UNIVERSITÁ DEGLI STUDI DI MILANO-BICOCCA
Dottorato di Ricerca in Statistica ed Applicazioni
XXVIII Ciclo

Joint models for time-to-event and multivariate longitudinal data

Tutor: Chiar.ma Prof.ssa Marcella Polisicchio

Ph.D. Thesis by:
Marcella Mazzoleni

Anno Accademico: 2014-2015

Contents

Abstract

Joint models for time-to-event and multivariate longitudinal data The joint models for longitudinal and time-to-event data are a recent family models that jointly analyse the longitudinal and the survival data.
The models are composed by two sub-models, the longitudinal and the survival sub-model. A proportional hazard model can be used for survival sub-model and it is expressed in function of the true and unobserved value of the longitudinal outcome, while concerning the longitudinal sub-model a linear mixed model is often proposed.
After analysing some univariate cases, it is interesting to study the situation in which one of the two sub-models or both are multivariate. Thus different scenarios are possible. Firstly it is possible to consider a situation in which only the longitudinal sub-model is multivariate, in this situation a multivariate linear mixed model or another type of multivariate longitudinal model can be considered. Choosing a multivariate linear mixed-effects model, a different longitudinal outcome must be considered for each linear predictor. Accordingly the survival sub-model is composed by several parameters that express the relation of each true and unobserved value of the longitudinal outcome with the hazard function.
Secondly a situation in which only the survival sub-model is multivariate is possible, thus the survival sub-model may consider two situations, competing risks or recurrent events. Lastly a situation in which both the longitudinal and the survival model are multivariate must be considered. The sub-models are composed by the multivariate longitudinal and survival sub-models which are jointly analysed.
The great problem related to the multivariate situation concerns the computational aspect of the estimation. In fact considering that the univariate case is computational demanding, increasing the number of the parameters or the dimension of the sub-models will lead to higher computational demanding situations. This problem could be solved with the implementation of some algorithms in the R software that could reduce the time and the memory requested.
In this thesis the focus is on the situation in which only the longitudinal sub-model is multivariate. The aim is to find new methods of estimation and some algorithms that could help to solve the problem of the computational aspect. At first a two stages approach is implemented as it permits to obtain very fast and significant estimations.
The most of the applications of joint models focus on the biostatistical area, thus the event analysed is death or the manifestation of a disease and the influence of some biomarkers on it.
In this thesis the focus is on the undergraduates' career, analysing the careers of the undergraduate students in an Italian university, using jointly the time to graduation and the student's path, focusing on the marks and on the number of exams that the student has already passed before a fixed time. The algorithms are implemented also on a well-known biostatistical data set available in the package JM of the software R the test the reliability and the efficiency.
Keywords: Joint Models, Proportional Hazard Models, Linear Mixed Models, Timing of Student Graduation

Chapter 1

Introduction

The joint models for longitudinal and time-to-event data are a recent family of models that jointly analyse longitudinal and survival data. The models are composed by two sub-models, the longitudinal and the survival sub-model.
In the classical joint model a proportional hazard model can be used for survival sub-model and it is expressed as a function of the true and unobserved value of the longitudinal outcome, while concerning the longitudinal sub-model a linear mixed model is often proposed.
These recent models had been analysed from several researchers, then there is a very extensive literature available. Then after analysing in deep the literature of the classical case in which both sub-models are univariate, it is interesting to study the possible extensions of these models. These extensions may concern different formulations of the sub-models, thus different formulation for the longitudinal or for the survival sub-model. Through the use of different formulation of the sub-models, it is possible to deal, for example, with heterogeneity in the sample or relations between the risk of the event and the longitudinal covariates which are not linear.
One of the most interesting extension concerns the situation in which one of the two sub-models or both are multivariate. Thus different scenarios are possible. Firstly a situation in which only the longitudinal sub-model is multivariate can be considered, then more than one longitudinal covariate may influence the risk of the event. Secondly the extension may concern the case in which the survival sub-model is multivariate, then there are more than one event analysed, that can be recurrent or terminal. Lastly a situation in which both sub-models are multivariate can be considered.
In this class of extensions we decided to focus on the situation in which only the longitudinal sub-model is multivariate, as a possible first step to analyse the extension in which one or both the sub-models are multivariate. This choice is related to the fact that we want to investigate the situations in which not only one but more than one longitudinal covariates influenced the event, in order to obtain better estimations. In fact we think that increasing the number of covariates that influence the event, will deal to better estimations of the risk of the event or of the survival function, as there is an increase of the information at disposal to investigate the event. In addition

it is possible to find different covariates, not only one, that may influence the event, considering also the interaction and the joint effect of these covariates on the risk of the event.
As for the case in which both sub-models are univariate, also for the extensions the literature is very extensive. In fact it is possible to find several researchers that focused on one or more extensions. Concerning the case in which the longitudinal sub-model is multivariate, several authors focused on possible assumptions on the distribution of the covariates, different formulation of longitudinal sub-models, or the use of Bayesian methods for the estimations.
The great problem related to the situation in which one or both sub-models are multivariate concerns the computational aspect of the estimation. The joint model with univariate sub-models is computationally demanding, then increasing the number of parameters that must be estimated or the dimensions of the sub-models will lead to methods of estimation more computationally demanding. The aim of this thesis is to find some methods of estimation and to implement some new algorithms in software R that could help to solve the problem of the computational aspect, focusing, as a first step, on the situation in which only the longitudinal sub-model is multivariate.
As a first possible method of estimation for the case in which only longitudinal sub-model is multivariate, we proposed a two-stage approach as it permits to obtain very fast and with desirable proprieties estimations. The two-stage approach is based on two steps. In the first step the parameter for the longitudinal sub-model are estimated using a maximum likelihood approach. Subsequently in the second step these estimates are used to impute appropriate values in the classical partial likelihood of the Cox model.
The most of the applications of joint models focused on the medical data, because in clinical trail it is very interesting to analyse two subgroups, for example placebo and treated, in order to study the longitudinal covariates that could influence the survival or the effect of a new drug. In this thesis the focus of the applications is on the undergraduates' career, jointly analysing the time to graduation and the student's path, in order to study the influence of some longitudinal covariates on the event graduation. This is one of the biggest novelty of this thesis, as the time to graduation had never been analysed through a joint model, neither in joint model in which the sub-models are univariate.
This thesis is organised as follow: after a brief introduction, in the follower two chapters there is a review of the longitudinal and survival models.
The fourth chapter reports a review of the definitions, method of estimations, diagnostics, and applications of the classical joint models.
Subsequently in the fifth chapter there is a review of the possible extensions of the classical joint models. The sixth chapter is given by a deep review of the extension with multivariate longitudinal sub-model and by a proposal of using the two-stage approach for the estimations of the joint model with multivariate longitudinal sub-model.
The seventh chapter concerns the results of this method applied to inves-

tigate the effect of some longitudinal covariate on the event graduation for university students. Particularly firstly the results with univariate longitudinal sub-models are shown, subsequently the results obtained by the two-stage method of estimation with a bivariate longitudinal sub-model are discussed.
The last chapter reports some conclusions and proposals for further work.

Chapter 2

Longitudinal models

The longitudinal models analyse data that are collected repeatedly in time, the outcome and the covariates are collected in different time points in order to evaluate the trend and the outcome change over time. The basic formulation used to analyse the longitudinal data is called marginal model which is based on a classical linear regression model:

$$\mathbf{Y}_i = X_i\beta + \epsilon_i \tag{2.1}$$

where $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in_i})$ and Y_{ij} is the value j observed on the subject i at time t_{ij}, X_i is a known matrix of covariates, β is a vector of parameters, and ϵ_i are the error terms. The most famous marginal model used to analyse the longitudinal data is the mixed model, that helps to evaluate also the individual influence on the evolution of the observed variables.

2.1 Mixed models

The linear mixed model is based on the idea that each individual has his own subject specific mean response profile over time. The mixed model can be represented by these equations introduced by Laird and Ware [29]:

$$\begin{cases} \mathbf{Y}_i = X_i\beta + Z_i b_i + \epsilon_i \\ b_i \sim N(\mathbf{0}, D) \\ \epsilon_i \sim N(\mathbf{0}, \Sigma_i) \\ \mathbf{b}_1, ..., \mathbf{b}_N, \epsilon_1, ..., \epsilon_N \quad \textit{independent} \end{cases} \tag{2.2}$$

where $\mathbf{Y}_i$ is the n_i-dimensional response vector for subject i $(i = 1, ..., N)$, N is the number of subjects measured at time t_{ij}, and n_i is the number of answers given by subject i. X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates, β is a p-dimensional vector containing the fixed effects, b_i is the q-dimensional vector containing the random effects, and ϵ_i is an n_i-dimensional vector of residual components or error terms. Finally, D is a general $(q \times q)$ covariance matrix with every (i, j) element that respects this relation $d_{ij} = d_{ji}$ and Σ_i is a $(n_i \times n_i)$ covariance matrix which depends on i only through its dimension n_i, i.e. the set of unknown

parameters in Σ_i will not depend on i.
Due to the distributions of error terms, it is possible to obtain that, conditional on the random effect b_i, $\mathbf{Y}_i$ is normally distributed with mean vector $X_i\beta + Z_i b_i$ and with covariance matrix Σ_i, which is often posed equal to $\sigma^2 I_{n_i}$. Further, b_i is assumed to be normally distributed with mean vector 0 and covariance matrix D.
Therefore this model analyses two type of effects, the fixed and the random effects. The fixed effects describe the impact of known measured covariates, where such effects are assumed to hold over all the population of individuals. On the other hand, the random effects measure the impact of known variables, where effects are assumed to vary across individuals in the population.
In addition the error terms could be split in two parts [56], $\epsilon_i = \epsilon_{(1)i} + \epsilon_{(2)i}$, thus the general model becomes:

$$\begin{cases} \mathbf{Y}_i = X_i\beta + Z_i b_i + \epsilon_{(1)i} + \epsilon_{(2)i} \\ b_i \sim N(\mathbf{0}, D) \\ \epsilon_{(1)i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}) \\ \epsilon_{(2)i} \sim N(\mathbf{0}, \tau^2 H_i) \\ \mathbf{b}_1, ..., \mathbf{b}_N, \epsilon_{(1)1}, ..., \epsilon_{(1)N}, \epsilon_{(2)1}, ..., \epsilon_{(2)N} \quad \textit{independent} \end{cases} \tag{2.3}$$

where H_i is the correlation matrix and each elements of this matrix is often created by this relation $h_{ijk} = g(|t_{ij} - t_{ik}|)$ where $g(\cdot)$ is a fixed function.
Let $f(\mathbf{y}_i|\mathbf{b}_i)$ and $f(\mathbf{b}_i)$ be the density functions of $\mathbf{y}_i|\mathbf{b}_i$ and $\mathbf{b}_i$, therefore the marginal density function of $\mathbf{Y}_i$ is given by:

$$f(\mathbf{y}_i) = \int f(\mathbf{y}_i|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \tag{2.4}$$

which is the density function of an n_i-dimensional normal distribution with mean vector $X_i\beta$ and with covariance matrix $V_i = Z_i D Z_i^{'} + \Sigma_i$. This marginal distribution is used for the maximum likelihood estimation, then starting from:

$$\mathbf{Y}_i \sim N(X_i\beta; Z_i D Z_i^{'} + \Sigma_i) \tag{2.5}$$

the classical likelihood function becomes:

$$\begin{aligned} L_{ML}(\theta) = \prod_{i=1}^{N} \Bigg\{ & (2\pi)^{-n_i/2} |V_i|^{1/2} \\ & \times \exp\left[-\frac{1}{2} (\mathbf{Y}_i - X_i\beta)' V_i^{-1}(\alpha) (\mathbf{Y}_i - X_i\beta) \right] \Bigg\} \end{aligned} \tag{2.6}$$

where θ denotes the parameters and α denotes the vector of all variance and covariance parameters found in V_i.
Supposing that α is known, it is possible to obtain the estimation of the β:

$$\hat{\beta}(\alpha) = \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i y_i \tag{2.7}$$

where $W_i = V_i^{-1}$. Note that this estimator is the same obtained by the generalized least square method. If α is not known but there is an estimation of it, it can be used and the estimator change accordingly.
The expected value of the estimator is:

$$\begin{aligned} E\left[\hat{\beta(\alpha)}\right] &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \sum_{i=1}^{N} X_i' W_i E(\mathbf{Y}_i) = \\ &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \sum_{i=1}^{N} X_i' W_i X_i \beta = \beta \end{aligned} \tag{2.8}$$

thus the β estimator is unbiased, and the variance and covariance matrix is:

$$\begin{aligned} Var\left[\hat{\beta(\alpha)}\right] &= \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' W_i Var(\mathbf{Y}_i) W_i X_i\right) \\ &\left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} = \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \end{aligned} \tag{2.9}$$

Concerning the random effects $\mathbf{b}_i$, they can be estimated using the Bayesian methods. In fact it can be found that [56]:

$$f(\mathbf{b}_i|\mathbf{y}_i) \equiv f(\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i) = \frac{f(\mathbf{b}_i|\mathbf{y}_i) f(\mathbf{b}_i)}{\int f(\mathbf{b}_i|\mathbf{y}_i) f(\mathbf{b}_i) d\mathbf{b}_i} \tag{2.10}$$

And often the estimations are made by the expected value of this probability, thus it becomes:

$$\hat{\mathbf{b}}_i(\theta) = E[\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i] = \int \mathbf{b}_i f(\mathbf{b}_i|\mathbf{y}_i) d\mathbf{b}_i = DZ_i' W_i(\alpha)(\mathbf{y}_i - X_i\beta) \tag{2.11}$$

But if α is not known or there is no estimation for it, the Restricted Maximum Likelihood (REML) method must be used to estimate the parameters. Therefore Verbeke and Molenberghs [56] introduces the matrix of error contrast $U = A'Y$ where A is a $(n \times (n-p))$ full-rank matrix with columns orthogonal to the columns of X. The vector U is distributed as a normal with mean vector 0 and variance and covariance matrix given by $A'V_i(\alpha)A$. From this the likelihood function becomes:

$$\begin{aligned} L(\alpha) =& (2\pi)^{-(n-p)/2} \left|\sum_{i=1}^{N} X_i' X_i\right|^{1/2} \times \left|\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right|^{-1/2} \\ & \prod_{i=1}^{N} |V_i|^{-1/2} \times \exp\left[-\frac{1}{2}(\mathbf{Y}_i - X_i\hat{\beta})' V_i^{-1} (\mathbf{Y}_i - X_i\hat{\beta})\right] \end{aligned} \tag{2.12}$$

This formulation can also be simplified and it becomes:

$$L(\alpha) = C \left|\sum_{i=1}^{N} X_i' W_i^{-1}(\alpha) X_i\right|^{-1/2} L_{ML}(\hat{\beta}(\alpha), \alpha) \tag{2.13}$$

where C is a constant that not depending on α.
The maximization of the likelihood or of the log-likelihood cannot be written in closed form, thus mathematical algorithms are used to obtain a numerical optimization. One of this method is the Expectation-Maximization (EM) algorithm [12] which is based on two steps, the expectation (E) step, which creates a function for the expectation of the likelihood using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected likelihood found on the E step. Another mathematical method is the Newton-Raphson which is an iterative methods used to find a numerical solution starting with a first value and then trying to approach the solution using the tangent.
As Verbeke and Molenberghs [56] and Hedecker and Gibbons [21] showed, the model can be also written in a matrix formulation:

$$\mathbf{Y} = X\beta + Z\mathbf{b} + \epsilon \tag{2.14}$$

where the vectors $\mathbf{Y}$, $\mathbf{b}$, and ϵ are composed by stacking the single vectors of the subjects $\mathbf{Y}_i$, $\mathbf{b}_i$, and ϵ_i, the matrix X is composed by the single matrices X_i, and Z is a block diagonal matrix with all the elements on the diagonal that are the matrices Z_i. The dimensions of the vector $\mathbf{Y}$ is given by $\sum_{i=1}^{N} n_i = n$.
An easier formulation of the model based only on the time is possible [21], accordingly the model becomes:

$$y_{ij} = \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij} \tag{2.15}$$

This equation can also be written highlighting the random and fixed effects and it becomes:

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\, t_{ij} + \epsilon_{ij} \tag{2.16}$$

where $b_i = (b_{i0}, b_{i1})^T$ are called the random effects and β_1 and β_0 are called the fixed effects.

