeated measurements data such as biomarkers are collected simultaneously for each subject over the same time period. As presented in Sections 2 and 3, there are standard methods for analysing these two types of data separately. Longitudinal data can be analysed with mixed models and time-to-event data can be analysed with survival models such as proportional hazard models to assess the hazard for an event based on predictor variables. However, studying these two types of data separately is sometimes not an optimal approach. Indeed, the longitudinal and time-to-event processes can be dependent and the interest can be to investigate their association. It is therefore necessary to study the two processes simultaneously.

An intuitive approach might be to incorporate a time-dependent covariate into the linear predictor of the proportional hazard Cox model. But this technique does not work with biomarkers, as discussed in section 3.3.

To remedy this, a new modelling framework called joint modelling for longitudinal and time-to-event data has been developed. The basic principle of joint modelling is first to model the survival data with a survival model (often, a proportional hazard model), then to model the longitudinal data using a mixed model, and finally to link these models using a common latent structure. Two types of joint model exist for longitudinal and time-to-event data namely the shared random-effect models (SREM) and the joint latent class models (JLCM). In literature, the shared random-effect approach has received the main interest in the joint modelling research. The SREMs introduce functions of the individual random-effects of the mixed model as predictors in the survival model defined for the time-to-event data. These random-effects capture the individual deviations to the mean trajectory of the biomarker. The JLCMs assume that the population of subjects is heterogeneous and can be divided in homogeneous subgroups. Each subgroup shares the same trajectory of the biomarker and the same hazard for the event. In this section, we describe the shared random-effect model in detail based on the following books and articles: Tsiatis and Davidian (2004), Rizopoulos (2012), Wu et al. (2012), Papageorgiou et al. (2018)

4.1 Longitudinal and survival submodels

As introduced in section 3.1, for subject i, $i = 1, \ldots, n$, we denote by T_i the observed event time defined as $T_i = \min(T_i^*, C_i)$ where T_i^* is the true event time and C_i is the censoring time. The value of the longitudinal outcome (e.g., a biomarker) observed at time point t for the subject i is denoted as $y_i(t)$. The n_i-dimensional

vector $y_i = (y_i(t_{i1}), \ldots, y_i(t_{in_i}))$ is observed on specific occasions t_{ij} during which $y_{ij} = y_i(t_{ij}), j = 1, \ldots, n_i$ measurements are collected.

As described above, the principle of the joint model is to model longitudinal response via a mixed model and to inject some of the information obtained into a survival model. However, we know that the longitudinal information y_i is collected intermittently and may be subject to measurement error. To resolve this, we consider the expected value of the longitudinal response at time t and introduce $\eta_i(t)$ such that $\eta_i(t) = g\{\mathbb{E}(y_i(t) \mid b_i)\}$. We denote $\mathcal{N}_i(t) = \{\eta_i(s), 0 \leq s < t\}$ the history of the true unobserved longitudinal process up to t. In order to properly include $\eta_i(t)$ in the survival model, we need to estimate it and reconstruct the complete longitudinal history $\mathcal{N}_i(t)$ for each subject using the conditional expectation of the observed response at time t, $y_i(t)$. We obtain the following generalized mixed effects model,

$$\begin{cases} \eta_i(t) = g\{\mathbb{E}(y_i(t) \mid b_i)\} = x_i^t(t)\,\beta + z_i^t(t)\,b_i \\ \\ b_i \sim N(0, D), \end{cases} \tag{4.1}$$

As the notations used in Section 2, β denotes the p-dimensional vector of fixed effects associated with the time-dependent design vector $x_i(t)$ and b_i denotes the the q-dimensional vector of random effects associated with $z_i(t)$. The presence of a time structure in the design vectors as well as the use of random effects allow this model to construct the full path of $\mathcal{N}_i(t)$ for each subject. The terms b_i allow each of the subjects to have a different response profile in time.

Then, to measure the effect of the longitudinal response to the hazard of an event, we build the following proportional hazard model:

$$\begin{aligned} h_i(t|\mathcal{N}_i(t), w_i) &= \lim_{dt \to 0} \frac{\Pr(t \leq T_i^* < t + dt \mid T_i^* \geq t, \mathcal{N}_i(t), w_i)}{dt} \\ &= h_0(t) \exp\{\gamma^t w_i + \alpha \eta_i(t)\}, \quad t > 0 \end{aligned} \tag{4.2}$$

where parameter α measures the strength of the association between the linear predictor at time t, $\eta_i(t)$ and the risk for an event at the same time. Specially, $\exp(\alpha)$ represents the relative increase in the risk for an event at time t that results from one unit increase in $\eta_i(t)$ at the same time t. The vector w_i represents the baseline covariates with the vector of regression coefficients γ. The associated survival function is defined by

$$S_i(t|\mathcal{N}_i(t), w_i) = \exp\Big(- \int_0^t h_0(s) \exp\{\gamma^t w_i + \alpha\eta_i(s)\}\ ds\Big) \tag{4.3}$$

Unlike the hazard model (4.2) which assumes that the risk for an event at time t depends only on the current value of $\eta_i(t)$, the survival function of the ith subject depends on the whole covariate history $\mathcal{N}_i(t)$ of this subject.

To complete the specification of the survival sub-model, we need to discuss to choice for the baseline hazard function $h_0(t)$. We have seen in section 3.1 that in survival analysis, the baseline hazard function is often left unspecified and that in this case a partial likelihood function can be used to estimate model parameters. However, in the joint modelling framework, leaving $h_0(t)$ completely unspecified may leads to an underestimation of the standard errors of the parameter estimates. Moreover, maximum likelihood estimation can no longer be obtain with a partial likelihood function in the joint modelling approach (Hsieh et al. 2006, Rizopoulos 2012).

To avoid this problem, it is advised to assume a parametric specification for the baseline hazard function. One approach is to choose a hazard function associated with a known parametric survival distribution such as Weibull, log-normal or Gamma distributions, among others. Another approach allows for more flexible specifications using step functions or B-splines approximation of the function $h_0(t)$.

Under the B-splines approach, the logarithm of the baseline hazard function is expressed as (Rizopoulos 2012)

$$\log\{h_0(t)\} = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, v) \tag{4.4}$$

where $B_q(t, v)$ denotes the qth basis function of a B-spline with knots $v_1, \ldots, v_Q$ and $\gamma_{h_0}^t = (\gamma_{h_0,0}, \gamma_{h_0,1}, \ldots, \gamma_{h_0,Q})$ are the spline coefficients and v. Choosing an ideal number of knots is an important but very complex task. Indeed, as the number of knots Q increase the approximation of $\log(h_0(t))$ becomes more flexible. However, too many knots can cause serious over-fitting. Moreover, the position of the knots may have a influence on estimation. An alternative method eliminates the need to make a precise choice of knots. This method includes several equally spaced knots and controls the regularity of the curve by using of a penalty on the B-spline coefficients. We obtain penalized B-splines, also called P-splines. These P-splines are described in the Bayesian framework by assuming that γ_{h_0} follows the prior (Lang and Brezger, 2004, Rizopoulos,

2016) :

$$p(\gamma_{h_0}|\tau) \propto \tau^{\rho(K)/2} \exp\big(-\frac{\tau}{2}\gamma_{h_0}^t \mathbf{K}\gamma_{h_0}\big) \tag{4.5}$$

where τ is the smoothing parameter and $\mathbf{K}$ denotes the penalty matrix of rank $\rho(K)$.

4.2 The joint model

The standard joint model introduced in Section 1 simultaneously analysing the longitudinal and survival sub-models given by (4.1) and (4.2) respectively. We assume that the association between the two sub-models is explained by the vector of time-independent random effects b_i. We also assume in Section 2 that these random effects explain the correlation between the repeated longitudinal measurements of a subject. Conditionally to the random effects, the longitudinal and time-to-event outcomes are independent and the repeated measurements are also independent of each other. The random effects are assumed to explain all the interdependencies. Formally we have, (Rizopoulos, 2006)

$$\begin{aligned} p(T_i, \delta_i, y_i|b_i;\theta) &= p(T_i, \delta_i|b_i;\theta)\; p(y_i|b_i;\theta) \\ p(y_i|b_i;\theta) &= \prod_j p(y_i(t_{ij})|b_i;\theta) \end{aligned} \tag{4.6}$$

where $\theta = (\theta_t^t, \theta_y^t, \theta_b^t)^t$ contains θ_t the parameters for the event time outcome, θ_y the parameters for the longitudinal outcome and θ_b the parameters of the random-effects covariance matrix. The function $p(.)$ denotes an appropriate probability density function.

Since random effects play an important role in joint models, it is necessary to appropriately specify their distribution. As defined in model (4.1), a common choice is to assume that the random effects follow a normal multivariate distribution with a mean zero and a variance covariance matrix D. Rizopoulos, Verbeke and Molenberghs (2008) has shown that this choice of distribution can be made safely because the higher the number of repeated measurements per subject, the smaller the effect of poor specification on parameter estimates and standard errors.
In addition, we assume that given the observed history, the

- censoring mechanism that describes how censoring occurs, and
- visiting process which generates the time points at which the longitudinal measurements are collected

are independent of the true event times and future longitudinal measurements. This assumption is no longer valid if one of the two processes depends on the random effects.

4.3 Estimation of Joint Models

4.3.1 Frequentist Estimation

Maximum likelihood estimation for joint models is a widely used traditional approach and is based on the maximization of the log-likelihood corresponding to the joint distribution of the observed outcomes $\{T_i, \delta_i, y_i\}$. This joint distribution is defined according to the assumptions formulated in (4.6) and the expression of the log-likelihood contribution for the i-th subject is written as

$$\begin{aligned} \log p(T_i, \delta_i, y_i; \theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta)\, db_i \\ &= \log \int p(T_i, \delta_i | b_i; \theta_t, \beta) \big[\prod_{j=1}^{n_i} p\big\{ y_i(t_{ij}) |\, b_i; \theta_y \big\} \big] p(b_i; \theta_b)\; db_i \end{aligned} \tag{4.7}$$

We recall that $\theta = (\theta_t^t, \theta_y^t, \theta_b^t)^t$ denotes the parameter vector. The log-likelihood function formulated for all subjects is obtained by $l(\theta) = \sum_i \log p(T_i, \delta_i, y_i; \theta)$.

Then we develop all terms of (4.7). The developed form of the conditional density for the survival part is written by

$$\begin{aligned} p(T_i, \delta_i | b_i; \theta_t, \beta) &= \{ h_i(T_i | \mathcal{N}_i(T_i); \theta_t, \beta) \}^{\delta_i}\, S_i(T_i | \mathcal{N}_i(T_i); \theta_t, \beta) \\ &= \Big[h_0(T_i) \exp\{ \gamma^t w_i + \alpha \eta_i(T_i) \} \Big]^{\delta_i} \\ &\quad \times \exp\Big(- \int_0^{T_i} h_0(s) \exp\{ \gamma^t w_i + \alpha \eta_i(s) \}\, ds \Big) \end{aligned} \tag{4.8}$$

The densities of the outcomes $y_i(t_{ij}) = y_{ij}$ are assumed to belong to the exponential family of distributions and so we have that

$$p(y_{ij} | b_i; \theta_y) = \exp\Big\{ \frac{y_{ij}\, \psi_{ij}(b_i) - c\{\psi_{ij}(b_i)\}}{\phi} - d(y_{ij}, \phi) \Big\} \tag{4.9}$$

where $\psi_{ij}(b_i)$ and ϕ are the natural and scale parameters respectively and $c(.)$ and $d(.)$ are known functions that identify the member of the exponential family. Since the random effects follow a normal multivariate distribution, their density is given by

$$p(b_i;\theta) = (2\pi)^{-q/2}\det(D)^{-1/2}\exp(-b_i^t\, D^{-1}b_i/2) \tag{4.10}$$

where q denotes the dimension of the random-effects vector.

Maximum likelihood estimators can be obtained either by directly maximizing the log-likelihood $l(\theta)$ or by using an algorithm that maximizes an approximated log-likelihood. This algorithm refers to the Expectation-Maximization (EM) algorithm. Maximization of the log-likelihood $l(\theta)$ with respect to θ is a computationally intensive task. Indeed, the integral with respect to the random effects in (4.7), and the integral include in the survival function in (4.8) do not have an analytical solution in general and they need to be approximated.

The integral in (4.8) can be approximated using the 7-point or 15-point Gauss-Kronrod rule. However, the integral with respect to the random effects is more complex to approximate. When this integration is performed using Gaussian quadrature rules (e.g. Gauss-Hermite rule), the computational complexity increases exponentially with the dimension of the random-effects vector. In order to decrease this computational burden, Rizopoulos (2012a) proposed to use a pseudo-adaptive Gauss-Hermite quadrature. The idea behind this quadrature is to re-center and re-scale the integrand for each subject using information from a separate fit of the generalized linear mixed effects models for the longitudinal outcome. This method requires much less quadrature points than the standard Gauss-Hermite rule and is therefore faster.

EM algorithm

The EM algorithm is widely used for likelihood inference of joint models. It is known for its numerical stability and to be robust to anomalous individual behaviour and to model misspecification. Moreover, we notice that a direct maximization of the observed data log-likelihood $l(\theta)$ by calculating the score vector requires similar computations to the E-step of the EM algorithm; the same integrals with respects to the random effects are calculated. For this reason, even if the EM algorithm was used to estimate the joint model, the standard errors can be directly estimated by calculating the information matrix from the observed data score vector (Rizopoulos et al. ,2009 and Viviani et al. , 2014).

The aim of the EM algorithm is to find the parameter values $\hat{\theta}$ that maximize the observed data log-likelihood $l(\theta)$. and is based on the idea that the log-likelihood corresponding to the complete data is much simpler to maximize that the observed log-likelihood. For each iteration, this algorithm alternates between executing an expectation step, which creates a function for the expectation of the complete log-likelihood,

and a maximization step, which maximizes the function created in the E-step.

In the joint modelling framework, the observed information is $\{T_i, \delta_i, y_i\}$ and to apply the EM algorithm, the random effects b_i are treated as 'missing data'. The values of the random effect are latent information. The full parameter vector is $\theta = (\theta_t^t, \theta_y^t, \theta_b^t)^t$ where $\theta_t^t = (\gamma^t, \alpha, \theta_{h0}^t)$ with θ_{h0} denoting the parameters in the $h_0(.)$ function, $\theta_y = (\beta^t, \sigma^2)^t$ and $\theta_b^t = \text{vech}(D)$.

In the E-step, the expected value of the complete log-likelihood at the i-th iteration is given by

$$\begin{aligned} Q(\theta|\theta^{(i)}) &= E\{\log p(T_i, \delta_i, y_i, b_i; \theta)|T_i, \delta_i, y_i; \theta^{(i)}\} \\ &= \sum_i \int \log p(T_i, \delta_i, y_i, b_i; \theta)\, p(b_i \mid T_i, \delta_i, y_i; \theta^{(i)})\, dbi \\ &= \sum_i \int \Big\{ \log p(T_i, \delta_i|b_i; \theta_t, \beta) + \log p(y_i|b_i; \theta_y) \\ &\qquad + \log p(b_i|\theta_b) \Big\}\, p(b_i|T_i, \delta_i, y_i; \theta^{(i)})\, db_i. \end{aligned} \tag{4.11}$$

As mentioned above, the integral with respect to the random effects and the integral involved in term $p(T_i, \delta_i|b_i; \theta_t, \beta)$ do not have closed-form solutions and need to be solved numerically.

During the M-step, this function is maximized in order to obtain the updated parameters by

$$\theta^{(i+1)} = \arg\max_{\theta} Q(\theta|\theta^{(i)}). \tag{4.12}$$

The algorithm iterates between the expectation and the maximization steps until the parameter estimates converge and then we obtain the maximum likelihood estimator of $\hat{\theta}$ of θ . However, a drawback of the EM algorithm is its slow convergence, especially near the maximum point.

4.3.2 Bayesian Estimation

We now propose a parameter estimation performed under the Bayesian framework. In this framework, both θ and the random effects $\{b_i, i = 1, \ldots, n$ are regarded as model parameters. All parameters of the model are assumed to be random variables with a certain prior distribution. We want to obtain information of the probability distribution of these parameters based on the observed data.

Bayesian estimation of joint model's parameters is based on the posterior distribution. The expression for the posterior distribution of the model parameters is derived under the assumptions defined in (4.6). We obtain,

$$\begin{aligned} p(\theta, b \,|\, T, \delta, y) &= \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{p(y_{ij}|\, b_i; \theta)\; p(T_i, \delta_i|\, b_i; \theta)\; p(b_i; \theta) p(\theta)}{p(T_i, \delta_i, y_{ij})} \\ &\propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} p(y_{ij}\,|b_i\,; \theta)\; p(T_i, \delta_i\,|\, b_i; \theta)\; p(b_i; \theta)\; p(\theta) \end{aligned} \tag{4.13}$$

where θ denotes the full parameter vector and $p(\theta)$ is the prior distribution for parameters in θ. The expressions of $p(y_{ij}\,|b_i\,; \theta)$, $p(T_i, \delta_i\,|\, b_i; \theta)$ and $p(b_i; \theta)$ are respectively given by (4.8), (4.9) and (4.10).

Markov chain Monte Carlo (MCMC) algorithms can be used to draw samples from the posterior distribution $p(\theta, b \,|\, T, \delta, y)$. With a sufficient number of samples, inferences can be made for all models parameters, conditional on the observed data. Indeed, the posterior means and variances of the parameters can be estimated based on these samples. Two widely used MCMC methods are the Gibbs sampler and the Metropolis algorithm. A number of early publications on joint modelling used the Gibbs sampler as the method of estimation (e.g. Faucett and Thomas (1996); Brown and Ibrahim (2003)). The R package `JMBayes` implemented in this thesis uses a random walk Metropolis algorithm to sample from the posterior conditional distributions of the model parameters. (Rizopoulos, 2016).