2.2 Missing data

During the longitudinal studies it often happens that someone not give the answer, then there are missing data. The biggest problems of missing data are related to the loss of efficiency, the influence on the design, and the creation of unbalanced data sets which can leads to bias and misleading inferences. In these cases the answers are split in two types, the y_i^o, which contains all the value of observed y_{ij}, and y_i^m, which contains the missing values.
There are two type of missing data: monotone and intermittent. The first one incorporated the case in which a subject withdraws from the study (drop-out) or when a subject comes into the study later (late entry). The second one represents the case in which a subject answers just intermittently, therefore one time he answers, the next one no, later he answers

again and so on.
If in the study there are some missing data, the most important probability analysed is the probability of missingness process r_i given the complete response vector $Y_i = (y_i^o, y_i^m)$, defined as:

$$p(r_i|y_i^o, y_i^m; \theta_r) \tag{2.17}$$

where θ_r is the vector of parameters and $r_i = (r_{i1}, ..., r_{in_i})'$ where r_{ij} is equal to 1 only if y_{ij} is observed, 0 otherwise.
Rubin [47] classified the different missing data mechanisms in:

1. Missing completely at random (MCAR): in this case r_i is independent of the complete response, so:

$$p(r_i|y_i^o, y_i^m; \theta_r) = p(r_i; \theta_r)$$

thus the observed and complete data have the same distribution. One example of this situation is when the researcher decides to stop the study after a fixed number of answers.

2. Missing at random (MAR): in this case the probability of missingness depends only on the observed data, defined as:

$$p(r_i|y_i^o, y_i^m; \theta_r) = p(r_i|y_i^o; \theta_r)$$

therefore it is possible to predict the missing values using the observed values. One example of this situation is when a subject decides to go out of the study if he has a specific characteristic observed. From the probability of missingness it is possible to obtain:

$$\begin{aligned} p(y_i^m|y_i^o, r_i; \theta) &= \frac{p(r_i, y_i^o, y_i^m; \theta)}{p(y_i^o, r_i; \theta)} = \frac{p(r_i|y_i^o, y_i^m; \theta_r)p(y_i^o, y_i^m; \theta_y)}{p(r_i|y_i^o; \theta_r)p(y_i^o, r_i; \theta_y)} \\ &= \frac{p(r_i|y_i^o; \theta_r)p(y_i^o, y_i^m; \theta_y)}{p(r_i|y_i^o; \theta_r)p(y_i^o, r_i; \theta_y)} = \frac{p(y_i^o, y_i^m; \theta_y)}{p(y_i^o, r_i; \theta_y)} \\ &= p(y_i^m|y_i^o; \theta_y) \end{aligned} \tag{2.18}$$

where θ is the parameter vector of the joint distribution of response vector and missingness process, and θ_y is the parameter vector of the distribution of response vector.
This type of missing data has a likelihood-based inferences, the ignorability, through which the likelihood function can be split in two parts:

$$\begin{aligned} L_i(\theta) &= \int p(y_i, r_i; \theta) dy_i^m = \int p(y_i^0, y_i^m; \theta_y) p(r_i|y_i^o, y_i^m; \theta_r) dy_i^m \\ &= \int p(y_i^0, y_i^m; \theta_y) p(r_i|y_i^o; \theta_r) dy_i^m = p(y_i^0; \theta_y) p(r_i|y_i^o; \theta_r) \\ &= L_i(\theta_y) L_i(\theta_r) \end{aligned}$$

3. Missing not at random (MNAR): in this case the missing depends on a subset of missing values, thus this probability of missingness is:

$$p(r_i|y_i^o, y_i^m; \theta_r) \text{ or } p(r_i|y_i^m; \theta_r)$$

An example of this situation is the fact that often the richest or the poorest decide to ignore the demands regarding the income.
This is the most difficult case to deal with because unlike the other situations in which a mixed model can be used, in this case it creates biased results. Therefore there are three different models to handle it, selection models, pattern mixture models, and shared-parameter models, in which cases the different formulations of the conjoint probabilities are respectively:

$$\begin{aligned}
p(y_i^o, y_i^m, r_i; \theta) &= p(y_i^o, y_i^m; \theta_y) p(r_i|y_i^o, y_i^m; \theta_r) \\
p(y_i^o, y_i^m, r_i; \theta) &= p(y_i^o, y_i^m|r_i; \theta_y) p(r_i; \theta_r) \\
p(y_i^o, y_i^m, r_i; \theta) &= \int p(y_i^o, y_i^m|b_i; \theta_y) p(r_i|b_i; \theta_r) p(b_i; \theta_b) db_i
\end{aligned}$$

Chapter 3

Survival models

The survival models analyse the time of an event, like the time of death, the time of degree, the time of appearance of a disease, and so on. Two functions characterise the survival models. The first one is the survival function that indicates the probability of not experiment the event before time t, or the probability of surviving at time t, defined as:

$$S(t) = Pr(T^* > t) \tag{3.1}$$

in addition it can be obtained from the distribution function $F(t)$, as long as it is absolutely continuous, because $S(t) = 1 - F(t)$.
The second function is the hazard function that indicates the instantaneous failure rate, hence it indicates the probability of experiment the event in the time interval $[t, t + dt)$ provided survival up to t, defined as:

$$h(t) = \lim_{dt \to 0} \frac{Pr(t \leq T^* < t + dt | T^* \geq t)}{dt} \qquad t > 0 \tag{3.2}$$

it can besides be evaluated from this relation:

$$h(t) = \frac{f(t)}{S(t)} \tag{3.3}$$

Furthermore even the cumulative hazard function can be evaluated that describes the accumulated risk until time t:

$$H(t) = \int_0^t h(s)ds \tag{3.4}$$

The survival function can be moreover expressed in term of the cumulative hazard function:

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int_0^t h(s)ds\right\} \tag{3.5}$$

3.1 Censored

As well as in the longitudinal models there are missing data, in survival models there are censored data, therefore the event of interest is not fully observed on all the subjects. There are three different types of censoring:

1. Right censoring: in this situation the event of interest is only known to occur after a certain time point, the true unobserved event is on the right of the censoring time. There are three classification of this type of censoring. The first one is Type I, when a subject may reach the end of the study without having experienced the event of interest. The second one is the Type II, when the study is terminated after a pre-specified number of events has recorded. The third one is random censoring, thus when a subject moves away from the study.

2. Left censoring: in this situation the event of interest is only known to occur before a certain time point, then the event has already occurred when observation time begins, true unobserved event is on the left of the time of study.

3. Interval censoring: in this situation the event of interest is only known to occur between two certain time points, for example the tests are made periodically but the researcher ignores when exactly the disease appeared. This type of censoring includes both the two previous situations, left and right censoring.

Another useful classification is based on the distinction between informative or non-informative censoring. The first type indicates the situation in which a subject withdraws from the study for reasons related to the failure time (similar to MNAR). The second type is the opposite, thus when the withdrawn is not relate to the event of interest, but it can depend on some covariates (similar to MCAR).

3.2 Estimation

The estimations for the survival function, if there is no censoring in the data, are based on the empirical estimate of the survival function that is the proportion of individuals with event times greater than t, expressed as:

$$\hat{S}(t) = \frac{number\ of\ indivisual\ with\ T^* \geq t}{n} \tag{3.6}$$

where n is the number of subject.
If instead there is censoring two methods of estimation are possible: the non-parametric or the parametric method. The parametric method can be used if the survival function assumes a specific parametric function, instead the non-parametric method is the only solution.
The most famous non-parametric method is the Kaplan-Meier estimator [28], defined as:

$$\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} \frac{r_i - d_i}{r_i} \tag{3.7}$$

where r_i denotes the number of subjects still at risk at the time t_i and d_i is the number of event at time t_i. This estimation is based on this relation:

$$\begin{aligned} S(T) &= p(T \geq t_{i+1}) = p(T \geq t_1, T \geq t_2, ..., T \geq t_{i+1}) \\ &= p(T \geq t_1) \prod_{j=1}^{i} p(T \geq t_{j+1} | T \geq t_j) = \prod_{j=1}^{i} [1 - p(T = t_{j+1} | T \geq t_j)] \\ &= \prod_{j=1}^{i} \left[1 - \frac{d_j}{r_j} \right] \end{aligned}$$

The estimator is approximately normal and asymptotically unbiased, but for evaluating the variance, the Greenwood's formula [19] is needed, therefore the variance estimation is:

$$var(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{i:t_i \leq t} \frac{d_i}{(r_i - d_i) r_i} = \left(\prod_{i:t_i \leq t} \frac{r_i - d_i}{r_i} \right)^2 \sum_{i:t_i \leq t} \frac{d_i}{(r_i - d_i) r_i} \tag{3.8}$$

It is obtained by the fact that [24]:

$$\ln(\hat{S}(t)) = \sum_{i:t_i \leq t} \frac{r_i - d_i}{r_i} = \sum_{i:t_i \leq t} \hat{p}_i$$

where $\hat{p}_i = \frac{r_i - d_i}{r_i}$. If t_i is considered as a Bernoulli distribution and $\hat{p}_i$ is the estimation of its probability, this estimator is unbiased, then his expected value is p_i, and his variance is $(\hat{p}_i(1 - \hat{p}_i))/r_i$.
Using the delta method, which said that if Y is normal with mean μ and variance σ^2, then $g(Y)$ is approximately normally distributed with mean $g(\mu)$ and variance $[g'(\mu)]^2 \sigma^2$, the variance of $\log(Y)$ becomes:

$$Var(\log(Y)) \cong \frac{1}{\mu_Y^2} \sigma_Y^2$$

Applying this method to $\log(\hat{p}_i)$ it is possible to obtain:

$$\hat{Var}(\log(\hat{p}_i)) \cong \frac{1}{\hat{p}_i^2} \frac{\hat{p}_i(1 - \hat{p}_i)}{r_i} = \frac{d_i}{r_i(r_i - d_i)} \tag{3.9}$$

From this the variance of $log(\hat{S}(t))$ is:

$$\hat{Var}(\log(\hat{S}(t))) = \sum_{i:t_i \leq t} \hat{Var}(\log(\hat{p}_i)) = \sum_{i:t_i \leq t} \frac{d_i}{r_i(r_i - d_i)} \tag{3.10}$$

Trough the delta method it is possible to obtain the $var(\exp(X))$:

$$Var(\exp(X)) \cong (\exp(\mu_X))^2 \sigma_X^2$$

Applying this to $\exp(\log(\hat{S}(t))) = \hat{S}(t)$ it becomes:

$$\hat{Var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{i:t_i \leq t} \frac{d_i}{(r_i - d_i) r_i} \tag{3.11}$$

Frequently even the variance of $\log(-\log(\hat{S}(t))) = \log(\hat{H}(t))$ is analysed:

$$\hat{Var}(\log(\hat{H}(t))) = \frac{1}{[\log(\hat{S}(t))]^2} \sum_{i:t_i \leq t} \frac{d_i}{r_i(r_i - d_i)} = \frac{\sum_{i:t_i \leq t} d_i/[r_i(r_i - d_i)]}{[\sum_{i:t_i \leq t} \log((r_i - d_i)/r_i]^2} \tag{3.12}$$

From the estimation of variance the confidence intervals for $\hat{S}(t)$ are built as it is asymptotically normal:

$$\hat{S}(t) \pm z_{1-\alpha/2} se(\hat{S}(t)) \tag{3.13}$$

As said, if the parametric form of the survival function is known, a parametric method could be used to estimate the parameters: the maximum likelihood. In this situation a subject that experiment the event contributes to the likelihood with the density function $f(T_i; \theta)$, while a subject that is censored is still survived up to the T_i point then he contributes to the likelihood with the survival function $S_i(T_i; \theta)$. Thus the log-likelihood function becomes:

$$l(\theta) = \sum_{i=1}^{n} [\delta_i \log f(T_i; \theta) + (1 - \delta_i) \log S_i(T_i; \theta)] \tag{3.14}$$

where δ_i is equal to 1 if the subject is not censored and 0 otherwise.
From the relations introduced before, which are $h(t) = \frac{f(t)}{S(t)}$ and $S(t) = \exp\{-H(t)\}$, it is possible to obtain:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} [\delta_i \log f(T_i; \theta) + \log S_i(T_i; \theta) - \delta_i \log S_i(T_i; \theta)] \\ &= \sum_{i=1}^{n} \left[\delta_i \log \frac{f(T_i; \theta)}{S_i(T_i; \theta)} + \log S_i(T_i; \theta) \right] \\ &= \sum_{i=1}^{n} \left[\delta_i \log h_i(T_i; \theta) - \int_0^{T_i} h_i(s; \theta) ds \right] \end{aligned}$$

Therefore it can be seen that the censored observations contribute with less information to the statistical inference than uncensored ones.