For the purpose of clarity, we denote all the parameters of interest of the joint model by the vector $\Theta = (\theta, b)^t$ and we denote the information from the n subjects by $\mathcal{D}_n = \{(T_i, \delta_i, y_i), i = 1, \ldots, n\}$. The MCMC algorithms allow to generate samples $\{\Theta^{(m)} : m = 1, \ldots, M\}$ from the joint posterior distribution.

The Metropolis algorithm begin by choosing a starting vector $\Theta^{(0)}$ from an arbitrary distribution. For each iteration, $t = 1, 2, \ldots$ the algorithm propose a new vector Θ_{prop}

from a symmetric proposal distribution. An usual proposal distribution choice is a multivariate Gaussian distribution which is centered on the previous accepted vector , $\Theta^{(t-1)}$. Next, the algorithm computes an appropriate rule to determine whether Θ_{prop} is accepted or rejected. We define

$$\text{prob} = \min\left\{1, \frac{p(\Theta_{prop}|\mathcal{D}_n)}{p(\Theta^{t-1}|\mathcal{D}_n)}\right\}.$$

The proposal is accepted with the probability prob and so, $\Theta^{(t)} = \Theta_{prop}$. Otherwise, $\Theta^{(t)} = \Theta^{(t-1)}$, the previous sample is kept.
We use standard prior distributions for the parameters θ. In details, independent univariate diffuse normal priors are used for each of β, γ, α and γ_{h_0}, the vector of spline coefficients for the baseline hazard. We assume an inverse Wishart prior for the covariance matrix of the random effects.

The deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) can be used to assess the goodness-of-fit measure in order to perform a comparison of joint model. The DIC is defined by

$$\text{DIC} = D(\bar{\theta}, \bar{b_i}) + 2p_D$$

where $D(\theta, b_i) = -2\sum_{i=1}^{n} \log p(\mathcal{D}_n|b_i, \theta)$ is the Bayesian deviance. The penalty term $p_D = \overline{D(\theta, b_i)} - D(\bar{\theta}, \bar{b_i})$ is the effective number of parameters and is calculated as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior means of the parameters. This criterion can be reformulated as $\text{DIC} = \overline{D(\theta, b_i)} + p_D$. It takes into account the adequacy of the model and the number of parameters required (Piulachs et al., 2017). The idea is that models with smaller DIC should be preferred to models with larger DIC.

4.4 Association structures

The standard joint model introduced in the Sections 4.1 and 4.2 can be written as

$$\begin{cases} h_i(t \,|\, \mathcal{N}_i(t), w_i) & = h_0(t) \exp\left[\gamma^t w_i + \alpha\, \eta_i(t)\right] \\ g\big[E(y_i(t) \,|\, b_i)\big] & = \eta_i(t) \\ & = x_i^t(t)\beta + z_i^t(t) b_i \end{cases}$$

where $\mathcal{N}_i(t) = \{\eta_i(s), 0, \leq s < t\}$. This model assumes that the hazard for an event at a given time t depends on the current underlying value of the longitudinal response at the same time point, given by $\eta_i(t)$. The parameter α provides a measure of the strength of the association between η_i and the time-to-event process at the same time point. This parameterization is very intuitive and easy to interpret but it is not always the most appropriate according to the data that we want to model. Thus, other types of association have been introduced in the literature (Papageorgiou et al., 2019; Rizopoulos, 2012; Gould et al., 2015).

In this section, we present some alternative parameterizations to associate the longitudinal and survival processes. The writing of the hazard model including a general formulation of association structure is given by

$$h_i(t) = h_0(t) \exp\left[\gamma^t w_i + f\big\{\mathcal{N}_i(t), b_i, \alpha\big\}\right].$$

Time-dependent slopes parameterization

In some situation, the rate of change of the longitudinal response at the time t may be associated with the risk of event at the same time. We therefore introduce a parameterization which postulates that the risk depends both on the current value of the true longitudinal response and on its slope.The hazard function becomes

$$h_i(t) = h_0(t) \exp\left[\gamma^t w_i + \alpha_1 \eta_i(t) + \alpha_2 \eta_i'(t)\right],$$

where

$$\eta_i'(t) = \frac{\mathrm{d}\{x_i^t(t)\beta + z_i^t(t) b_i\}}{\mathrm{dt}}.$$

The parameters α_1 and α_2 respectively measure the association between the value and the rate of change of the true longitudinal trajectory and the time-to-event process and the same time. For individuals having the same level of $\eta_i(t)$, the hazard ratio for one-unit increase of the current rate of change $\eta_i'(t)$ is equals to $\exp(\alpha_2)$. This association structure makes it possible to distinguish two subjects who have the same true response values at a given time, but who differ in the rate of change of the marker.

Cumulative effects parameterization

The two parameterizations we have seen suppose that the risk for an event and the true response values depends only at the same single time point. However, in some situations, it could be interesting to associate the risk of an event with the whole history of the longitudinal response. To do this, the integral of the longitudinal trajectory from time zero up to time t is included in the linear predictor of the hazard model.

$$h_i(t) = h_0(t) \exp\left[\gamma^t w_i + \alpha \int_0^t \eta_i(s)\, \mathrm{d}s\right].$$

Parameter α measures the association between the risk for an event at t and the area under the longitudinal trajectory until the same time t. This structure of association makes it possible to analyse the cumulative effect of the longitudinal response on the hazard fon an event.

Random-effects parameterization

The last type of association structure we consider connects the time-to-event and the longitudinal submodels only through random effects. Indeed, the random effects are included in the linear predictor of the survival submodel,

$$h_i(t) = h_0(t) \exp\left[\gamma^t w_i + \alpha^t b_i\right],$$

where α is a vector whose size is equal to the number of random effects introduced. Each element of α measures the association between the associated random effect and the risk of an event. It is very relevant to use this parameterization when a simple random-intercepts and random-slopes structure is assumed for the longitudinal submodel. In this case, the association structure suggests that the values of the subjects-specific deviations from the average intercept and average slope are associated to the risk that the subject experience the event.
On the other hand, when the longitudinal submodels have a more elaborate structure of random effects (for example, use of splines or polynomials), their interpretations are much more complicated. The association parameters α are also more difficult to interpret. This parameterization is therefore not advantageous in this case.

5 Data application

In this section, we apply the previously introduced methods to build and fit the joint on clinical data. After a description and an exploratory analysis of the data, we will applied joint modelling methods on survival and longitudinal data simultaneously. Initially, the longitudinal data studied will be binary data and then we will study longitudinal data following a Poisson distribution.

5.1 Intensive care data description

The medical description of the data is based on the work of Duc Nam Nguyen, a former student in biostatistics during the academic year 2017-2018.

The data analysed on this work contains medical information on 248 critically ill patients admitted to intensive care units (ICU). One of the most common clinical features in intensive care unit patients is the development of delirium. A delirium is a brain dysfunction that is sudden, fluctuating and usually reversible. It is characterized by inability to pay attention, disorientation, anxiety and fluctuations in consciousness. Up to 70% of critically ill patients admitted in the ICU may suffer from delirium. It can be diagnosed by a cognitive evaluation that should be regularly performed. The development of delirium can be induced by predisposal factors such as that age or previous brain disorder. These types of factors make the brain more vulnerable to precipitating factors. Triggers for the development of delirium may include prolonged hypotension or the administration of excessive sedation. The hypotension corresponds to a prolonged low arterial blood pressure and is known to be a common risk factor of brain dysfunction. Indeed, persistent low arterial blood pressure can induce a shortening of oxygen delivery to organs including the brain.

However, the impact of hypotension on delirium is still unclear and will therefore be investigated in this work. We want to assess whether the occurrence of diastolic hypotension episodes is associated with the development of delirium. An episode of diastolic hypotension occurs when a patient has a diastolic blood pressure (DBP) below 50mmHg. The DBP is a component of the arterial blood pressure.

More precisely, the main interest of the analysis is first to investigate whether the presence of at least one episode of diastolic hypotension in a day could contribute to the development of delirium in patients admitted to the intensive care unit. Then, the second interest is to investigate whether the number of diastolic hypotension episode in a day is associated the to development of delirium.

The values of the longitudinal responses are recorded at the end of the day for each patient. The variable `Timevisit` contains these time points. One unit of this variable correspond the one day. We also include in the analyze the possible following risk factors of the development of delirium:

- **Age of the patient**:

  Delirium may occurs at any age but is more common among older. They are susceptible to have more fragile brain function.

  `age` is a quantitative variable of the age of the patient.

- **Gender of the patient**:

  $\texttt{gender} = \begin{cases} \text{M} & \text{if the patient is a man} \\ \text{F} & \text{if the patient is a woman} \end{cases}$

- **Duration of the sedation** :

  The prolonged used of sedatives could disturb the brain function.

  `lengthsed` is a quantitative variable of the duration of the sedation in days.

- **Apache III**:

  Clinical score is use to assess the severity of an illness.

  `apache` is a quantitative variable of the score.

- **History of brain dysfunction**:

  $\texttt{histneuro} = \begin{cases} 0 & \text{if the patient has never had a brain dysfunction before} \\ 1 & \text{if the patient has had a previous brain dysfunction} \end{cases}$

- **History of arterial hypertension:**

  Patients who had history of arterial hypertension may have high baseline arterial blood pressure and may be more sensitive to hypotension than others.

  $\texttt{ath} = \begin{cases} 0 & \text{if the patient has never had arterial hypertension} \\ 1 & \text{if the patient has had a previous arterial hypertension} \end{cases}$

5.2 Exploratory analysing

The study contains the information of 248 patients admitted to the ICU. Among the 248 patients analysed, 138 (56%) developed delirium during the ICU stay.

There are 167 (67%) men against 81 (33%) women in the data. In addition, 103 (61%) men and 35 (43%) woman developed delirium. The median age of the patients is equal to 70 years old (IQR: 60 - 78). In the Figure 1, we can see that the median age of the group of patients who developed delirium is higher than the median age of those who did not.

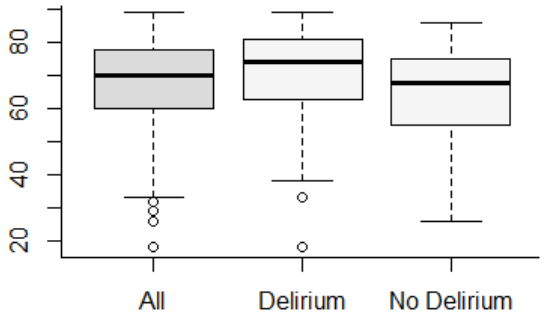

Figure 1: Age distribution for all patients, for patients that developed delirium and for patients without delirium.

In Figure 2, we see that the Apache III scores of the two groups of patients are similar. The median value of the score is 76 for patients with delirium and 75 for those without. For the duration of the sedation shown in Figure 3, no major difference is observed between the two groups but the duration values for the patients who developed delirium are slightly higher.

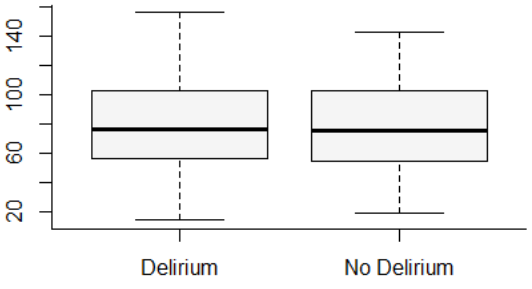

Figure 2: Apache III score

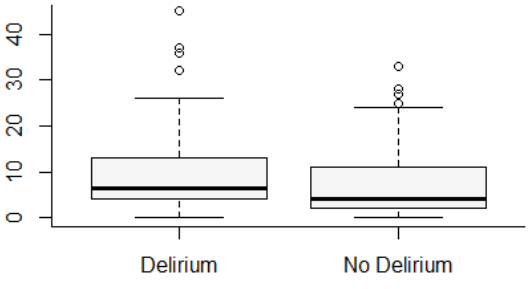

Figure 3: Duration of the sedation

There are 63 (0.25%) patients who have ever had brain dysfunction and 57% of them have developed delirium. Moreover, there are 144 (0.58%) patients who had history of arterial hypertension and 57% of them have developed delirium.

We now describe the longitudinal responses under the two aspect that we will use in the following analyses. First, a value of the longitudinal count outcome represent the number of diastolic hypotension episode occur for a subject in a certain day. The evolution of the number of hypotension episodes per day can be seen in Figure 4. The boxplots for each day are built on the same number of patients. The median number of episodes slowly decreases and becomes zero from Day 3.

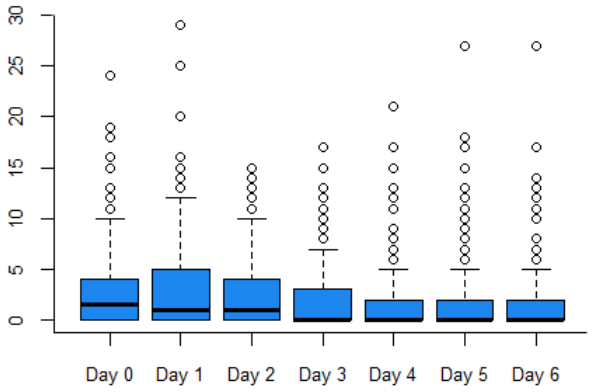

Figure 4: Evolution of the number of hypotension episodes per day, for all patients

Then, a value of the longitudinal binary outcome is equal to one if a subject having at least one hypotension episode during a certain day, otherwise the value is equal to zero for this day. Figure 5 represents for each day the number of patients having at least one hypotension episode. We obtain that 66.5% of the patients have at least one episode during their Day 0, the day of their admission in the ICU. This percentage decreases and from day 4, less than half of patients have at least one episode of hypotension during the day.

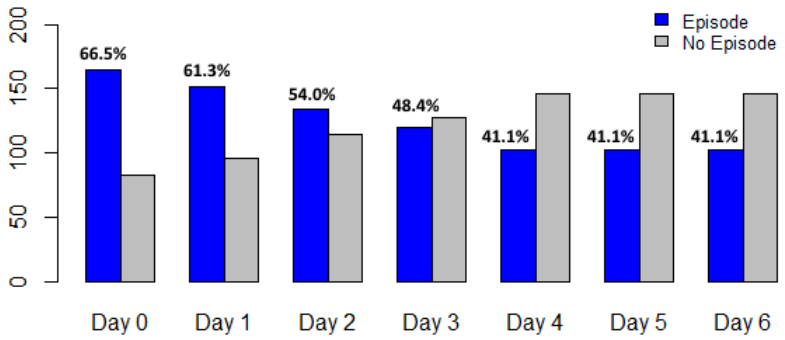

Figure 5: Evolution of the number of patients having at least one hypotension episode during the day, for all patients

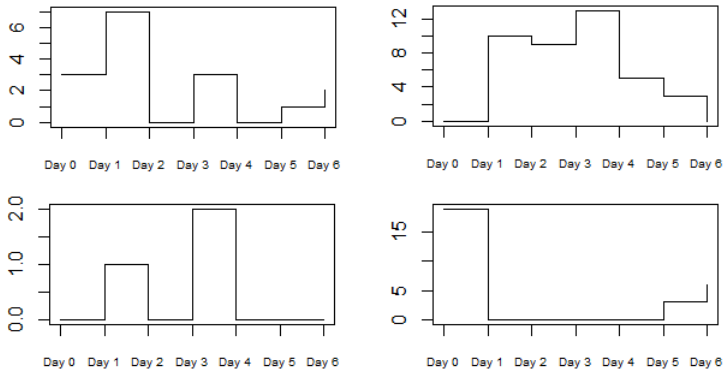

Figure 6: Evolution of longitudinal count values for four patients

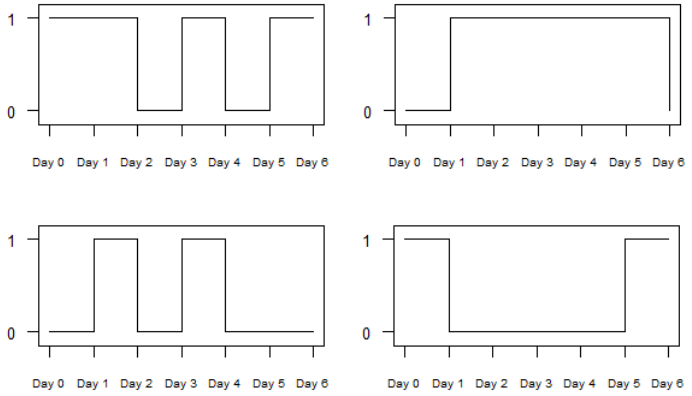

Figure 7: Evolution of longitudinal binary values for four patients

Figure 6 represents the evolution of the longitudinal count responses for four patients chosen at random. The patients shown in the two graphs above both develop delirium during their stay in ICU. This is not the case for the patients represented two graphs below. Figure 7 represents the evolution of the longitudinal binary responses for the same four patients. We notice that for some patients (eg lower left), the response value changes almost every day while for other patients (eg lower right), it remains constant for several days.

5.3 Joint modelling in R

The joint modeling is carried out with the software R and we use the `R` package **JMbayes** in order to fit joint models under a Bayesian approach using Markov chain Monte Carlo algorithms (Rizopoulos D., 2016).
An advantage of this package is that it makes it possible to fit both joint models for continuous or categorical longitudinal responses. To be able to fit the joint model using **JMbayes** package, we must first build two sub-models whose information will be used as main arguments to the basic function of the package, called `jointModelBayes()`. Indeed, this function accepts as main arguments a generalized linear mixed effects object as returned by function `glmmPQL` of package **MASS** (Venables and Ripley 2002), and a survival object as returned by function `coxph()` from package **survival** (Therneau and Lumley 2016).

5.4 Joint modelling for binary responses

Longitudinal model selection.