3.3 Cox and extended Cox models

A particular class of survival model is the Cox model [6], also known as relative risk or proportional hazards model. In this model the covariates have a multiplicative effect on the hazard function and it is defined as:

$$h_i(t|\omega_i) = \lim_{dt \to 0} \frac{p(t \leq T^* < t + dt | T^* \geq t, \omega_i)}{dt} = h_0(t) \exp(\gamma' \omega_i) \tag{3.15}$$

where $\omega_i' = (\omega_{i1}, ..., \omega_{ip})$ denotes the vector of covariates associated with the hazard function, γ indicates the coefficient vector for the covariates, and $h_0(t)$ is called the baseline hazard. The baseline hazard can have different

formulations, it can also be constant or correspond to a known parametric distribution, like for example Weibull, log-normal, or Gamma.
The model could also be written in log scale:

$$\log h_i(t|\omega_i) = \log h_0(t) + \gamma_1\omega_{i1} + \gamma_2\omega_{i2} + ... + \gamma_p\omega_{ip} \tag{3.16}$$

The survival function can be obtained through the relation presented above $S(t) = \exp[-H(t)]$:

$$S(t|\omega_i) = \exp[-h_0(t)\exp(\gamma'\omega_i)] = [S_0(t)]^{\exp(\gamma'\omega_i)} \tag{3.17}$$

where $S_0(t) = \exp[-h_0(t)]$.
Note that using this model the ratio of hazard between two subjects, i and j, has a particular formulation and it is constant, as does not depend on time, in fact:

$$\frac{h_i(t|\omega_i)}{h_k(t|\omega_k)} = \exp[\gamma'(\omega_i - \omega_k)] \tag{3.18}$$

From the classical likelihood function it is possible to obtain the estimation of the parameters, dealing also with the censored data [24], because for the censored observations the value of $f(t,\gamma)$ is considered while for the uncensored observations the value of $S(t,\gamma)$ is used:

$$L(\gamma) = \prod_{i=1}^{n} \left\{ f(t_i,\gamma)^{\delta_i} S(t_i,\gamma)^{1-\delta_i} \right\} \tag{3.19}$$

where $\delta_i = 1$ if uncensored and 0 otherwise. From this log-likelihood function becomes:

$$l(\gamma) = \sum_{i=1}^{n} \{\delta_i \log[f(t_i,\gamma)] + (1-\delta_i)\log[S(t_i,\gamma)]\} \tag{3.20}$$

But Cox [6] introduces an easier method for estimate the parameter γ, the partial log-likelihood, defined as:

$$pl(\gamma) = \sum_{i=1}^{n} \delta_i \left\{ \gamma'\omega_i - \log\left[\sum_{T_j \geq T_i} \exp(\gamma'\omega_j) \right] \right\} \tag{3.21}$$

This method is based on the partial likelihood, expressed as:

$$L(\gamma) = \prod_{i=1}^{n} \left\{ \frac{exp(\gamma'\omega_i)}{\sum_{T_j \geq T_i} \exp(\gamma'\omega_j)} \right\}^{\delta_i}$$

which, applying the logarithm function, becomes:

$$\begin{aligned} l(\gamma) &= \log \left\{ \left[\prod_{i=1}^{n} \frac{exp(\gamma'\omega_i)}{\sum_{T_j \geq T_i} \exp(\gamma'\omega_j)} \right]^{\delta_i} \right\} \\ &= \sum_{i=1}^{n} \delta_i \left\{ \gamma'\omega_i - \log\left[\sum_{T_j \geq T_i} \exp(\gamma'\omega_j) \right] \right\} \end{aligned}$$

Then the estimations can be obtained by differentiating the partial likelihood function with respect to the parameter γ:

$$\frac{\partial pl(\gamma)}{\partial \gamma'} = \sum_{i=1}^{n} \delta_i \left\{ \omega_i - \frac{\sum_{T_j \geq T_i} \omega_j \exp(\gamma' \omega_j)}{\sum_{T_j \geq T_i} \exp(\gamma' \omega_j)} \right\} = 0 \tag{3.22}$$

In order to evaluate the variance, the inverse of expected information matrix $E[I(\hat{\gamma})]^{-1}$ is needed, where:

$$I(\hat{\gamma}) = -\sum_{i=1}^{n} \frac{\partial^2 pl_i(\gamma)}{\partial \gamma' \partial \gamma}\bigg|_{\gamma=\hat{\gamma}} \tag{3.23}$$

The extended Cox models include in the hazard function also exogenous time-dependent covariates, where an exogenous variables is not affected by the occurrence of failure, it is a predictable process, in contrast with the endogenous ones that are not predictable and are in function of the event of failure. In the extended Cox models the formulation for the hazard function is based on the work of Andersen and Gill [2]:

$$h_i(t|y_i(t), \omega_i) = h_0(t) R_i(t) \exp[\gamma' \omega_i + \alpha y_i(t)] \tag{3.24}$$

where $y_i(t)$ indicates the time-dependent covariate and α is his coefficient, $R_i(t)$ assumes value 1 if subject i is at risk at time t and 0 otherwise, and all the other elements are the same indicated in the classical Cox models. In addition the element $N_i(t)$ must be considered that indicates the number of events for subject i by time t. The authors used the partial log-likelihood function for estimating the parameters γ and α:

$$\begin{aligned} pl(\gamma) = \sum_{i=1}^{n} \int_0^{\infty} \Bigg\{ & R_i(t)[\gamma' \omega_i + \alpha y_i(t)] \\ & - log \left[\sum_j R_j(t) \exp[\gamma' \omega_j + \alpha y_j(t)] \right] \Bigg\} dN_i(t) \end{aligned} \tag{3.25}$$

Andersen and Gill [2] demonstrated that estimated parameters are asymptotically normal.
The limit of this type of model is that it considers only exogenous covariates, thus the values of each covariates are known at every failure time, moreover no longitudinal error is considered. This is one of the reason for the introduction of the joint models.

Chapter 4

Joint models

The joint models for longitudinal and time-to-event data are a recent family of models that analyse jointly the longitudinal and the survival data. With the joint models it is possible to analyse a repeatedly measured outcome and his association with a time-to-event outcome. Nevertheless joint models allow to analyse the time-to-event outcomes considering their association with the repeatedly measured outcomes, where the longitudinal time-dependent covariates can be measured with errors. Alternately the aim of these models can be the study of the relationship between time-to-event and repeatedly measure outcomes, that is the association between survival and longitudinal processes.

4.1 Model definitions and estimations

In 1988 Wu and Carroll [61] proposed a method that helps to solve the problem related to right censored data using a linear random effects model, therefore for the first time a longitudinal model is considered jointly with an event. Moreover the authors used likelihood ratio test for analysing informativeness and maximum likelihood for estimating the response parameters. As long as the primary right censoring process coefficients are derived under probit model, where primary right censoring is the right censoring caused by the participant's death or withdrawal.
The first formulation of the joint model was presented by De Gruttola and Tu [9, 10]. The authors considered a population of n subjects, indexed by i, each of whom has m_i observations of a marker of disease progression. Let y_i be an $m_i \times 1$, whose elements y_{ij} are the values of the marker for the person i on the occasion j of measurement for $i = 1, ..., n$ and $j = 1, ..., m_i$. Let $\tau_i = \min(x_i, c_i)$, where x_i and c_i denote respectively the survival and censoring times for the subject i. The authors modelled the progression of the marker y_i using a random effects model:

$$y_i = T_i\alpha + Z_i b_i + \epsilon_i \quad i = 1, ..., n \tag{4.1}$$

where α is a $p \times 1$ vector of unknown parameters, T_i is a known full-rank $m_i \times p$ design matrix linking α to y_i, $b_i \sim N(0, D)\, iid$ denotes a $k \times 1$ vector

of unknown individual effects, Z_i is a known $m_i \times k$ design matrix linking b_i to y_i, $\epsilon_i \sim N(0, \sigma^2 I)$ is a vector of residuals, and I is an $m_i \times m_i$ identity matrix. In addition De Gruttola and Tu [9, 10] proposed a transformation of the survival time:

$$x_i = w_i'\xi + \lambda' b_i + r_i \tag{4.2}$$

where ξ is a q length vector of unknown parameters, w_i is a q length design vector linking x_i to ξ, λ is a k length vector of unknown parameters linking b_i to x_i, and $r_i \sim N(0, s^2)$ iid are the residuals.
The authors assumed that all the censoring is non informative and the missing data are missing at random, from these assumptions is possible to obtain the likelihood contribution for any subject:

$$L = \phi(y_1|b) \prod_{j=2}^{M} [\phi(y_j|b) g(t_j|y_1, ..., y_{j-1}; t_j < x)]^{I(t_j<x)} [\phi(x|b)]^{\rho} [1 - \Phi(c|b)]^{1-\rho} \tag{4.3}$$

where c is a constant, and $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the probability density and the cumulative distribution function of the standard Normal distribution. The joint log-likelihood function can be written using the previous equation as:

$$\begin{aligned} L_{obs} = &\sum_{i=1}^{N^o} log \left[\int_{b_i} \phi(y_i|b_i, \alpha, \sigma^2) \phi(b_i|D) \phi(x_i|b_i, \xi, s^2) db_i \right] \\ &+ \sum_{i=1}^{N^c} log \left\{ \int_{b_i} \phi(y_i|b_i, \alpha, \sigma^2) \phi(b_i|D) \left[1 - \Phi(c_i|b_i, \xi, s^2) \right] db_i \right\} \end{aligned} \tag{4.4}$$

where $\Phi(c_i|b_i, \xi, s^2) = \int_{-\infty}^{c_i} \phi(x|b_i, \xi, s^2) dx$, and N^o and N^c denote the number of failed and censored individuals respectively. If there is no censoring, the equation is formed by only the first part. In this case the EM algorithm for the estimations can be used as the joint distribution of longitudinal and survival data is supposed to follow a multivariate normal distribution. The sufficient statistics required are then:

$$\sum_{i=1}^{n} b_i b_i', \ \sum_{i=1}^{n} \epsilon_i' \epsilon_i, \ \sum_{i=1}^{n} r_i^2, \ \sum_{i=1}^{n} Z_i b_i', \ \sum_{i=1}^{n} x_i b_i, \ and \ \sum_{i=1}^{n} w_{ij} b_i \ (1 \le j \le q)$$

The estimations obtained from log-likelihood maximisation are:

$$\begin{aligned} \hat{\sigma}^2 &= \left(\sum_{i=1}^{n} m_i \right)^{-1} E \left(\sum_{i=1}^{n} \epsilon_i' \epsilon_i | Y, \hat{\theta} \right) \\ \hat{s}^2 &= n^{-1} E \left(\sum_{i=1}^{n} r_i^2 | Y, \hat{\theta} \right) \\ \hat{D} &= n^{-1} E \left(\sum_{i=1}^{n} b_i b_i' | Y, \hat{\theta} \right) \end{aligned}$$

From these estimation the estimations of the other parameters become:

$$\hat{\lambda} = \left(\sum_{i=1}^{n} b_i b_i'\right)^{-1} \sum_{i=1}^{n} \left(x_i - w_i'\hat{\xi}\right) b_i$$
$$\hat{\alpha} = \left(\sum_{i=1}^{n} T_i' T_i\right)^{-1} \sum_{i=1}^{n} T_i' \left(y_i - Z_i \hat{b}_i\right)$$
$$\hat{\xi} = \left(\sum_{i=1}^{n} w_i w_i'\right)^{-1} \sum_{i=1}^{n} w_i \left(x_i - \hat{\lambda}'\hat{b}_i\right)$$

If there is some censoring, the authors supposed that they are non informative and with some modifications the estimation of the parameters is possible by the EM algorithm. In the application a comparison between the joint maximization and the two steps maximisation is proposed. Concerning the two steps maximisation, the first step fits the growth curve model using maximum likelihood and the second one is ordinary least squares or proportional hazards regression of the survival times applied on the estimated random effects obtained from the first step. This comparison shows that the joint maximization creates more unbiased and efficient results. The same results are obtain from a simulation study.
Another model is based on the work of Tsiatis, DeGruttola, and Wulfsohn [55]. The authors proposed a two-stage approach. In the first stage, the longitudinal time dependent data are modelled using repeated measures random components models, while in the second stage methods are developed for estimating the parameters in a Cox proportional hazard model.
The model was based on the hypothetical true value of the longitudinal data $Z^*(t)$ and the history up to time t, $\overline{Z}^*(t)$, was introduced in the hazard function in order to analyse the relationship between survival and the longitudinal values:

$$\lambda(t|\overline{Z}^*(t)) = \lim_{h\to 0} \frac{1}{h} P[t \leq T \leq t+h | T \geq t, \overline{Z}^*(t)] = \lambda_0(t) f(\overline{Z}^*(t), \beta) \quad (4.5)$$

The estimation was based on the classical partial likelihood proposed by Cox [6], but the effect of measurement error must be consider, then the model becomes:

$$Z(t) = Z^*(t) + e(t) \quad (4.6)$$

where $e(t)$ is a zero mean error with $var(e(t)) = \sigma^2$ and $cov(e(s), e(t)) = 0$ with $s \neq t$. From the equation (4.5) it is possible to obtain:

$$\begin{aligned} \lambda(t|\overline{Z}(t)) &= \int \lambda(t|\overline{Z}(t), \overline{Z}^*(t)) dP(\overline{Z}^*(t)|\overline{Z}(t), X \geq t) \\ &= \lambda_0(t) E[f(\overline{Z}^*(t), \beta)|Z(t_1), ...Z(t_j), X \geq t] \end{aligned} \quad (4.7)$$

and the parameter β can be again estimated with the classical partial likelihood proposed by Cox [6]. A formulation of the hazard function in relation with the current value $Z^*(t)$ can be considered:

$$f(Z^*(t), \beta) = \exp(\beta Z^*(t)) \quad (4.8)$$

and the estimation can be evaluated with the conditional expectation and the partial likelihood. Concerning the longitudinal analysis, the mixed model [29] is used:

$$Z_i(u) = Z_i^*(u) + e_i(u), \quad u \leq t \tag{4.9}$$

where $Z_i^*(u) = \theta_{0i} + \theta_{1i}u$. From this each hazard function can be modelled by the parameter that indicates the past history θ_i:

$$\lambda_i(t) = \lambda_0(t) \exp[\beta_1(\theta_{0i} + \theta_{1i}u) + \beta_2\theta_{1i}] \tag{4.10}$$