As required by the **JMbayes** package, we have to fit the submodel with the `glmmPQL` function for the longitudinal part of the data. This function fit a generalised linear mixed effect model using a penalized Quasi-Likelihood estimation method as described in Section (2.3). In the `glmmPQL` function, we need to specify the argument `family = binomial` to take into account that the longitudinal responses are Bernoulli distributed. Since the longitudinal response is binary and the repeated measurements for the same individual are generally correlated, we use a generalized linear mixed model to describe the evolution of the occurrence of episodes of diastolic hypotension in patients admitted to intensive care.

Let y_{ij} be the binary responses indicating the presence of a diastolic hypotension episode for patient i at the day j. The association between repeated measurements will be modelled through inclusion of random terms. A simple and widely used setting is the random-intercepts model. Conditional on random-effects, we assume that Y follows a Bernoulli distribution with success probability $P(Y = 1|b) = \pi$ and $E(Y) = \pi$, this model is given by,

$$\begin{cases} \text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 t_{ij} + b_{i0} \\ \\ b_{i0} \sim N(0, \sigma_b^2), \end{cases}$$

where b_{i0} is the random intercept for the subject i and is assumed to be normally distributed with mean zero and variance σ_b^2. We denote by t_{ij} the time (in days) at which the j-th measurement of patient i is taken. This model means that in the logistic scale, the probability π_{ij} of having a hypotension episode at a certain time differ at baseline for each patient but all patients have the same evolution in time. The parameters β_0 and β_1 are the fixed-effects regression coefficients and describe the mean evolution of the log odds in the population.

The results given in Table 1 show that $\beta_1 = -0.254$. That means that, if we compare the patient i at day d and the patient l at day $d + 1$ such that they have the same random-effect value, the odds ratio for unit (a day) increase of `TimeVisit` is equal to

$$\frac{\pi_{lj}/(1 - \pi_{lj})}{\pi_{ij}/(1 - \pi_{ij})} = \exp(\beta_1) = 0.776$$

where the patient i and l have the same random-effect value.

Table 1: Results of the random-intercepts model

Random effects : ID

Fixed effects	**Estimation (se)**	**p-v**	$\sigma^2_{b_0}$	**Residual**
β_0 `(Intercept)`	0.762 (0.131)	0	1.89	0.896
β_1 `(TimeVisit)`	-0.254 (0.029)	0		

It is possible to extend the random-intercepts model by allowing patient-specific random slopes. This model is called a random-intercepts and random-slopes model and is given by

$$\begin{cases} \text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} \\ b_i \sim N(0, D), \end{cases}$$

The rate of change in the log odds of having a hypotension episode is now different from patient to patient. The parameters estimates for the this model are presented in the Table 2.

The odds ratio defined above becomes equal to $\exp(\beta_1) = 0.750$ and is therefore less high than the odds ratio corresponding to the model in which we did not include random slopes. Moreover, we can observe in matrix D that this model offers more variability between patients in the baseline levels ($\sigma^2_{b_0} = 2.610$) of the log odds than in the evolutions in time ($\sigma^2_{b_1} = 0.159$).

Table 2: Results of the random-intercepts and random-slopes model

Random effects :ID, ID*TimeVisit

Fixed effects	**Estimation (SE)**	**p-v**	**D matrix**	**Residual**
β_0 `(Intercept)`	0.814 (0.139)	0	$\begin{pmatrix} 2.610 & -0.268 \\ -0.268 & 0.159 \end{pmatrix}$	0.825
β_1 `(TimeVisit)`	-0.288 (0.041)	0		

Time-to-event model selection.

We fit a Cox proportional hazards (PH) model on data from intensive care patients to describe the factors that affect survival and measure their impacts. The event of interest is the appearance of delirium in patients and the factors studied are age (`age`), sex (`gender`), duration of sedation (`lengthsed`), apache score (`apache`), history of neurological diseases (`histneuro`) and history of hypertension (`aht`). We use the function `coxph` from the **survival** package. The function fit a Cox proportional hazards regression model as described in the Section 3.2. Since the events are observed on a discrete scale, we use the partial likelihood adjusted by Breslow to take into account the probable tied event times. We specify the argument `ties= "breslow"`. We also need to specify `x = TRUE` to include in the returned object the design matrix of the model.

We first construct a Cox PH model (`ModA`) which assumes that all of these covariates have a multiplicative effect on the hazard of delirium. The results are displayed in Table 3. The column `Coeff` contains the estimated regression coefficients $\hat{\gamma}$. The hazard ratios (HR) are equal to the exponential of the coefficients and give the effect size of the covariates. The column marked `z` gives the Wald statistic value which evaluates whether the coefficient of a given variable is significantly different from 0. We see that only the factors `age` and `gender` are significant. The other four factors do not seem to have a significant effect on survival, their estimated coefficient is very low. For example, among these four covariates, the least non-significant is the covariate `apache`. Its hazard ratio is equal to HR = 0.996, with a 95% confidence interval of 0.99 to 1.01. It means that, holding the other covariates constant, an additional unit of score apache reduce the daily hazard of delirium by about 0.4%, which is not a significant contribution.

We therefore fit a new Cox PH model (`ModB`) containing only the covariates `age` and `gender`. For the subject i, this model is written as

$$h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i\}$$

where `genderM` is the indicator variable for the male gender. The associated survival curve is shown in the Figure 8. This curve is produced for a pseudo-subject with other covariates fixed to their average values (if they are continuous variables) or to their lowest level (if they are discrete variables). The parameter estimates obtained are displayed in Table 4 and the two covariates remain significant. The association between the age and the risk of delirium is expressed by the fact that an additional year of age of the patient corresponds to a 1.03-fold increase in the risk of delirum. In addition, males have 64% higher risk of delirium compared to females. This result is illustrated

Table 3: Parameter estimates from the Cox PH model - `ModA`

	coef	HR (95% CI)	se(coef)	z	P
`age`	0.026	1.026 (1.010-1.043)	0.008	3.217	0.001
`genderM`	0.503	1.654 (1.136-2.467)	0.196	2.569	0.009
`lengthsed`	-0.008	0.992 (0.970-1.017)	0.012	-0.680	0.571
`apache`	-0.004	0.996 (0.990-1.002)	0.003	-1.389	0.171
`histneuro1`	-0.044	0.956 (0.638-1.380)	0.195	-0.228	0.746
`aht1`	0.096	1.101 (0.786-1.585)	0.178	0.541	0.537

in Figure 9 which shows the predicted survival proportion at any given point in time for males and females, given that the `age` covariate are fixed to its average value. The female survival curve is always above that of men.

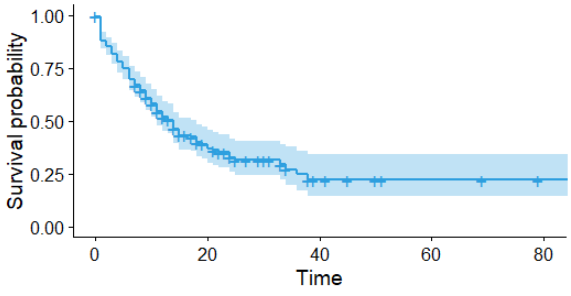

Figure 8

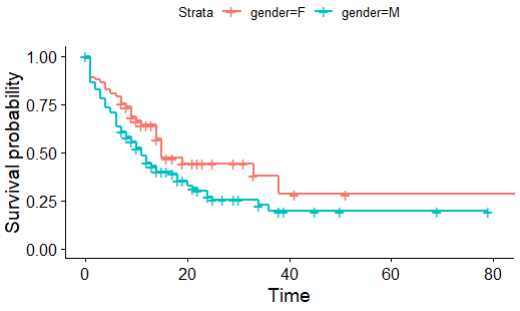

Figure 9

To detect possible violations of the PH assumption, we can use the scaled Schoenfeld residuals to test the assumption both graphically and via a formal statistical test. More information on the equation of these residuals is given in Therneau and Grambsch (2000). A plot that shows a non-random pattern against time is evidence of violation of the PH assumption. Moreover, the null hypothesis of the formal test supposes that the correlation between the scaled Schoenfeld residuals and time is zero. The figure 10 shows the graphs of the scaled Schoenfeld residuals for the covariate `age` and `gender`. The red dots represent the residuals, the solid line is a smooth fit and the dashed lines are the confidence bands at two standard errors. For both covariates, it seems that there is no pattern over time. We also notice that the test for both covariates is not statistically significant, and the global test is also not significant. Therefore,

Table 4: Parameter estimates from the Cox PH model - `ModB`

	coef	exp(coef)	se(coef)	z	P
γ_1 `(age)`	0.026	1.027	0.007	3.537	0.001
γ_2 `(genderM)`	0.495	1.640	0.194	2.460	0.014

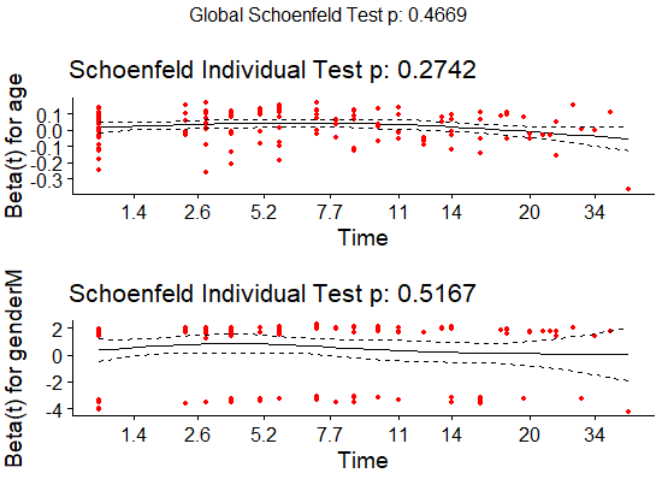

Figure 10: Scaled Schoenfeld residuals against time for `ModB`

the proportional hazards assumption seems to be supported for covariates `age` and `gender`.