The parameters can be estimated imputing the empirical Bayes estimates of the individual random effects in the partial likelihood. An adjustment of the model for the missing data pattern considering the hazard also as a function of the timing of the events can be introduced:

$$\lambda(t|\overline{Z}(t), \overline{Z}^*(t)) = \lambda_0(t) \exp[\beta_0 Z^*(t) + \gamma D(t)] \tag{4.11}$$

where $D(t) = t - t_j$.
A similar model was proposed by Self and Pawitan [48], but instead of a typical hazard function, the authors proposed a hazard function which has a linear relation with the random effects, then the exponential function was replaced with a linear function of the longitudinal values.
Faucett and Thomas [15] in 1996 used the Markov chain Monte Carlo techniques of Gibbs sampling to estimate the joint posterior distribution of the unknown parameters of the model. The authors modelled a continuous covariate over time and simultaneously related the covariate to disease risk. The model considers censoring of the survival time and allows for unequally spaced or missing covariate data, so different numbers of observations per subject are possible.
Gibbs sampling is used to fit simultaneously the covariate tracking model and the disease risk model, which are the two sub-models. The covariate tracking model is a random components model with normal errors defined as:

$$z_{ij} = x_i(t_{ij}) + \epsilon_{ij} \tag{4.12}$$

where $x_i(t)$ is the true unobserved value of covariates, z_{ij} is a continuous time dependent covariate, t_{ij} are the observational times, and ϵ_{ij} are the error terms assumed independent and normally distributed with mean 0 and variance σ_ϵ^2. While the disease risk model is a proportional hazard model defined as:

$$\lambda_i(t) = \lambda_0(t) \exp[\gamma x_i(t)] \tag{4.13}$$

where $\lambda_0(t)$ is the baseline hazard of disease and γ is the regression coefficient for estimation. It is posed that the baseline hazard is a step function, therefore $\lambda_0(t) = \lambda_k$.
From this the posterior distribution of model parameters was estimated, starting from the random effects using Bayes' rule. Subsequently the estimation of the means and covariance matrix, of the error variance, of the baseline hazard and of the disease risk parameter can be obtained. The authors highlighted that the combined analysis is a feasible approach, and the

parameter estimations are more efficient, also if Gibbs sampling is computationally intensive.
Another model is based on the work of Wulfsohn and Tsiatis [62]. The authors criticized the two-stage approach and presented a new approach in which only the random effects are normally distributed. In the model there was T_i that indicates the survival time for individual i, $X_i = \min(T_i, C_i)$ where C_i corresponds to the censoring time, Δ_i is the failure indicator. In addition m_i indicates the number of measurements of the covariate and n indicates the number of individuals, t_{ij} is the time of observation for subject i for $j = 1, ..., m_i$ and Z_i indicates the values observed in different times. The first model used for the longitudinal data is a linear growth model with random intercept and slope:

$$Z_{ij} = \theta_{0i} + \theta_{1i} t_{ij} + \epsilon_{ij} \tag{4.14}$$

where ϵ_{ij} is a $N(0, \sigma_\epsilon^2)$, and $\boldsymbol{\theta}_i \sim N(\boldsymbol{\theta}, \boldsymbol{\sigma})$. For the survival model a Cox model is used:

$$\lambda(t|\theta_i, Z_i, t_i) = \lambda_0(t) \exp[\beta(\theta_{0i} + \theta_{1i} t)] \tag{4.15}$$

The observed data likelihood was given by:

$$\prod_{i=1}^{n} \left\{ \int_{-\infty}^{+\infty} \left[\prod_{j=1}^{m_i} f(z_{ij}|\theta_i, \sigma_\epsilon^2) \right] f(\theta_i|\theta, V) f(X_i, \Delta_i|\theta_i, \lambda_0, \beta) d\theta_i \right\} \tag{4.16}$$

where:

$$f(z_{ij}|\theta_i, \sigma_\epsilon^2) = (2\pi\sigma_\epsilon^2)^{-1/2} \exp\left[-\frac{(z_{ij} - \theta_{0i} - \theta_{1i} t_{ij})^2}{2\sigma_\epsilon^2} \right]$$
$$f(\theta_i|\theta, V) = (2\pi|V|)^{-1/2} \exp\left[-\frac{(\theta_i - \theta)' V^{-1} (\theta_i - \theta)}{2} \right]$$

and

$$f(X_i, \Delta_i|\theta_i, \lambda_0, \beta)$$
$$= \left\{ \lambda_0(X_i) \exp[\beta(\theta_{0i} + \theta_{1i} X_i)]^{\Delta_i} \exp\left[- \int_0^{X_i} \lambda_0(u) \exp[\beta(\theta_{0i} + \theta_{1i} X_i)] du \right] \right\}$$

Subsequently the EM algorithm is used for the estimations and it is possible to obatin the estimations for all the parameter except for the β for which the one-step Newton-Raphson method is used. The results are therefore:

$$\begin{aligned}
\hat{\theta} &= \sum_{i=1}^{n} E_i(\theta_i)/n \\
\hat{V} &= \sum_{i=1}^{n} E_i(\theta_i - \hat{\theta})(\theta_i - \hat{\theta}) \\
\hat{\sigma}_e^2 &= \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} E_i(z_{ij} - \theta_{0i} - \theta_{1i} t_{ij})^2}{\sum_{i=1}^{n} m_i} \\
\hat{\lambda}_0(u) &= \sum_{i=1}^{n} \frac{\Delta_i I(X_i = u)}{\sum_{j=1}^{n} E_j\{\exp[\beta(\theta_{0j} + \theta_{1j} u)]\} Y_j(u)}
\end{aligned} \tag{4.17}$$

where $Y_j(u)$ is the at risk indicator equal to $I(X_j \geq u)$. Concerning β an iterative method was used and the estimation at the iteration k was:

$$\hat{\beta}_k = \hat{\beta}_{k-1} + I^{-1}_{\hat{\beta}_{k-1}} S_{\hat{\beta}_{k-1}} \tag{4.18}$$

where $S_{\hat{\beta}_{k-1}}$ is the score for β at the iteration $k-1$, while $I_{\hat{\beta}_{k-1}}$ is its information for the same iteration.
Another important model is based on the work of Henderson, Diggle, and Dobson [22]. The authors considered a set of m subjects observed over interval $[0, \tau)$ that provide a set of measurements $\{y_{ij} : j = 1, ..., n_i\}$ at times $\{t_{ij}, j = 1, ..., n_i\}$, in addition there were also the realizations of a counting process $\{N_i(u) : 0 \leq u \leq \tau\}$ for the events and a predictable zero-one process $\{H_i(u) : 0 \leq u \leq \tau\}$ indicating whether the subject is at risk of experiencing the event. The authors proposed to model the joint model with an unobserved zero-mean bivariate Gaussian process $W_i(t) = \{W_{1i}(t), W_{2i}(t)\}$ independent in different subjects. The two sub-models are:

1. measurement sub-model composed by:

$$Y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + Z_{ij} \tag{4.19}$$

   where $Z_{ij} \sim N(0, \sigma_z^2)$ is a sequence of errors and $\mu_i(t_{ij})$ is the mean response and assumed that can be described by a linear model $\mu_i(t_{ij}) = x_{1i}(t)'\beta_1$, where $x_{1i}(t)$ are the explanatory variables and β_1 are their regression coefficients.

2. intensity sub-model composed by:

$$\lambda_i(t) = H_i(t)\alpha_0(t)\exp[x_{2i}(t)'\beta_2 + W_{2i}(t)] \tag{4.20}$$

   where $x_{2i}(t)$ are the explanatory variables and β_2 are their regression coefficients. These variables and coefficients may have some elements in common with $x_{1i}(t)$ and β_1.

The main objective of the joint analysis was to show the relation between the two latent variables W_{1i} and W_{2i}: if there is no association, the joint model is senseless. From these considerations a linear random effect model [29] was proposed $W_{1i} = U_{1i} + U_{2i}t$, in conjunction with a proportionality assumption $W_{2i} = \gamma W_{1i}$, where (U_{1i}, U_{2i}) are bivariate Gaussian random effects. It is possible to vectorise the model, thus the vector of all the subjects is:

$$W_1(t) = U_1 + U_2 t \tag{4.21}$$

where (U_1, U_2) are zero mean bivariate Gaussian variable with variances σ_1^2 and σ_2^2 and ρ correlation coefficient. So the formulation of $W_2(t)$ becomes:

$$W_2(t) = \gamma_1 U_1 + \gamma_2 U_2 + \gamma_3(U_1 + U_2 t) + U_3 \tag{4.22}$$

where $U_3 \sim N(0, \sigma_3^2)$ is independent of (U_1, U_2). The authors showed that the likelihood can be factorised in two part:

$$L = L_Y \times L_{N|Y} \tag{4.23}$$

and it is possible to obtain the estimation with the EM algorithm applying the conditional expectations presented by Wulfsohn and Tsiatis [62]. Lastly a simulation study was presented to show the estimation performance.
In 1997 Hogan and Laird [23] focused on the joint distribution of repeated vector Y and failure time D. The authors proposed two different models, the selection model where $f_{Y,D} = f_{D|Y} f_Y$, and the mixture model where $f_{Y,D} = f_{Y|D} f_D$.
Similarly in 2001 Xu e Zeger [64] focused on the joint probability, so the authors analysed how the distribution of the responses change with covariates $[T, Y|X]$ where T are times to clinical events, Y are repeated measures, and X are the covariates. The authors recalled the three classes of model already presented [23]: selection model $[T, Y|X] = [T|Y, X][Y|X]$, pattern-mixture model $[T, Y|X] = [Y|T, X][T|X]$, and latent variable model $[T, Y|X] = \int [T, Y|\eta, X] d[\eta|x]$ where η is a unobserved latent variable. Subsequently the authors focused on the latest and use a Markov chain Monte Carlo algorithm to estimate the model parameters and the distributions.
Likewise in 2003 Slasor and Laird [49] used the mixture model where the authors proposed to use $f(T|Z)f(Y|T, Z)$ for the conjoint distribution and $f(T|Z)$ was modelled as a piecewise exponential distribution. Through a simulation study and an application, the authors showed that this hypothesis gives more efficient results than using the standard hazard model and the gain in efficiency is higher than the gain of the other joint models that suppose a multinomial distribution instead of a piecewise exponential distribution.
In 2001 Wang and Taylor [57] proposed a classical joint model of two linked sub-models for longitudinal and survival data, but they added an integrated Ornstein-Uhlenbeck stochastic process, so the model for longitudinal data becomes:

$$\begin{cases} Y_i(t_{ij}) = Z_i(t_{ij}) + e_i(t_{ij}) \\ Z_i(t) = a_i + bt + \beta X_i(t) + W_i(t) \end{cases} \tag{4.24}$$

where $W_i(t)$ is the integrated Ornstein-Uhlenbeck stochastic process that accounts for the random fluctuation of the marker around the population average. Subsequently the authors used a Bayesian techniques for the estimations.
Tsiatis and Davidian [53] in 2001 proposed a simpler method for estimating the proportional hazards model parameters that requires no assumption on the distribution of random effects. The authors proposed to use the conditional score method, which is a semi parametric method. At first the authors estimated the baseline hazard and replacing it in the estimating equations, it is possible to estimate the other parameters by the large sample theory. The authors verified the model with a simulation study, also if they obtain that in some situations the results are not so good.
In 2008 Diggle, Sousa and Chetwynd [13] presented a review of the methods and described in detail a fully parametric approach. The authors presented three approaches for the joint model: the random effects model [62, 64], the semi parametric models and a simple transformation model, which is

a fully parametric model where $(Y, \log F) \sim MVN(\mu, \Sigma)$, where Y are the repeated measures while F are the times to event.
When the approaches are compared the choice depends on the context. If the aim of the research is to analyse the individual effect, the first model is the best also if it is computationally intense and it requests some assumptions concerning the distribution, while the second model offers a sharp focus on properties of the mean response profiles of the measurement process, delivering consistent estimation under minimal assumptions and self-evident robustness to non-Gaussian behaviour. Lastly the authors focused on the third method as it is the simplest one because the likelihood based inference, and the accommodation of both interval-censored and right-censored event times are simpler.
In 2010 Ibrahim, Chu and Chen [27] presented basic concepts of the joint model of longitudinal and survival data, using the classical formulation of a joint model, composed by the linear mixed model and the hazard function. In addition a review of the applications already presented concerning the HIV/AIDS and the cancer data set are exposed.
In same year Wu, Liu and Hu [59] proposed a new method of estimation for the parameters, using the Laplace approximation and the Monte Carlo EM algorithm (MCEM). The MCEM was first proposed for the joint models in 2008 from Wu, Hu e Wu [58]. The EM algorithm iterates between an E-step and an M-step: the E-step computes the conditional expectation of the complete data log-likelihood given the observed data, and the M-step gives updated parameter estimates by the conditional expectation in the E-step. When the conditional expectation in the E-step is difficult to evaluate analytically, Monte Carlo approximations may be used, which leads to a Monte Carlo EM algorithm.
In 2011 Sweeting and Thompson [51] proposed a joint model with shared random effects and made a comparison with a two-stage approach through a simulation study. The authors performed a Bayesian approach to joint modelling the two sub-models using MCMC methods.
In 2013 McCrink, Marshall and Cairns [34] presented a review that highlights the benefits of joint modelling, introducing the possible survival submodels and the different estimation methods. Two different formulation of the joint model are presented:

1. in the first case the time-to-event process is influenced by a longitudinal time dependent covariate that is measured with error [43]:

$$h_i(t) = h_0(t) \exp\{x_{2i}\beta_2 + \alpha m_i(t)\} \tag{4.25}$$

   where x_{2i} represents the baseline covariates with the corresponding regression parameters β_2, $h_0(t)$ represents the baseline hazard, and α represents the effect of the true longitudinal response $m_i(t)$ on the survival process.

2. in the second case the longitudinal process assumes informative censoring or the focus is on both process, so the longitudinal random

effects is inserted into the survival function:

$$h_i(t) = h_0(t) \exp\{x_{2i}\beta_2 + \gamma_0 U_{0i} + \gamma_1 U_{1i} t\} \tag{4.26}$$

where U_{0i} and U_{1i} represent the random intercept and slope effects and γ_0 and γ_1 represent their effect on the survival process.

In addition the authors showed an application using two different packages of the software R, the JM and joineR package.
In 2015 Barrett, Diggle, Henderson and Taylor-Robinson [3] proposed a discretisation of the time scale of time-to-event outcome, using a probit model for the discrete hazard function. The aim of the study was to reduce the computational time because, as others researchers have already shown, the joint model is very computationally demanding. The authors highlighted that there was of course a lost of information due to the decision to apply a discretisation to the survival observations, but a simulation and an application study showed that this lost was not so big, while the reduction in the computational time was high with respect to the time needed for the classical method.
Tsiatis and Davidian [54] presented one of the best overview of the joint models in 2004. The authors showed how the longitudinal process $X_i(u)$ could be express as a function of the time and of some parameters, as several researchers have already done:

$$X_i(u) = f(u)'\alpha_i \tag{4.27}$$

In addition a mean-zero stochastic process $U_i(u)$ can be considered:

$$X_i(u) = f(u)'\alpha_i + U_i(u) \tag{4.28}$$

and the use of hypothesis on the distribution of random effects is very common. Concerning the survival data, an hazard function in relation with the observed longitudinal values is proposed.
The authors mentioned the estimation approaches presented before, and then suggested two other methods: the semi parametric likelihood and the conditional score method [53], which are very easy to compute.
Some of the joint model presented above can be classified as shared parameter model, which is a type of model where the two sub-models share the same random effects. This model is often use when there is informative right censoring [60], informative missing data [17] or to model drop out mechanism [33].

4.2 Rizopoulos' formulation

The joint model presented by Rizopoulos [43] is formed by two sub-models, one sub-model is the survival model as a function of the $m_i(t)$ which denote the true and unobserved value of the longitudinal outcome and the other

one is a longitudinal mixed model.
The survival sub-model is expressed as a function of $m_i(t)$:

$$\begin{aligned} h_i(t|M_i(t),\omega_i) &= \lim_{dt\to 0} \frac{p[t \le T_i^* < t+dt | T_i^* \ge t, M_i(t), \omega_i]}{dt} \\ &= h_0(t)\exp[\gamma'\omega_i + \alpha m_i(t)], \qquad t > 0 \end{aligned} \tag{4.29}$$

where $M_i(t) = \{m_i(s), 0 \le s < t\}$ denotes the history of the true unobserved longitudinal process up to time t, T_i^* is the true event time for the subject i, T_i is the observed event time, defined as the minimum of the potential censoring time C_i and T_i^*, and α quantifies the effect of the underlying longitudinal outcome to the risk of an event. The other elements are the same introduces in the classical Cox model, where $h_0(t)$ indicates the baseline hazard function and ω_i are the covariates that influence the risk of the event with coefficient γ. In addition $\exp(\gamma_j)$ denotes the ratio of hazards for a one unit change in ω_{ij} at time t, and $\exp(\alpha)$ denotes the relative increase in the risk for an event at time t that results from a one unit increase in $m_i(t)$ at the same time point [43].
Contrarily to the classical survival model, the evaluation of the survival function is given by:

$$\begin{aligned} S_i(t|M_i(t),\omega_i) &= p[T_i^* > t | M_i(t), \omega_i] \\ &= \exp\left\{ -\int_0^t h_0(s)[\gamma'\omega_i + \alpha m_i(s)]ds \right\} \end{aligned} \tag{4.30}$$

The baseline risk function $h_0(\cdot)$ must be specified, otherwise an underestimation of standard error (SE) could be found. To overcome this underestimation, the first option is to define the baseline risk function as a known parametric distribution, for example Weibull, log-normal, or Gamma. Alternatively, a parametric but flexible specification could be applied, such as step-function and linear splines, B-spline approximation, or cuBIC splines. Another option is the piecewise-constant model:

$$h_0(t) = \sum_{q=1}^{Q} \xi_q I(\upsilon_{q-1} < t \le \upsilon_q) \tag{4.31}$$

where $0 = \upsilon_0 < \upsilon_1 < ... < \upsilon_Q$ denotes a split in the time scale, and ξ_q denotes the value of the hazard in the interval $(\upsilon_{q-1}, \upsilon_q]$. Alternately a regression spline model could be used:

$$\log h_0(t) = \kappa_0 + \sum_{d=1}^{m} \kappa_d B_d(t,q) \tag{4.32}$$

where $\kappa' = (\kappa_0, \kappa_1, ..., \kappa_m)$ are the spline coefficient, q denotes the degree of the basis functions of the B-splines $B(\cdot)$, and $m = \ddot{m} + q - 1$ where $\ddot{m}$ denotes the number of interior knots.

Concerning the longitudinal sub-model, Rizopoulos [43] proposed a linear mixed model:

$$\begin{cases} y_i(t) = m_i(t) + \epsilon_i(t) \\ m_i(t) = x_i'(t)\beta + z_i'(t)b_i \\ \epsilon_i(t) \sim N(0, \sigma^2) \\ b_i \sim N(0, D) \\ b_1, ..., b_N, \epsilon_1, ..., \epsilon_N \quad \textit{independent} \end{cases} \tag{4.33}$$

Note that $y_i(t)$ is composed by the $m_i(t)$ and a random error term, and as in the classical longitudinal model, β are the fixed effects for $x_i(t)$, and b_i are the random effects for $z_i(t)$. The author highlighted that sometimes subjects show highly non-linear longitudinal trajectories, in these cases a flexible representations for the covariates was introduced. An alternative formulation, that helps to consider the highly non linear shapes, is based on adding a stochastic term that aims to capture the remaining serial correlation:

$$y_i(t) = m_i(t) + u_i(t) + \epsilon_i(t) \tag{4.34}$$

where $u_i(t)$ is a mean-zero stochastic process independent of the random effects and of the error term.

4.3 Rizopoulos' estimation

Rizopoulos [43] proposed two kinds of estimation, the two-stage approach and the joint likelihood formulation. The two-stage approach is biased but less computationally demanding, while the joint likelihood is more efficient but computationally slower.
The two-stage approach is based on two steps. In the first one: the random effects are estimated using a least-squares approach, while in the second step the estimates previously found are used to impute appropriate values of $m_i(t)$ that are substituted in the classical partial likelihood of the Cox model. The joint likelihood could be based on maximum likelihood, or a Bayesian estimation of joint models using MCMC, or some hypothesis concerning the normal distribution of random effects or of covariates.
Rizopoulos [43] proposed a new method of estimation based on the joint likelihood formulation. The author supposed that the vector of random effects b_i underlies both the longitudinal and survival processes. As a results:

$$p(T_i, \delta_i, y_i|b_i; \theta) = p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta) \tag{4.35}$$

and

$$p(y_i|b_i; \theta) = \prod_j p[y_i(t_{ij})|b_i; \theta] \tag{4.36}$$

where $\delta_i = I(T_i^* \leq C_i)$ is the event indicator and $\theta = (\theta_t', \theta_y', \theta_b')'$ denotes the full parameter vector, with θ_t denoting the parameters for the event time outcome, θ_y the parameters for the longitudinal outcomes, and θ_b the unique parameters of the random-effects covariance matrix.

In addition some assumptions are made: given the observed history, the censoring mechanism and the visiting process are independent of the true event times and of the future longitudinal measurements. Where the visiting process is the mechanism that generates the time points at which longitudinal measurements are collected, and the observed history, for any time point t, is formed by all available information for the longitudinal process prior to t. These assumptions mean that the possible withdrawn of a subject depends only on the observed past history.
Accordingly the log-likelihood contribution for the subject i is:

$$\begin{aligned}\log p(T_i,\delta_i, y_i;\theta) &= \log \int p(T_i, \delta_i, y_i, b_i;\theta) db_i \\ &= \log \int p(T_i, \delta_i|b_i;\theta_t,\beta) \left\{ \prod_j p[y_i(t_{ij})|b_i;\theta_y] \right\} p(b_i;\theta_b) db_i \end{aligned} \tag{4.37}$$

obtained by the equations before, and where:

$$\begin{aligned} p(T_i, \delta_i|b_i;\theta_t,\beta) =& h_i[T_I|M_i(T_i);\theta_t,\beta]^{\delta_i} S_i[T_i|M_i(T_i);\theta_t,\beta] \\ =& \{h_0(T_i)\exp[\gamma'\omega_i + \alpha m_i(T_i)]\}^{\delta_i} \\ & \times \exp\left\{ -\int_0^{T_i} h_0(s)\exp[\gamma'\omega_i + \alpha m_i(s)] ds \right\} \end{aligned} \tag{4.38}$$

where $h_0(\cdot)$ is the baseline hazard. In addition:

$$\begin{aligned} p(Y_i|b_i;\theta)p(b_i;\theta) =& \prod_j p[y_i(t_{ij})|b_i;\theta_y] p(b_i;\theta_b) \\ =& (2\pi\sigma^2)^{n_i/2} \exp\left[-\frac{\| y_i - X_i\beta - Z_i b_i \|^2}{2\sigma^2} \right] \\ & \times (2\pi)^{q_b/2} \det(D)^{-1/2} exp\left(-\frac{b_i' D^{-1} b_i}{2} \right) \end{aligned} \tag{4.39}$$

where $\| \cdot \|$ denotes the Euclidean vector norm.
For the maximization standard algorithm can be used, like EM or Newton-Raphson algorithm. In order to make the estimation easier the score vector

corresponding to $l(\theta)$ can be so simplified:

$$\begin{aligned}
S(\theta) =& \sum_i \frac{\partial}{\partial \theta'} \log \int p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\
=& \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \frac{\partial}{\partial \theta'} \int p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta) db_i \\
=& \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \int \frac{\partial}{\partial \theta'} [p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta)] db_i \\
=& \sum_i \int \left\{ \frac{\partial}{\partial \theta'} \log [p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta)] \right\} \\
& \times \frac{p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) p(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)} db_i \\
=& \sum_i \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i
\end{aligned} \tag{4.40}$$

where $A(\theta, b_i) = \partial [\log p(T_i, \delta_i | b_i; \theta) + \log p(y_i | b_i; \theta) + \log p(b_i; \theta)] / \partial \theta'$. Rizopoulos [43] showed besides the EM algorithm applied to the joint model, beginning from the joint model:

$$\begin{cases}
h_i(t) = h_0(t) \exp \{\gamma' \omega_i + \alpha [x_i'(t) \beta + z_i'(t) b_i]\} \\
y_i(t) = x_i'(t) \beta + z_i'(t) b_i + \epsilon_i(t) \\
b_i \sim N(0, D) \\
\epsilon_i \sim N(0, \sigma^2)
\end{cases} \tag{4.41}$$

where $\theta = (\theta_t', \theta_y', \theta_b')'$, with $\theta_y = (\beta', \sigma^2)'$, $\theta_t = (\gamma', \alpha', \theta_{h_0}')'$ where θ_{h_0} denote the parameters in the baseline risk function $h_0(\cdot)$, and $\theta_b = vech(D)$. For the E-step this equation can be used:

$$\begin{aligned}
Q(\theta | \theta^{(it)}) =& E[\log p(y; \theta) | y^0; \theta^{(it)}] \\
=& \int p(y^m, y^o; \theta) p(y^m | y^o; \theta^{(it)}) dy^m \\
=& \sum_i \int \log p(T_i, \delta_i, y_i, b_i; \theta) p(b_i | T_i, \delta_i, y_i; \theta^{(it)}) db_i \\
=& \sum_i \int [\log p(T_i, \delta_i | b_i; \theta_t) + \log p(y_i | b_i; \theta_y) \\
& + \log p(b_i; \theta_b)] p(b_i | T_i, \delta_i, y_i; \theta^{(it)}) db_i
\end{aligned} \tag{4.42}$$

where y^o is the vector of observed data, while y^m is the vector of missing data. For the M-step the log-likelihood could be split into three part:

$$\log p(T_i, \delta_i, y_i, b_i; \theta) = \log p(T_i, \delta_i | b_i; \theta_t) + \log p(y_i | b_i; \theta_y) + \log p(b_i; \theta_b)$$

From the maximization of $Q(\theta|\theta^{(it)})$ it is possible to obtain:

$$\begin{aligned}\hat{\sigma}^2 &= N^{-1}\sum_i \int (y_i - X_i\beta - Z_i b_i)'(y_i - X_i\beta - Z_i b_i)p(b_i|T_i,\delta_i,y_i;\theta)db_i \\ &= N^{-1}\sum_i (y_i - X_i\beta)'(y_i - X_i\beta - 2Z_i\tilde{b}_i) + tr(Z_i'Z_i\tilde{\nu b}_i) + \tilde{b}_i'Z_i'Z_i\tilde{b}_i \\ \hat{D} &= n^{-1}\sum_i \nu\tilde{b}_i + \tilde{b}_i\tilde{b}_i{}'\end{aligned}$$

where $N = \sum_i n_i$, $\tilde{b}_i = E(b_i|T_i,\delta_i,y_i;\theta^{(it)}) = \int b_i p(b_i|T_i,\delta_i,y_i;\theta^{(it)})db_i$, and $\nu\tilde{b}_i = var(b_i|T_i,\delta_i,y_i;\theta^{(it)}) = \int (b_i - \tilde{b}_i)^2 p(b_i|T_i,\delta_i,y_i;\theta^{(it)})db_i$. For the remaining parameters there is no close solution, so an iteration solution must be considered:

$$\begin{aligned}\hat{\beta}^{(it+1)} &= \hat{\beta}^{(it)} - [\partial S(\hat{\beta}^{(it)})/\partial\beta]^{-1}S(\hat{\beta}^{(it)}) \\ \hat{\theta}_t^{(it+1)} &= \hat{\theta}_t^{(it)} - [\partial S(\hat{\beta}^{(it)})/\partial\beta]^{-1}S(\hat{\theta}_t^{(it)})\end{aligned}$$

where $\hat{\beta}^{(it)}$ and $\hat{\theta}_t^{(it)}$ denotes the values of β and θ_t at the current iteration and $\partial S(\hat{\beta}^{(it)})/\partial\beta$ and $\partial S(\hat{\beta}^{(it)})/\partial\beta$ represents the corresponding blocks of the Hessian matrix. The score vectors become:

$$\begin{aligned}S(\beta) =& \sum_i X_i'(y_i - X_i\beta - Z_i\tilde{\beta}_i)/\sigma^2 + \alpha\delta_i x_i(T_i) \\ & - \exp(\gamma'\omega_i)\int\int_0^{T_i} h_0(s)\alpha x_i(s)\exp\{\alpha[x_i'(s)\beta + z_i'(s)b_i]\} \\ & \times p(b_i T_i,\delta_i,y_i;\theta)dsdb_i \\ S(\gamma) =& \sum_i \omega_i\Bigg\{\delta_i - \exp(\gamma'\omega_i)\int\int_0^{T_i} h_0(s)\exp[\alpha(x_i'(s)\beta + z_i'(s)b_i)] \\ & \times p(b_i|T_i,\delta_i,y_i;\theta)dsdb_i\Bigg\}\end{aligned}$$

$$\begin{aligned}S(\alpha) =& \sum_i \delta_i[x_i'(T_i)\beta + z_i'(T_i)\tilde{b}_i] \\ & - \exp(\gamma'\omega_i)\int\int_0^{T_i} h_0(s)[x_i'(s)\beta + z_i'(s)b_i]\exp\alpha[x_i'(s)\beta + z_i'(s)b_i] \\ & \times p(b_i|T_i,\delta_i,y_i;\theta)dsdb_i] \\ S(\theta_{h_0}) =& \sum_i \delta_i\frac{\partial log h_0(T_i;\theta_{h_0})}{\partial\theta_{h_0}'} \\ & - \exp(\gamma'\omega_i)\int\int_0^{T_i}\frac{\partial h_0(s;\theta_{h_0})}{\theta_{h_0}'}\exp\{\alpha[x_i'(s)\beta + z_i'(s)b_i]\} \\ & \times p(b_i|T_i,\delta_i,y_i;\theta)dsdb_i\end{aligned}$$