Joint model.

Based on the generalized linear mixed-effects and time-to-event models fitted in the previous sections, we can now introduce the joint models for binary longitudinal and survival outcomes.

First, we recall the general definition of a joint model for longitudinal and survival outcomes,

$$\begin{cases} h_i(t) &= h_0(t)\exp\left[\gamma^T w_i + f\{\mathcal{N}_i(t), b_i, \alpha\}\right] \\ \eta_i(t) &= g\big[E(y_{ij}|b_i)\big] \\ &= x_i^T(t)\beta + z_i^T(t)b_i \end{cases}$$

where $g\big[E(y_{ij}|b_i)\big] = \text{logit}(\pi_{ij})$ and π_{ij} is the probability for the patient i of having an episode of diastolic hypotension at time j.

As introduced in Section 4.4, there are several parameterizations allowing to analyse in several different ways the association between the two sub-models. We discuss two frequently used association functions $f(.)$ and compare the different joint models obtained.

The most standard approach to connect the longitudinal and time-to-event submodels is the "current value" association. This parameterization assume that the hazard for an event at a given time t is associated with the current estimated value of the longitudinal response at the same time. For our data, the survival submodel with this parameterization is written as

$$h_i(t) = h_0(t)\exp\big(\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha\eta_i(t)\big),$$

with $h_0(t)$ denoting the baseline hazard function. The subject-specific linear predictor $\eta_i(t)$ is equal to $\text{logit}(\pi_{ij}) = \log(\pi_{ij}/(1-\pi_{ij}))$ and represents the log of the odds of having an episode of hypotension. The constant parameter α quantifies the strength of association. In particular, the quantity $\exp(\alpha)$ returns the hazard ratio for one-unit increase in the value $\log(\pi_{ij}/(1-\pi_{ij}))$ at time t.

The second type of parameterization that we discuss is called the "Random-effects" parameterization. This method includes in the linear predictor of the hazard model only the random effects of the longitudinal submodel. This type of association is meaningful when a simple random-intercepts and random-slopes structure is assumed for the longitudinal submodel. Assuming that the longitudinal submodel contains two

vectors of random effects (intercepts and slopes), the related survival submodel can be written as

$$h_i(t) = h_0(t) \exp\big(\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1}\big),$$

where α_1 and α_2 are the parameters of association of b_{i0} and b_{i1} which are defined in as the individual deviations from the sample average intercept and slope values, respectively.

We therefore have two ways to construct the survival submodel depending on these two parameterizations. Moreover, we also have two ways to construct the longitudinal submodel, namely the random-intercepts model and the random-intercepts and random-slopes model. So, we can fit four joint models defined by:

$$\text{Model A} = \begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha \eta_i(t)\} \\ \eta_i(t) = \beta_0 + \beta_1 t + b_{i0} \end{cases}$$

$$\text{Model B} = \begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha \eta_i(t)\} \\ \eta_i(t) = \beta_0 + \beta_1 t + b_{i0} + b_{i1} t \end{cases}$$

$$\text{Model C} = \begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha_1 b_{i0}\} \\ \eta_i(t) = \beta_0 + \beta_1 t + b_{i0} \end{cases}$$

$$\text{Model D} = \begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1}\} \\ \eta_i(t) = \beta_0 + \beta_1 t + b_{i0} + b_{i1} t \end{cases}$$

where the baseline hazard function $h_0(t)$ is no longer left unspecified at this stage of the construction of joint models. As specified in the Section 4.1, the baseline hazard is approximated using penalized B-splines. This method correspond to the default value of the argument `baseHaz` in the function `jointmodelBayes()`. The type of parameterization used is by default the "current value" parameterization. In order to use the "random-effects" parameterization, we need to specify the argument `param = "shared-RE"`. Finally, we specify the argument `densLong`. It depends on the type of the longitudinal responses by accepting a function that calculates the probability density function of the longitudinal outcome.

These four joint models are fitted using Markov chain Monte Carlo algorithms. The posterior summaries of the parameters can be find in Table 5 for the two models with the current value parameterization and in Table 6 for the two models associated with the random-effects parameterization. Mean and 95% credible interval are sampled for

each parameter from the corresponding posterior distribution.

Before interpreting the results, we investigate the convergence of the chains of the MCMC sampling. A common graphical convergence diagnostic method is the trace plot. It is a time series plot that shows the realizations of the Markov chain at each iteration against the iteration numbers. The method allows to visualize how well the chain is mixed. Across the iterations, the estimated values of the parameters should move quickly but remain within the ranges and should not suddenly go beyond the range. The trace plots for the parameters of the fitted joint model are shown in Figure 11 for the joint model C and in Figure 12 for the joint model D.

We can now interpret and compare the results obtained for the four joint models. For the model A and B, the estimate parameter α measures the strength of the association between the hazard of delirium at time t and the current underlying value of a function of the longitudinal response at the same time point, namely the log odds of having an episode of hypotension. The model A suggests that each unit increase in the current value of $\log(\pi_{ij}/(1-\pi_{ij}))$ involves a $\exp(\alpha_A) = 1.19$ - fold increase (95% CI: 1.04, 1.40) in the subject's delirium hazard, whereas this association parameter leads to a $\exp(\alpha_B) = 1.08$ - fold increase (95% CI: 1.03, 1.15) if we assume a random-intercepts and random-slopes longitudinal submodel (Model B). So when $\text{logit}(\pi_{ij})$ increases, the hazard of having a delirium increase at the same time. Furthermore, since the logit(.) function is strictly increasing, the hazard of having a delirium increases when the probability π_{ij} increase.

The posterior means obtained for all parameters of the Model C are almost similar to those obtained for Model A. Indeed, the model C assumes the random effects parameterization and includes only one vector of random effects (random-intercepts) in the longitudinal submodel. Only one vector of random effects is include in the linear term of the survival term. Therefore, the only difference in construction between these two models is that, unlike Model A, Model C does not include the fixed-effects terms of the longitudinal submodel in the linear predictor of the survival submodel.

In the model D, the parameters α_1 and α_2 define the strength of the association between the random intercept and the random slope effect, respectively. But, the 95% credible interval for the association parameter α_1 contain zero. That means that, the subject-specific deviations from the average sample intercept are not predictive of the hazard of delirium. However, the subject-specific deviations from the average sample slope are predictive of the hazard of delirium. For individuals having the same deviation from the average intercept, the hazard ratio for a one unit increase in individual deviation from the average slope is equal to $\exp(\alpha_2) = 2.20$, (95% CI: 1.28, 4.23).

We finally compare statistically the four estimated joint models. We rely on the Deviance Information Criterion (DIC) for which smaller values indicating better model adjustments to the data. The log-pseudo-marginal-likelihood and the DIC values are reported in the Tables 5 and 6. As we said, Model A and Model C give similar results. Their DIC values are very close. Among the four joint models, the DIC value is minimum for the simplest model, model A.