From these estimations and from the score vector $S(\theta)$, the Hessian matrix can be evaluated from the fact that:

$$\begin{aligned}\frac{\partial S_i(\theta)}{\partial \theta} =& \frac{\partial}{\partial \theta}\int A(\theta, b_i)p(b_i|T_i, \delta_i, y_i; \theta)db_i \\ =& \int \frac{A(\theta, b_i)}{\partial \theta}p(b_i|T_i, \delta_i, y_i; \theta)db_i \\ &+ \int A(\theta, b_i)\frac{p(b_i|T_i, \delta_i, y_i; \theta)}{\partial \theta}db_i\end{aligned} \tag{4.43}$$

In addition it is possible to pose:

$$\begin{aligned}I_1 =& \int A(\theta, b_i)\frac{p(b_i|T_i, \delta_i, y_i; \theta)}{\partial \theta}db_i \\ =& \int A(\theta, b_i)\left\{\frac{\log p(b_i|T_i, \delta_i, y_i; \theta)}{\partial \theta}\right\}' p(b_i|T_i, \delta_i, y_i; \theta)db_i \\ =& \int A(\theta, b_i)\Bigg\{\frac{\partial[\log p(T_i, \delta_i|b_i; \theta) + \log p(y_i|b_i; \theta) + \log p(b_i; \theta)]}{\partial \theta} \\ &- \frac{\partial \log p(T_i, \delta_i, y_i; \theta)}{\partial \theta}\Bigg\}' p(b_i|T_i, \delta_i, y_i; \theta)db_i \\ =& \int A(\theta, b_i)[A(\theta, b_i) - S_i(\theta)]' p(b_i|T_i, \delta_i, y_i; \theta)db_i\end{aligned}$$

Through this the standard errors for the parameters estimates can be easily estimated as:

$$\hat{var} = [I(\hat{\theta})]^{-1}$$

where

$$I(\hat{\theta}) = -\sum_{i=1}^{n}\frac{\partial S_i(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}}$$

Focusing on the estimation of the random effects, a Bayesian paradigm can be used, assuming that $p(b_i; \theta)$ is the prior distribution, and that the conditional likelihood part is $p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta)$, it is possible to obatin the posterior distribution:

$$\begin{aligned}p(b_i|T_i, \delta_i, y_i; \theta) &= \frac{p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta)p(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)} \\ &\propto p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta)p(b_i; \theta)\end{aligned} \tag{4.44}$$

As the number of longitudinal measurement increases, this distribution will converge to a normal distribution. For the estimation the mean or the mode can be used:

$$\begin{aligned}\overline{b}_i &= \int b_i p(b_i|T_i, \delta_i, y_i; \theta)db_i \\ \hat{b}_i &= \arg\max \log p(b|T_i, \delta_i, y_i; \theta)\end{aligned} \tag{4.45}$$

In 2009 Rizopoulus, Verbeke and Lesaffre [45] proposed another method of estimation, the Laplace approximation, using this method of approximation in the M-step of the EM algorithm in order to approximate the integral. Later, in 2012, Rizopoulous [42] presented another numerical method of estimation, the Gauss-Hermite quadrature rule, but this method is too much computationally demanding, so a new method was proposed, the pseudo-adaptive Gauss-Hermite quadrature. The main idea behind this rule is to re-scale and re-center the integrand for each subject using information from a separate fit of the linear mixed effects models for the longitudinal outcome, this considerably decreases the computational burden without sacrificing accuracy.

4.4 Asymptotic inference

Through the asymptotic inference is possible to analysis the hypothesis testing for the parameters. The classical hypothesis is:

$$H_0 : \theta = \theta_0 \qquad versus \qquad H_\alpha : \theta \neq \theta_0 \tag{4.46}$$

Rizopoulos [43] proposed to use three different tests:

1. Likelihood Ratio Test with the test statistic defined as:

   $$LRT = -2[l(\hat{\theta}_0) - l(\hat{\theta})]$$

   where $\hat{\theta}_0$ and $\hat{\theta}$ denote the maximum likelihood estimates under the null and alternative hypothesis.

2. Score test with the test statistic defined as:

   $$U = S'(\hat{\theta}_0)[I(\hat{\theta}_0)]^{-1}S(\hat{\theta}_0)$$

   where $S(\cdot)$ and $I(\cdot)$ denote the score function and the observed information matrix of the model under the alternative hypothesis.

3. Wald test with the test statistic defined as:

   $$W = (\hat{\theta} - \theta_0)'[I(\hat{\theta})](\hat{\theta} - \theta_0)$$

Under the null hypothesis, the asymptotic distribution of each of these tests is a chi-squared distribution with p degrees of freedom, where p denoting the number of parameters being tested. For a single parameter θ_j the Wald test is equivalent to $(\hat{\theta}_j - \theta_{0j})/\hat{se}(\hat{\theta}_j)$, which under the null hypothesis follows an asymptotic standard normal distribution. These tests are asymptotically equivalent but the likelihood test ratio is preferred than the Wald test because in finite samples its chi-squared distribution is not certain.
In addition a linear combination of coefficients can be tested, so the hypothesis is:

$$H_0 : L\theta = 0 \qquad versus \qquad H_\alpha : L\theta \neq 0 \tag{4.47}$$

where L specifies the linear combination of coefficients tested.
Another type of hypothesis test is based on the comparison of two nested models. There are two common tests, Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC):

$$AIC = -2l(\hat{\theta}) + 2n_{par}$$
$$BIC = -2l(\hat{\theta}) + n_{par}\log(n)$$

where n_{par} denotes the number of parameters in the model. For these tests the rule is "smaller is better". The differences of the two tests is based on the fact that AIC tends to select more elaborate models than BIC because the latter penalizes much more heavily for the complexity of the model.
Through the asymptotic inference is also possible to built the confidence intervals. The asymptotic 95% confidence intervals for the parameter of interest can be based on Wald statistics, so it is $\hat{\theta} \pm 1.96\hat{se}(\hat{\theta})$.
In addition the asymptotic confident intervals for the fitted values can be based on the asymptotic normal distribution of the MLEs. For example Rizopoulos [43] proposed the asymptotic confidence intervals for the average longitudinal evolutions $\mu = X\beta$ in the longitudinal process:

$$\hat{\mu} \pm 1.96\hat{se}(\hat{\mu})$$
$$X\hat{\beta} \pm diag[X\hat{var}(\hat{\beta})X']^{1/2}$$

where X denotes the design matrix of interest, and $\hat{var}(\hat{\beta})$ the block of observed Hessian matrix corresponding to $\hat{\beta}$.

4.5 Diagnostic

Rizopoulos [43] recalled that there are different type of residuals because there are different models involved, residuals for the longitudinal sub-model and residuals for the survival sub-model.
The first type of residuals are split in two, the subject specific and the marginal residuals. The subject specific residuals aim to validate these assumptions:

$$\begin{cases} y_i & = X_i\beta + Z_i b_i + \epsilon_i \\ b_i & \sim N(0, D) \\ \epsilon_i & \sim N(0, \sigma^2) \end{cases} \tag{4.48}$$

and the corresponding residuals are:

$$r_i^{ys}(t) = [y_i(t) - x_i'(t)\hat{\beta} - z_i'(t)\hat{b}_i] \tag{4.49}$$

with the corresponding standardized version:

$$r_i^{yss}(t) = \frac{[y_i(t) - x_i'(t)\hat{\beta} - z_i'(t)\hat{b}_i]}{\hat{\sigma}}$$

On the other hand the marginal residuals focus on validate the marginal model assumptions:

$$\begin{cases} y_i &= X_i\beta + \epsilon_i^* \\ \epsilon_i^* &\sim N(0, Z_i D Z_i' + \sigma^2 I_{n_i}) \end{cases} \tag{4.50}$$

and the corresponding residuals are:

$$r_i^{ym} = y_i - X_i\hat{\beta} \tag{4.51}$$

with the corresponding standardized version:

$$r_i^{ysm}(t) = \hat{V}_i^{-1/2}(y_i - X_i\hat{\beta})$$

where $\hat{V}_i = Z_i\hat{D}Z_i' + \hat{\sigma}^2 I_{n_i}$ denotes the estimated marginal covariance matrix of y_i.
Concerning the residuals for the survival data, the martingale residuals can be used:

$$\begin{aligned} r_i^{tm}(t) =& N_i(t) - \int_0^t R_i(s) h_i(s|\hat{M}_i(s); \hat{\theta}) ds \\ =& N_i(t) - \int_0^t R_i(s) \hat{h}_0 \exp[\hat{\gamma}'\omega_i + \hat{\alpha}\hat{m}_i(s)] ds \end{aligned} \tag{4.52}$$

where $N_i(t)$ is the counting process denoting the number of events for subject i by time t, $R_i(t)$ is the left continuous at risk process with $R_i(t) = 1$ if subject i is at risk at time t, 0 otherwise, $\hat{m}_i(t) = x_i'(t)\hat{\beta} + z_i'(t)\hat{b}_i$, and $\hat{h}_0(\cdot)$ denotes the estimated baseline risk function. The aims of these residuals is to identify excess events and to evaluate the functional form for the covariate.
An alternative type of residuals for the survival model is the Cox-Snell residuals:

$$\begin{aligned} r_i^{tcs} &= \int_0^{T_i} h_i(s|\hat{M}_i(s); \hat{\theta}) ds \\ &= \int_0^{T_i} \hat{h}_0(s) \exp[\hat{\gamma}\omega_i + \hat{\alpha}\hat{m}_i(s)] ds \end{aligned} \tag{4.53}$$

and thus $r_i^{tcs} = N_i(T_i) - r_i^{tm}(T_i)$.
Subsequently Rizopoulos [43] presented several plots that are able to investigate the residuals and to display the trend of the observed data.
The non-random drop-out in the longitudinal outcome can be a problem, so Rizopoulos, Verbeke, and Molenberghs [46] noted that the reference distribution of statistics, such as the residuals, in missing data settings is not directly available and complex calculations are required to derive it, so a multiple-imputation-based approach can be used. The basis of the method is the hypothesis that visit times were pre-specified by the protocol and all patients adhere to them.

Dobson and Henderson [14] presented a paper concerning the diagnostics of joint longitudinal and drop-out time modelling. The authors had three aims, at first to explore if there is any association between responses and drop-out time, secondary to study conditional residual analysis methods for longitudinal data with drop-out, and lastly analyse the individual case influence for joint modelling.
The robustness of the model is very important, thus Li, Elashoff and Li [31] proposed to analyse the robustness of the joint model and proposed a linear mixed effects sub-model for the longitudinal outcome and a proportional cause-specific hazards frailty sub-model for the competing risks data, linked together by latent random effects. Instead of the usual normality assumption for measurement errors in the linear mixed effects sub-model, the authors adopted a t-distribution which has a longer tail and thus is more robust to outliers.
Also Hsieh, Tseng and Wang [25] focused on the robustness, in fact they reviewed the merits of the joint modelling approach in Wulfsohn and Tsiatis [62] by providing a theoretical explanation of the robustness features observed in the literature. In addition the authors demonstrated the missing information and implicit profile features in joint modelling, and proposed to use the Fisher information for estimating the standard errors of the EM estimators.

4.6 Predictions and accuracy

As there are two types of residuals, there are also two types of predictions. The first type concerns the dynamic predictions of survival probabilities. The aim is to predicting survival probabilities for a new subject i that has provided a set of longitudinal measurements $Y_i(t) = \{y_i(s); 0 \leq s < t\}$, so Rizopoulos [43, 41] focused on the conditional probability of surviving time $u > t$ given survival up to t:

$$\pi_i(u|t) = p[T_i^* \geq u | T_i^* > t, Y_i(t), \omega_i, D_n; \theta^*] \qquad t > 0 \tag{4.54}$$

where ω_i denotes the baseline covariates, θ^* denotes the true parameter values, and $D_n = \{T_i, \delta_i, y_i; i = 1, ..., n\}$ denotes the random sample. Then the probability can be formulated as:

$$\begin{aligned} & p[T_i^* \geq u | T_i^* < t, Y_i(t); \theta] \\ & \quad = \int p(T_i^* \geq u | T_i^* > t, Y_i(t), b_i; \theta) p(b_i | T_i^* > t, Y_i(t); \theta) db_i \\ & \quad = \int p(T_i^* \geq u | T_i^* > t, b_i; \theta) p(b_i | T_i^* > t, Y_i(t); \theta) db_i \\ & \quad = \int \frac{S_i\left[u | M_i(u, b_i, \theta); \theta\right]}{S_i\left[t | M_i(t, b_i, \theta); \theta\right]} p(b_i | T_i^*, Y_i(t); \theta) db_i \end{aligned} \tag{4.55}$$

where $S_i(\cdot)$ denotes the survival function and $M_i(\cdot)$ indicates the longitudinal story. In order to estimate this function Rizopoulos [43, 41] used the

empirical Bayes estimates for b_i:

$$\tilde{\pi}_i(u|t) = \frac{S_i\left[u\middle|M_i(u, \hat{b}_i^{(t)}, \hat{\theta}); \hat{\theta}\right]}{S_i\left[t\middle|M_i(t, \hat{b}_i^{(t)}, \hat{\theta}); \hat{\theta}\right]} + O\{[n_i(t)]^{-1}\}$$

where $\hat{\theta}$ denotes the maximum likelihood estimates, $\hat{b}_i^{(t)}$ denotes the mode of the conditional distribution of $\log p(b_i|T_i^* < t, Y_i(t); \hat{\theta})$, and $n_i(t)$ denotes the number of longitudinal responses for subject i by time t.
In addition the author suggested to consider the posterior expectation of $\pi_i(u|t)$:

$$p(T_i^* \geq u|T_i^* > t, Y_i(t), D_n) = \int p(T_i^* \geq u|T_i^* > t, Y_i(t); \theta)p(\theta|D_n)d\theta$$

The estimations are then possible through the Monte Carlo method with a simulation scheme.
The other type of predictions concerns the dynamic prediction for the longitudinal outcome [43, 41]. In particular, for a specific subject i who is still alive by follow-up time t, the interest is in the expected value of the longitudinal outcome at time $u > t$ given the observed responses up to that time point $Y_i(t)$, which is formulated as:

$$\omega_i(u|t) = E[y_i(u)|T_i^* > t, Y_i(t), D_n; \theta^*] \qquad u > t \tag{4.56}$$