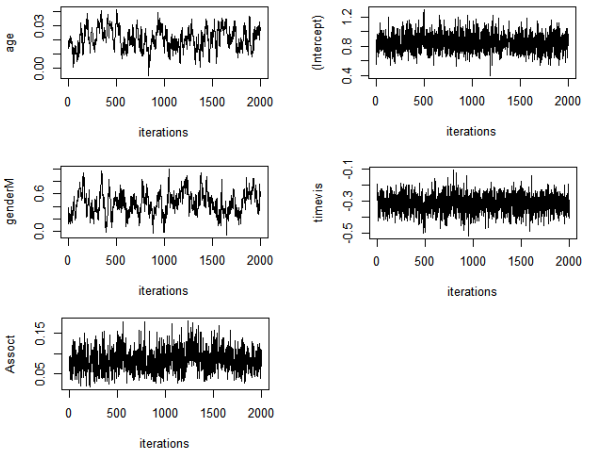

Figure 11: Trace plot for the parameters of the joint model C

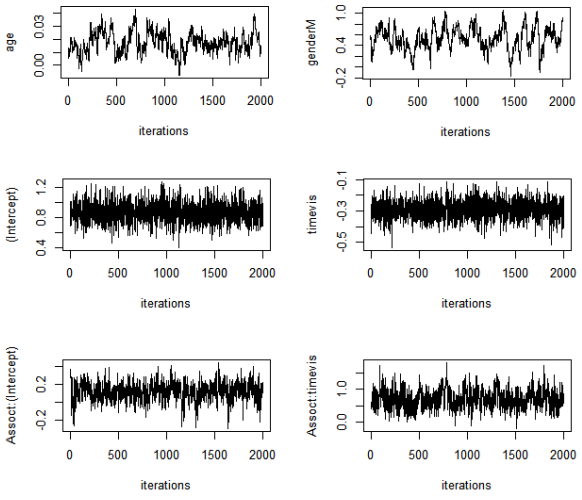

Figure 12: Trace plot for the parameters of the joint model D

Table 5: Parameter estimates and 95% credible intervals under the joint modeling analysis with a current value parameterization

Parameterization: Current value		
Mod.A: *Random-intercepts submodel*		
Parameters	Estimate	95%CI
Longitudinal submodel		
β_0 `(Intercept)`	0.810	[0.564; 1.064]
β_1 `(TimeVisit)`	-0.255	[-0.312; -0.202]
Survival submodel		
γ_1 `(Age)`	0.022	[0.007; 0.038]
γ_2 `(GenderM)`	0.605	[0.198; 0.990]
Association		
α	0.177	[0.037; 0.337]
Goodness of fit		
DIC	3694.8	
LPML	-1918.8	
Variance Components		
$\sigma^2_{b_0}$	2.376	

Mod.B: *Random-intercepts and random-slopes submodel*		
Parameters	Estimate	95%CI
Longitudinal submodel		
β_0 `(Intercept)`	0.845	[0.626; 1.075]
β_1 `(TimeVisit)`	-0.314	[-0.425; -0.213]
Survival submodel		
γ_1 `(Age)`	0.022	[0.006; 0.037]
γ_2 `(GenderM)`	0.517	[0.108; 0.970]
Association		
α	0.082	[0.032; 0.142]
Goodness of fit		
DIC	4193.8	
LPML	-2229.9	
Variance Components		
D matrix	$\begin{pmatrix} 1.963 & -0.108 \\ -0.108 & 0.486 \end{pmatrix}$	

Table 6: Parameter estimates and 95% credible intervals under the joint modeling analysis with a random effects parameterization

Parameterization: Random effects		
Mod.C: *Random-intercepts submodel*		
Parameters	Estimate	95%CI
Longitudinal submodel		
β_0 (Intercept)	0.806	[0.550; 1.058]
β_1 (TimeVisit)	-0.255	[-0.311; -0.201]
Survival submodel		
γ_1 (Age)	0.022	[0.008; 0.038]
γ_2 (GenderM)	0.581	[0.188; 0.938]
Association		
α_1	0.186	[0.027; 0.369]
Goodness of fit		
DIC	3693.3	
LPML	-1924.1	
Variance Components		
$\sigma^2_{b_0}$	2.344	
Mod.D: *Random-intercepts and random-slopes submodel*		
Parameters	Estimate	95%CI
Longitudinal submodel		
β_0 (Intercept)	0.862	[0.640; 1.118]
β_1 (TimeVisit)	-0.279	[-0.388; -0.167]
Survival submodel		
γ_1 (Age)	0.016*	[-0.001; 0.031]
γ_2 (GenderM)	0.567	[0.218; 0.927]
Association		
α_1	0.126*	[-0.157; 0.351]
α_2	0.787	[0.246;1.443]
Goodness of fit		
DIC	4201.3	
LPML	-2235.1	
Variance Components		
D matrix	$\begin{pmatrix} 2.077 & -0.135 \\ -0.135 & 0.516 \end{pmatrix}$	

5.5 Joint modelling for count responses

In this section, we analyse an other type of longitudinal responses, the longitudinal count responses. We now focus on the number of episodes observed per day for each patient and over a 7 day period. The value y_{ij} represents the total number of hypotension episodes the patient i had during the day j. We would like to assess the strength of the association between this hypotension episodes count at day t and the hazard of having a delirium at the same time point. We also compare the extended Cox model and the joint modelling approach.

Time-dependent Cox model.

We suppose that the hypotension episode count is taken as an exogenous time-dependent covariate. We fit the extended Cox model in which we include as time-independent covariates the age and the gender, and as time-dependent the hypotension episodes count. This model is expressed as

$$h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{age}_i + \gamma_2 \texttt{genderM}_i + \alpha y_i(t)\}$$

where $y_i(t)$ denotes the observed level of the number of hypotension episodes at the end of day t. The parameters estimated are shown in Table 7. The number of hypotension episodes does not appear to be significantly associated with the hazard of delirium. Indeed, the 95% confidence interval corresponding to $\exp(\alpha)$ contains 1, (95% CI: 0.911, 1.03).

Table 7: Parameter estimates from the Time-Dependent Cox model

	coef	HR (95% CI)	se(coef)	z	P
γ_1 (age)	0.036	1.037 (1.016-1.058)	0.010	3.588	0.001
γ_2 (genderM)	0.595	1.814 (1.097-2.998)	0.256	2.322	0.020
α	-0.029	0.971 (0.911-1.035)	0.032	-0.913	0.361

Longitudinal model selection.

Conditional on random-effects, we assume that Y follows a Poisson distribution with $E(Y) = \lambda$, the random-intercepts model is given by

$$\begin{cases} \log(\lambda_{ij}) = \beta_0 + \beta_1 t_{ij} + b_{i0} \\ b_{i0} \sim N(0, \sigma_b^2), \end{cases}$$

The notations used are the same as in Section 5.2. This model postulates that the logarithm of the expected number of hypotensive episodes at a certain time differs at baseline for each patient.

We fit this model with the `glmmPQL` function and we need to specify the argument `family = poisson` to take into account that the longitudinal responses are Poisson distributed. The results obtained are shown in Table 8. We see that $\exp(-\beta_1) = 1.13$. That means that an unit increase in time leads to a diminution of 13% for the expected hypotension episode count, for subjects with the same random-effect value.

Table 8: Results of the random-intercepts model

Random effects : ID

Fixed effects	**Estimation (se)**	**p-v**	$\sigma^2_{b_0}$	**Residual**
β_0 `(Intercept)`	0.741 (0.077)	0	0.926	1.486
β_1 `(TimeVisit)`	-0.125 (0.014)	0		

We also fit the random-intercepts and random-slopes model which is defined as

$$\begin{cases} \log(\lambda_{ij}) = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} \\ \\ b_i \sim N(0, D), \end{cases}$$

The parameters estimated of this model are shown in Table 9. This model leads that an unit increase in time (in days) corresponds to a decrease of 24% in the expected number of hypotension episodes, for subjects with the same random-effect value.

Table 9: Results of the random-intercepts and random-slopes model

Random effects : ID, ID*TimeVisit

Fixed effects	**Estimation (SE)**	**p-v**	**D matrix**	**Residual**
β_0 `(Intercept)`	0.832 (0.074)	0	$\begin{pmatrix} 0.884 & -0.021 \\ -0.021 & 0.059 \end{pmatrix}$	1.274
β_1 `(TimeVisit)`	-0.221 (0.024)	0		

Time-to-event model selection.

The survival data submodel is the same as that introduced in the joint model analysis for binary responses. We fit the same survival model which is given by

$$h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{age}_i + \gamma_2 \texttt{genderM}_i\}$$

and thus obtain the same result as in Table 4.

Joint model.

We construct two joint models, one is based on the random-intercepts submodel and the other is based on the random-intercepts and random-slopes submodel. We only consider the "current value" parameterization. We obtain two joint models given by

$$\text{Model E} = \begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha \eta_i(t)\} \\ \eta_i(t) = \beta_0 + \beta_1 t + b_{i0} \end{cases}$$

$$\text{Model F} = \begin{cases} h_i(t) = h_0(t) \exp\{\gamma_1 \texttt{Age}_i + \gamma_2 \texttt{GenderM}_i + \alpha \eta_i(t)\} \\ \eta_i(t) = \beta_0 + \beta_1 t + b_{i0} + b_{i1} t \end{cases}.$$

where $\eta_i(t) = \log(\lambda_{ij})$ represents the logarithm of the expected number of hypotension episodes. The quantity $\exp(\alpha)$ returns the hazard ratio for one-unit increase in the value $\log(\lambda_{ij})$ at time t. The baseline hazard function $h_0(t)$ is again approximated using penalized B-splines. We still use the function `jointmodelBayes()` to adjust these models. We need to modify the argument `densLong` adequately in order to be able to calculate the probability density function of the Poisson longitudinal outcome. The posteriors summaries of the parameters of the joint models with the current value parameterization can be find in Table 11.

We first notice that for Model E, the 95% credible interval for the association parameter α contains zero. This means that this model does not lead to a significant association between the logarithm of the expected number of episodes and the hazard of delirium.

The estimate of the parameter α in Model F is significantly different from zero but has a low value. The negative sign of the association parameter shows that a relatively high number of episodes in a day is instantaneously associated with a decreased risk of delirium. Specifically, this model suggests that each unit increase in the current value of the logarithm of the expected count results in a $\exp(-\alpha) = 1.05$-fold decrease of

the hazard of delirium for a subject.

With respect to DIC values, Model F appears to be better than Model E at fitting a joint model for longitudinal count and survival outcomes. We also see that in both models, the variance components for random intercepts are very high.

Table 10: Parameter estimates and 95% credible intervals under the joint modeling analysis with a current value parameterization - Poisson responses

Parameterization: Current value		
Mod.E *Random-intercepts submodel*		
Parameters	Estimate	95%CI
Longitudinal submodel		
β_0 (`Intercept`)	2.082	[1.631; 2.535]
β_1 (`TimeVisit`)	-0.257	[-0.297; -0.221]
Survival submodel		
γ_1 (`Age`)	0.027	[0.014; 0.041]
γ_2 (`GenderM`)	0.537	[0.197; 0.990]
Association		
α	-0.028*	[-0.105; 0.048]
Goodness of fit		
DIC	7361.2	
LPML	-3955.7	
Variance Components		
$\sigma^2_{b_0}$	10.011	
Mod.F *Random-intercepts and random-slopes submodel*		
Parameters	Estimate	95%CI
Longitudinal submodel		
β_0 (`Intercept`)	1.948	[1.383; 2.512]
β_1 (`TimeVisit`)	-0.386	[-0.522; -0.255]
Survival submodel		
γ_1 (`Age`)	0.028	[0.013; 0.044]
γ_2 (`GenderM`)	0.544	[0.181; 0.970]
Association		
α	-0.053	[-0.083; -0.010]
Goodness of fit		
DIC	5680.7	
LPML	-1694.7	
Variance Components		
D matrix	$\begin{pmatrix} 14.741 & -2.031 \\ -2.031 & 0.821 \end{pmatrix}$	

6 Conclusion

In this thesis we studied the diastolic blood pressure of critically ill patients and their risk of developing a delirium during their stay in the intensive care unit. We defined an episode of hypotension when a patient presents a diastolic blood pressure below 50 mmHg. Patients admitted to ICU are likely to develop cerebral dysfunction and, in particular, delirium. The objective of this work was therefore to use the joint modelling framework to assess whether there is an association between hypotension and the risk of developing delirium.

The primary focus was on analysing the impact of the occurrence of at least one hypotension episode per day. Results showed that the logarithm of the odds of having a hypotension episode decrease over time. Also, among several time-independent factors, only the patients' age and sex have a significant effect on the risk of developing delirium during the ICU stay. Men are at higher risk than women and being older also increases the risk of developing delirium. A joint modelling approach was then used to assess the association between binary longitudinal and survival outcomes.

We examined the association between the two outcomes using current value and random-effects parameterizations. We found that an increase in the current value of the logarithm of the odds of having a hypotension episode leads to a significant increase in the hazard of developing a delirium. Using the random-effects parameterization, we found that subject-specific random-slopes have a significant positive impact on the hazard of delirium.

The secondary interest of the thesis was to consider Poisson distributed longitudinal data. The objective of this analysing was to look at the longitudinal data from another perspective and also to compare the result of an extended Cox model and a joint model. We considered the number of episodes observed per day instead of considering only whether at least one episode occurred on that day or not. We found that the logarithm of the expected number of hypotension episodes decreased over time.
In order to evaluate the association between the longitudinal and the survival outcomes, we first fitted an extended Cox model in which we included the time-dependent covariate of counting hypotension episodes. This model led to an insignificant effect of the logarithm of the number of episodes to the hazard of developing a delirium.

We then implemented a joint model and obtained an association significantly different from zero, but only if we included both random-intercept and random-slopes in the longitudinal submodel. In this case, an increase in the logarithm of the expected hypotension episodes count leads to a decrease to the hazard of developing delirium.

Also, the effect obtained is significant but small.

In conclusion, we discovered that when the focus is on occurrence of at least one episode per day, the association effect is positive on the hazard of developing delirium, whereas when the focus is on the number of episodes per day, this association effect is negative.

Discussion and outlook

In section 4.2, we have assumed that the visiting process which generates the time points at which the longitudinal measurements are collected, is independent to the time at which the event occurs. In some situations, this assumption is not entirely realistic. For example, we consider a group of patients recovering from cancer for whom measurements are taken every three months. If one of these patients' condition worsens, it is necessary to follow this patient more closely and therefore the frequency of his measurements increases.
On the other hand, this hypothesis makes sense for the data studied in this thesis. The patients in intensive care all have advanced health conditions. Blood pressure is continuously monitored by an arterial catheter and its value is recorded every hour for each patient. The visiting process is therefore highly controlled.

A possible extension of this work is the inclusion of other association structures such as parameterizations of cumulative or lagged effects. With the latter case, the hazard of the event occurring at time t is associated with the level of the longitudinal measurement at a earlier time point $t - c$, where c indicates the time lag of interest. The cumulative effects structure allows the hazard to depend on the whole history up to time t of the longitudinal outcome. The use of these parameterizations may explain why we obtained the surprising result that an increase in the logarithm of the number of hypotension episodes leads to a decreased risk of developing delirium. Indeed, considering that only the current value of the time-dependent covariate impacts the current hazard of an event may lead to medically inconsistent results in some situations.

Finally, all joint models presented in this thesis were fitted with the `R` package **JMbayes** under a Basysian approach using Markov chain Monte Carlo algorithms. It is also possible to fit joint models in `SAS`. Garcia-Hernandez and Rizopoulos (2018) proposed `%JM`, a `SAS` macro that fits joint models for longitudinal and survival outcomes. This macro can handle normal, binary, binomial and Poisson longitudinal responses. It also offers several association structures such as those presented in this

thesis. A disadvantage of this macro is that its execution time is greater than that of the `jointModelBayes` function of the **JMbayes** package, especially when the structure of the model becomes complex.

References

Albert, P. and Follmann, D. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, 56, 667 – 677.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10, 1100 – 1120.

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25.

Brown, E. and Ibrahim, J. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59, 221 – 228.

Brown, H. and Prescott, R. (2006). *Applied Mixed Models in Medicine. Second edition.* New-York : John Wiley and Sons.

Elashoff R, Li G and Li N. (2016). *Joint modeling of longitudinal and time-to-event data.* Boca Raton, FL: Chapman and Hall/CRC.

Faucett, C., Schenker, N., and Elashoff, R. (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association,* 93, 427 – 437.

Faucett, C. and Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine,* 15, 1663 – 1685.

Fieuws S and Verbeke G. (2008). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, 9, 419-431.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008). *Longitudinal data analysis.* London : Chapman and Hall.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied Longitudinal Analysis. Second Edition.* Hoboken : John Wiley and Sons.

Garcia-Hernandez, A. and Rizopoulos, D. (2016). %JM: A SAS macro to fit jointly generalized mixed models for longitudinal data and time-to-event responses. *Journal of Statistical Software*

Gould, A. L., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S. and Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, 34(14), 2181–2195.

Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62, 1037 – 1043.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis.* New-York : Springer-Verlag.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data, 2nd edition.* New York : John Wiley and Sons.

Klein, J. and Moeschberger, M. (2003). *Survival Analysis - Techniques for Censored and Truncated Data.* New York : Springer-Verlag.

Lang, S. and Brezger, A. (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York : Springer-Verlag.

Moore, D. F. (2016). *Applied survival analysis using R.* Switzerland: Springer.

Papageorgiou, G., Mauff, K., Tomer, A. and Rizopoulos, D. (2019). An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes. *Annual Review of Statistics and Its Application*, 6(1), 223-240.

Piulachs X., Alemany R., Guillen M. and Rizopoulos D. (2017) Joint models for longitudinal counts and left-truncated time-to-event data with applications to health insurance. *SORT-Stat Oper Res Trans*, 41(2), 347–72.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95, 63 – 74.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data, with Applications in R.* Boca Raton : Chapman & Hall/CRC.

Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56, 491 – 501.

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software*, 72(7), 1–45.

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64, 583–639.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model.* New York : Springer-Verlag.

Therneau T, Lumley T (2016). *survival: Survival Analysis Including Penalised Likelihood.* R package version 2.39-5

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and timeto-event data: An overview. *Statistica Sinica*, 14, 809 – 834.

Tuerlinckx, F., Rijmen, F., Verbeke, G. and De Boeck, P. (2006). Statistical inference in generalized linear mixed models: a review. *Br J Math Stat Psychol*, 59(Pt 2), 225-55.

Venables W, Ripley B (2002). Modern Applied Statistics with S. 4th edition. *Springer-Verlag,* New York.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.

Viviani, S., Alfó, M. and Rizopoulos, D. (2014) Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, 24, 417–427.

Wang, Y. and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96, 895 – 905.

West, B. T., Welch, K. B. and Gałlecki, A. T. (2006). *Linear mixed models. A practical guide using statistical software.* Boca Raton : Chapman Hall/CRC.

Wu L., Liu W., Yi GY., Huang Y. (2012) Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics.*, 640153.

Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 330 – 339.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/sc**