For estimating it, Rizopoulos [43] suggested again to use the Bayesian method, and calculated the expectation of $\omega_i(u|t)$ with respect to the posterior distribution of the parameter $\{\theta|D_n\}$:

$$E[y_i(u)|T_i^* > t, Y_i(t), D_n] = \int E[y_i(u)|T_i^* > t, Y_i(t); \theta]p(\theta|D_n)d\theta \tag{4.57}$$

In addition it must be considered that:

$$\begin{aligned}
&E[y_i(u)|T_i^* > t, Y_i(t); \theta] \\
&\quad = \int E[y_i(u)|T_i^* > t, Y_i(t), b_i; \theta]p(b_i|T_i^* > t, Y_i(t); \theta)db_i \\
&\quad = \int E[y_i(u)|b_i]p(b_i|T_i^* > t, Y_i(t); \theta)db_i \\
&\quad = [x_i'(u)\beta + z_i'(u)b_i]p(b_i|T_i^* > t, Y_i(t); \theta)db_i \\
&\quad = x_i'(u)\beta + z_i'(u)\overline{b}_i^{(t)}
\end{aligned}$$

where:

$$\overline{b}_i^{(t)} = \int b_i p(b_i|T_i^* > t, Y_i(t); \theta)db_i$$

Under these assumptions a straightforward estimator of $\omega_i(u|t)$ was obtained by replacing θ with $\hat{\theta}$, and calculating the mean of the posterior distribution

of $p(b_i|T_i^* > t, Y_i(t); \hat{\theta})$. In the same spirit, a similar estimator is derived when instead of the mean $\overline{b}_i^{(t)}$ of the posterior distribution its mode $\hat{b}_i = \arg\max_b \log p(b|T_i^* > t, Y_i(t); \hat{\theta})$ is used, so:

$$\tilde{\omega}_i(u|t) = x_i'(u)\hat{\beta} + z_i'(u)\hat{b}_i^{(t)} + O(n_i^{-1})$$

considering the relation: $\overline{b}_i^{(t)} = \hat{b}_i^{(t)} + O\{[n_i(t)]^{-1}\}$. From this the estimations are possible through the Monte Carlo method with a simulation scheme.
The time-dependent accuracy measures for the longitudinal markers [43, 41, 67] are very important. The information at disposition is the collected set of longitudinal measurements $Y_i(t) = \{y_i(s); 0 \leq s < t\}$ upon time t for subject i. Considering a vector of threshold values c, these new functions can be defined:

$$\wp_i^s(t, k, c) = \{y_i(s) \geq c_s; k \leq s \leq t\}$$

as a "success", i.e., the marker indicates that the event will occur, and:

$$\wp_i^f(t, k, c) = \Re^{r(k,t)} \{y_i(s) \geq c_s; k \leq s \leq t\}$$

as a "failure", where $\Re^n$ denotes the n-dimensional Euclidean space, and $r(k, t)$ denotes the number of longitudinal measurements taken in the interval $[k, t]$. The value of $k \geq 0$ specifies which past marker values of the longitudinal history contribute to the rule, and c_s denotes the threshold value at time point s. The convention with these prediction rules is that larger values for the marker are associated with higher risk for death.
From these the sensitivity, the probability that the marker correctly classifies a subject as diseased, can be defined as:

$$TP_t^{\Delta t}(c) = p\{\wp_i^s\ (t, k, c)|T_i^* > t, T_i^* \in (t, t + \Delta t]; \theta^*\} \tag{4.58}$$

also known as the true positive rate. In addition the respectively, the probability that the marker correctly classified a subject as non-disease, can be defined as:

$$1 - FP_t^{\Delta t}(c) = p\{\wp_i^f\ (t, k, c)|T_i^* > t, T_i^* > t + \Delta t; \theta^*\} \tag{4.59}$$

where $FP_t^{\Delta t}(c)$ denotes the false positive rate.
The overall discrimination capability of the longitudinal marker for all possible thresholds $c \in \Re_y$ the corresponding Receiver Operating Characteristic (ROC) curve can be assessed:

$$ROC_t^{\Delta t}(p) = TP_t^{\Delta t}[(FP_t^{\Delta t})^{-1}(p)]$$

where p is in $[0, 1]$ and $(FP_t^{\Delta t})^{-1}(p) = \inf_c\{c : FP_t^{\Delta t}(c) \leq p\}$. In order to summarise this characteristics, the area under the ROC curve (AUC) can be used:

$$AUC_t^{\Delta t} = \int_0^1 ROC_t^{\Delta t}(p)dp$$

As Rizopoulos [43, 41] highlighted these values describe how well a marker can discriminate between patients at a specific follow-up time t. Therefore, at different time points the marker may exhibit different levels of discrimination, and thus a relevant question the author posed is how the discriminative capability of the marker over the whole follow-up period can be summarized. So a dynamic discrimination index based on the AUC was proposed:

$$C_{dyn}^{\Delta t} = \int_0^{\infty} AUC_t^{\Delta t} u(t) dt$$

where

$$u(t) = \frac{p(T_i^* > t)}{\int p(T_i^* > t) dt}$$

is the weight function and where $p(T_i^* > t)$ is the marginal survival probability. The attention will be typically restricted to a fixed follow-up period $(0, \tau)$. In this case the $C_{dyn}^{\Delta t}$ index can be modified to account for finite follow-up:

$$[C_{dyn}^{\Delta t}]^{\tau} = \int_0^{\tau} AUC_t^{\Delta t} u^{\tau}(t) dt$$

where $u^{\tau}(t) = u(t) / \int_0^{\tau} u(t) dt$. So this index is connected to the probability that the predictions for a random pair of subjects are concordant with their outcomes, but given that the smaller event time occurs within the interval $(0, \tau)$.
Focusing on the sensitivity some simplifications of the true positive rate can be made:

$$\begin{aligned} TP_t^{\Delta t}(c) &= p\{\wp_i^s(t, k, c) | T_i^* > t, T_i^* \in (t, t + \Delta t]; \theta^*\} \\ &= \frac{p\{\wp_i^s(t, k, c), T_i^* \in (t, t + \Delta t] | T_i^* > t; \theta^*\}}{1 - p(T_i^* > t + \Delta t | T_i^* > t; \theta^*)} \end{aligned}$$

where θ^* denotes the true parameter values. The numerator can be simplified and becomes:

$$\begin{aligned} & p\{\wp_i^s(t, k, c), T_i^* \in (t, t + \Delta t] | T_i^* > t; \theta^*\} \\ &= \int p\{\wp_i^s(t, k, c), T_i^* \in (t, t + \Delta t] | T_i^* > t, b_i; \theta^*\} p(b_i | T_i^* > t; \theta^*) db_i \\ &= \int p\{\wp_i^s(t, k, c) | b_i; \theta^*\} p\{T_i^* \in (t, t + \Delta t] | T_i^* > t, b_i; \theta^*\} p(b_i | T_i^* > t; \theta^*) db_i \end{aligned}$$

where

$$p\{\wp_i^s(t, k, c) | b_i; \theta^*\} = \prod_{s=k}^{t} \Phi \left[\frac{c_s - m_i(s, b_i, \beta^*)}{\sigma^*} \right]$$

and

$$p\{T_i^* \in (t, t + \Delta t] | T_i^* > t, b_i; \theta^*\} = 1 - \frac{S_i[t + \Delta t | M_i(t + \Delta t, b_i); \theta^*]}{S_i[t | M_i(t, b_i); \theta^*]}$$

While the denominator becomes:

$$\begin{aligned} & p(T_i^* > t + \Delta t | T_i^* > t; \theta^*) \\ & = \int p(T_i^* > t + \Delta t | T_i^* > t, b_i; \theta^*) p(b_i | T_i^* > t; \theta^*) db_i \\ & = \int \frac{S_i[t + \Delta t | M_i(t + \Delta t, b_i); \theta^*]}{S_i[t | M_i(t, b_i); \theta^*]} p(b_i | T_i^* > t; \theta^*) db_i \end{aligned}$$

Thus the sensitivity can be rewritten as the ratio of two expected values:

$$\begin{aligned} E_1(b_i, \theta) = & \left\{ \prod_{s=k}^{t} \Phi \left[\frac{c_s - m_i(s, b_i, \beta^*)}{\sigma^*} \right] \right\} \\ & \times \left\{ 1 - \frac{S_i[t + \Delta t | M_i(t + \Delta t, b_i); \theta^*]}{S_i[t | M_i(t, b_i); \theta^*]} \right\} \end{aligned}$$

and

$$E_2(b_i, \theta) = \frac{S_i[t + \Delta t | M_i(t + \Delta t, b_i); \theta^*]}{S_i[t | M_i(t, b_i); \theta^*]}$$

with respect to the marginal posterior distribution $p(b_i | T_i^* > t; \theta^*)$.
Following a simulation scheme this Monte Carlo estimate of sensitivity can be obtained:

$$\hat{p}\{\wp_i^s(t, k, c) | T_i^* > t, T_i^* \in (t, t + \Delta t]\} = \frac{\sum_i E_1(b_i^{(l)}, \theta^{(l)})}{L - \sum_l E_2(b_i^{(l)}, \theta^{(l)})}$$

The same procedure can be used to obtain the parallel simplified results for the specificity.

4.7 Review of the applications in biostatistic area

The joint model is very useful in medical area because in clinical trail is very interesting to analyse in the two subgroups, placebo and treated, in order to study the longitudinal covariates that influence the survival or the effect of a new drug.
The first examples regarding the study of AIDS data set. A model that study the relationship between longitudinal and survival data was proposed in 1993 [11] as an application to analyse the lymphocyte CD4 in the AIDS studies. The aim was to analyse the effect of the CD4 lymphocyte on the risk of death and the effect of an antiviral treatment, the zidovudine. The relation between CD4 lymphocytes and the hazard of death at time t showed that the risk of death is greater for patients with lower CD4 lymphocytes counts. The main usefulness of the joint model is that it permits calculation of the distributions of expected CD4-lymphocyte count among surviving patients and the expected survival distributions that correspond to a given

CD4 trajectory.
A similar application was made by DeGruttola and Tu [9, 10], defining the $\log CD4$ as:

$$\begin{aligned}\log CD4_{ij} =& \alpha_1 + \alpha_2 I_{(t_{ij}\geq 8)} + \alpha_3 (t-8) I_{(t_{ij}\geq 8)} \\ &+ b_{1i} + b_{2i} I_{(t_{ij}\geq 8)} + b_{3i} I_{(t_{ij}\geq 8)} + \epsilon_{ij}\end{aligned}$$

where $I_{(t\geq a)}$ is an indicator function, b_i are the random effects, and ϵ_{ij} are the residuals. The results showed that although the entry level CD4 count is not related to survival, greater initial rise in CD4 and less steep decline in CD4 are associated with longer survival.
Tsiatis, DeGruttola, and Wulfsohn [55] applied a two-stage model to the same data set obtaining that the hazard rate increases as CD4 declines, with the greatest effect occurring for patients with CD4 counts less than 50, they also demonstrated that the hazard rate increases as a function of time from treatment initiation even after adjusting for CD4 values.
The AIDS data set was analysed also by Self and Pawitan [48], but instead of the CD4 the ratio of T4 and T8 and the effect on the time from seroconversion to diagnosis with AIDS is studied, using a two steps method with a linearised relation between hazard and random effects.
Also Faucett and Thomas [15] used the data set composed by the data of the immunologic marker CD4 to analyse the time of diagnosis of AIDS, using the Markov chain Monte Carlo techniques of Gibbs sampling to estimate the joint posterior distribution of the unknown parameters of the model.
With reference to the same data set Wu, Hu e Wu [58] showed that patients' viral loads after initiating antiviral treatment declined in the early period and some patients' viral loads rebounded in the later period. This is probably caused by the fact that HIV virus was sensitive to the antiviral treatment in the initial period but developed drug resistance subsequently. There is a substantial variation between patients and some patients do not experience viral rebound during the study period. Some patients with faster initial viral decays appear to have earlier viral rebound. In addition smaller baseline CD4 values are associated with earlier viral rebound.
Subsequently Wu, Liu and Hu [59] suggested to consider in addition a new covariates, a joint model with shared random effects is applied to the AIDS data set in order to analyse three different models, the non-linear mixed effects model for viral dynamics, the non-parametric mixed model for CD4 process, and the parametric event-time model for the ratio CD4/CD8 decline. These three model are liked by the same random effects.
Wang and Taylor [57] used the AIDS data set to investigate the effect of treatment and risk factors on the markers and on the development of AIDS or death. The estimations showed that the CD4 levels before infection have a significant positive effect on CD4 after infection, and higher CD4 before infection leads to higher CD4 after infection. Analysing the adequacy of the model with the comparison between observed and predicted values, the model with IOU or Brownian process fits better the data. In add a simulation study showed that the biggest differences are noticeable for the hazard

model parameters comparing the joint model and an alternative approach in which the longitudinal and the hazard models are fitted separately.
Henderson, Diggle, and Dobson [22] proposed to use a different data set composed by schizophrenia trial data analysing the development of mean scores for each of three treatment groups on a particular measure of psychiatric disorder (Positive and Negative Symptom Scale, PANSS), the three treatment groups are Haloperidol, placebo, and Risperidone. A comparison between different model with different relations between the latent components is made showing that when there is a latent association, there is a substantial improvement in combined likelihood, and the improvement becomes greater when the relation between the latent components become more elaborate.
The same data set was used by Xu e Zeger [64] that analyse the effect of risperidone on the PANSS score and his effect on time of survival. The parameters showed a big effect for both and in add this model permits a more precise estimation of the survival function.
In 2008 Diggle, Sousa and Chetwynd [13] used fully parametric approach to analyse the schizophrenia trial data set. The goal of the trial was to compare the efficacy of the different treatments in reducing the mean PANSS score.
Subsequently Ibrahim, Chu and Chen [27] applied joint model to analyse the trade-off between quality of life and survival in cancer clinical trials. The joint modelling approach typically gives unbiased and larger estimates of the treatment effect when the longitudinal data is associated with survival.
In 2011 Sweeting and Thompson [51] used the multicentre aneurysm screening study data to analyse the association between abdominal aortic aneurysm diameter and the hazard of his rupture. Several models are implemented, the classical time-dependent, the classical two-stage, the classical shared random effects, and the Bayesian shared random effects models. The results showed a strong association between risk of rupture and current diameter.
In 2012 Das, Li, Huang, Gai, and Wu [8] applied a classical joint model to the data set composed by the quantitative trait loci in order to control development processes and the timing of development and their casual correlation over time.
In 2013 McCrink, Marshall and Cairns [34] used joint model to analyse the factors that affect the survival of end-stage renal disease patients. The factors analysed are the glomerular filtration rate and the changing haemoglobin levels. The first factor is analysed with the package JM of the software R because the aim of the study is to analyse the survival end-stage renal disease patients, while the second one is analysed with the package joineR of the same software because it focuses on the link between the two processes with shared latent random effects.
In 2015 Barrett, Diggle, Henderson and Taylor-Robinson [3] used the cystic fibrosis patients data set in order to analyse the relation between disease progression and survival. Four random-effects models are applied to the data: a stationary Gaussian process, a stationary Gaussian process with one time lag in the survival model, a random intercept and slope model,

and a stationary Gaussian process plus random intercept and slope.
Rizopoulos [43] analysed four data set. The first data set is primary biliary cirrhosis where the outcome of primary interest was patient survival and whether this could be prolonged by D-penicillamine. The interest was in serum bilirubin level and his association with survival because it is considered a strong indicator of disease progression .
The second data set is the AIDS data set, and the aim of this study was to compare the efficacy and safety of two alternative antiretroviral drugs, namely didanosine and zalcitabine, in the time-to-death, focusing on studying the association structure between the CD4 count and the risk for death for these advanced HIV-infected patients.
The third data set in the liver cirrhosis data set were patients randomized to a treatment with prednisone and the remaining receiving a placebo, but the interest was on the association between the prothrombin index and the risk for death, investigating also the capability of the prothrombin index in discriminating between subjects who died within a medically relevant time interval after their last assessment and subjects who lived longer than that.
The last data set is the aortic valve data set where the interest was in the association between the aortic jet velocity (aortic gradient) and the risk for death or re-operation.

Chapter 5

Extensions

After analysing the simplest joint models in which both the sub-model are univariate and of linear form, it is interesting to analyse different extensions which deal with hazard formulation, the heterogeneity in the sample, multiple failure times, accelerated failure time, categorical or multiple longitudinal outcomes, and joint cure model.

5.1 Different hazard function parametrizations

Rizopoulos [43] presented different parametrizations for the hazard function. The generalization of these parametrization is based on:

$$h_i(t) = h_0 \exp\{\gamma'\omega_{i1} + f(m_i(t-c), b_i, \omega_{i2}; \alpha)\} \tag{5.1}$$

where $f(\cdot)$ is a function of the true level of the marker $m_i(\cdot)$, of the random effects b_i, and of the extra covariates ω_{i2}.
The first parametrization is based on the interaction effects that is used when the effect of the true level of the marker is different between the subgroups of the target population, so Rizopoulos [43] proposed to include in the linear predictor of the relative risk model interaction terms of the marker with the baseline covariates of interest:

$$h_i(t) = h_0(t) \exp\{\gamma'\omega_{i1} + \alpha'[m_i(t)\omega_{i2}]\}$$

where ω_{i1} is used to accommodate the direct effects of baseline covariate to the risk for an event, and ω_{i2} contains interaction terms that expand the association of $m_i(t)$ in different subgroups in the data.
Another parametrization is based on lagged effects, that supposes that the current value of the time dependent covariates does not affect the current risk for an event, but it is influenced by the values of the covariates at a precedent time, so it is possible to use time-lagged covariates:

$$h_i(t) = h_0(t) \exp\{\gamma'\omega_{i1} + \alpha m_i[\max(t-c, 0)]\}$$

where c specifies the time lag of interest, so the relative risk indicates that the risk at time t depends on the true value of the longitudinal marker at time $t - c$.
The third parametrization is based on the time-dependent slope which is based on the work of Ye, Lin and Taylor [65] who supposed that in the joint model the risk depends on both the current true value of the trajectory and the slope of the true trajectory at time t:

$$h_i(t) = h_0(t) \exp[\gamma' \omega_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)]$$

where

$$m_i'(t) = \frac{d}{dt} m_i(t) = \frac{d}{dt} [x_i'(t)\beta + z_i'(t) b_i]$$

and where the parameter α_2 measures how strongly associated is the value of the slope of the true longitudinal trajectory at time t with the risk for an event at the same time point, provided that $m_i(t)$ remains constant. This parametrization is applied to the study of cancer recurrence in prostate cancer after radiation therapy, analysing the dependence of the risk of cancer recurrence in the prostate-specific antigen level and the rate change of the prostate-specific antigen level.
Subsequently Rizopoulos [43] suggested that the risk for an event at a specific time depends on an elaborate function of the longitudinal marker history, like the cumulative effect of the longitudinal outcome:

$$h_i(t) = h_0(t) \exp \left[\gamma' \omega_i + \alpha \int_0^t m_i(s) ds \right]$$

where for any particular time point t, α measures the strength of the association between the risk for an event at time point t and the area under the longitudinal trajectory up to the same time t, with the area under the longitudinal trajectory regarded as a suitable summary of the whole trajectory. In add sometimes is better to consider a weight function for evaluating the past values of the marker:

$$h_i(t) = h_0(t) \exp \left[\gamma' \omega_i + \alpha \int_0^t \varpi(t - s) m_i(s) ds \right]$$

where $\varpi(\cdot)$ denotes the weight function.
Rizopoulos [43] introduced in the hazard function some time dependent covariates considering the case in which there are some exogenous time-dependent covariates that influenced the risk of failure:

$$h_i(t) = h_0(t) \exp[\gamma' \omega_i(t) + \alpha m_i(t)]$$

Rizopoulos [43] applied these extensions to the primary biliary cirrhosis and to the liver cirrhosis data set.

5.2 Heterogeneity in the sample

If the sample considered comes from a non-homogeneous population, so the population is divided in multiple strata, then Rizopoulos [43] introduced a stratified relative risk model. Accordingly the risk for patient i belonging to stratum k becomes:

$$h_{ik}(t) = h_{0k}(t) \exp[\gamma' \omega_i + \alpha m_i(t)]$$

where $h_{0k}(t)$ denoting the baseline hazard function for stratum k. This model is applied to the primary biliary cirrhosis data set, stratifying the relative risk sub-model in two part.
If, instead, the interest is in recovering latent heterogeneity which is not captured by any of the observed covariates, a latent class joint models can be used. The population is divided in class and the model postulates that patients in different latent groups have both different longitudinal evolutions and different risks for an event, so the joint model becomes:

$$\begin{cases} h_i(t|c_i = g) = h_{0g}(t) \exp(\gamma_g' \omega_i) \\ [y_i(t)|c_i = g] = x_i'(t)\beta_g + z_i'(t) b_{ig} + \varepsilon_i(t) \\ \varepsilon_i(t) \sim N(0, \sigma^2) \\ P(c_i = g) = \frac{\exp(\lambda' u_i)}{\sum_{l=1}^{G} \exp(\lambda_l' u_i)} \end{cases} \tag{5.2}$$

where c_i is the class indicator. The application concerns the data set of AIDS, and he obtained, obtaining that the population can be split in three distinct sub-population.
This model was proposed by Proust-Lima, Joly, Dartigues, and Jacqmin-Gadda [38] in order also to deal with Multiple longitudinal outcomes.

5.3 Multiple failure times

There are two type of multiple failure the competing risks, different causes of failure, and the recurrent event, single event that may occur several times. In the first case there are K different causes of failure so the standard relative risk model becomes:

$$h_{ik}(t) = h_{0k}(t) \exp[\gamma_k' \omega_i + \alpha_k m_i(t)]$$

where $k = 1, ..., K$ and the parameters are different for each cause of failure. This model is applied to the primary biliary cirrhosis data set, in order to consider three different state, dead, transplanted and alive.
In the second case two different type of risk model can be used, one for the recurrent events and one for the terminal events:

$$\begin{cases} r_i(t) = & r_0(t) \exp[\gamma_r' \omega_{ri} + \alpha_r m_i(t) + v_i] \\ h_i(t) = & h_0(t) \exp[\gamma_h' \omega_{hi} + \alpha_h m_i(t) + \zeta v_i] \end{cases} \tag{5.3}$$

where ω_{ri} denotes the baseline covariates affecting the risk for a recurrent event, and ω_{hi} denotes the baseline covariates affecting the risk for the terminating event, with corresponding regression coefficients γ_r and γ_h. Parameters α_r and α_h measure the strength of the association between the current value of the longitudinal marker and the risk for a recurrent or a terminal event, respectively. Term v_i is a random effect that accounts for the correlation in the recurrent events. Parameter ζ in the terminating event relative risk model measures how strongly associated is the risk of a terminating event with the risk for a recurrent event.

5.4 Accelerated failure time

This model, presented by Tseng, Hsieh and Wang [52], specifies that predictors act multiplicatively on the failure time or additively on the log failure time, so the proportionality assumption fails to capture the relationship between the survival time and longitudinal covariates. AFT models are typically defined as:

$$\log T_i^* = \gamma' \omega_i + \sigma_t \varepsilon_{ti} \tag{5.4}$$

where parameter σ_t is a scale parameter and ε_{ti} is assumed to follow a specific distribution. Every single parameter γ_j denotes the change in the expected log failure time for a unit change in the corresponding covariate ω_{ij}. In add the risk rate function for subject i becomes:

$$h_i(t|M_i(t), \omega_i) = h_0(V_i(t)) \exp[\gamma' \omega_i + \alpha m_i(t)] \tag{5.5}$$

where:

$$V_i(t) = \int_0^t \exp[\gamma' \omega_i + \alpha m_i(s)] ds$$

A simulation study is made in order to analyse the performance of the EM procedure used to estimate the parameters. In addition an application of this model to the data set composed by some female Mediterranean fruit flies, and the number of eggs produced daily is made, jointly analysing the fecundity curves and the longevity. This model reflects covariate risks on an accelerated time scale and involves the cumulative reproductive effects and not just the daily effects.

5.5 Categorical or multiple longitudinal outcomes

To handle with categorical longitudinal outcomes Rizopoulos [43] proposed to use generalized linear mixed model (GLMM). This type of model uses

the classical formulation and the joint model becomes:

$$\begin{cases} p(y_i(t)|b_i) = \exp\left\{\sum_{j=1}^{n_i} \frac{y_{ij}\psi_{ij}(b_i) - c[\psi_{ij}(b_i)]}{a(\varphi)} - d(y_{ij}, \varphi)\right\} \\ m_i(t) = E(y_i(t)|b_i) = g^{-1}[x_i'(t)\beta + z_i'(t)b_i] \\ b_i \sim N(0, D) \\ h_i(t) = h_0(t)\exp\{\gamma'\omega_{i1} + f[m_i(t-c), b_i, \omega_{i2}; \alpha]\} \end{cases} \tag{5.6}$$

where the formulation for the survival model $f[m_i(t-c), b_i, \omega_{i2}; \alpha]$ can have the simplest formulation, or one of the form presented above.
While to solve the problem of multiple longitudinal outcome a multivariate generalized linear mixed-effects model can be used where the linear predictor is given by:

$$g_q[E(y_{iq}(t)|b_{iq}] = x_{iq}'(t)\beta + z_{iq}'(t)b_{iq} \tag{5.7}$$

where $g_q(\cdot)$ denotes a known one-to-one monotonic link function, and $y_{iq}(t)$ denotes the value of the qth longitudinal outcome for the subject i at time point t, and accordingly the relative risk becomes:

$$h_i(t) = h_0(t)\exp\left\{\gamma'\omega_{i1} + \sum_q f_q[m_{iq}(t-c), b_{iq}, \omega_{i2q}; \alpha_q]\right\} \tag{5.8}$$

where q indicates the number of longitudinal covariates and b_{iq} random effects vector member of the exponential family.
Another approach for handling multiple longitudinal outcomes is proposed by Proust-Lima, Joly, Dartigues, and Jacqmin-Gadda [38], these authors proposed a multivariate joint model in which the longitudinal outcomes are considered as realizations of a single latent process which is defined in continuous time and represents the common unobserved factor that drives the observed longitudinal trajectories. In addition different error terms for each longitudinal outcome are considered which account for the extra correlation in the repeated measurement of the longitudinal outcome not captured by the random effect.

5.6 Joint cure model

In the joint-cure model, presented by Law, Taylor, and Sandler [30] and by Yu, Law, Taylor, and Sandler [66], joint modelling of the disease progression marker and the failure time process is done in a cure model setting. They assumed that a fraction of the patients are cured by the treatment and are immune from recurrence, the event of interest.
There are three sub-models, the incidence model which indicates the probability of an individual i to be in the susceptible group, that is given by the logistic function:

$$p(D_i = 1|b, Z_i) = \frac{\exp(b'Z_i)}{1 + \exp(b'Z_i)} \tag{5.9}$$

while the longitudinal model is given by:

$$\log(Y_{ij+1}) = \log(Y_{ij}^* + 1) + \epsilon_{ij} \tag{5.10}$$

and the conditional failure time model is:

$$\lambda(t|D_i = 1, R_i, Z_i) = \lambda_0(t|\eta) \exp[\gamma log(Y_i^*(t) + 1) + \beta' Z_i^*] \tag{5.11}$$

where Z_i denote the $q+1$ fixed baseline covariates for subject i, Y_i the vector of longitudinal values, T_i indicates the observed follow-up time, and δ_i is the corresponding censoring indicator. The cure group indicator is denoted by D_i, thus if a subject i is in the susceptible group, D_i is equal to 1; otherwise, it is equal to 2. Y_{ij}^* is the true longitudinal value at t_{ij} , $\epsilon_{ij} \sim N(0, \sigma_e^2) iid$ is the measurement error at time t_{ij}, $\lambda_0(t|\eta)$ is the unspecified baseline hazard function at time t. The EM and the Markov chain Monte Carlo algorithm are used for the estimation.
There are three major advantages of adding a longitudinal component to the cure model: at first it is possible to reduce the bias due to informative censoring, secondly there is an increase in individual predictions, and lastly it is possible to observe a reduction of the identifiability problems in a cure model.
This model is applied to a data set composed by patients who had carcinoma of the prostate and who were treated with radiation therapy. The endpoint of interest is clinical recurrence, but a fraction of the patients are cured by the treatment and are immune of recurrence, considering the prostate-specific antigen as a biomarker.

Chapter 6

The case of multivariate longitudinal

This thesis focuses on the extension in which only the longitudinal sub-model is multivariate, as a first possible evolution of the classical joint model in which both sub-models are univariate.
This chapter is organised as follows. At first a review of the literature concerning joint models with multivariate longitudinal sub-model is presented. Subsequently the new method of estimation is proposed: the two-stage approach.
We propose this method as a first approach to solve the computational problem. As already said, the joint model is computationally demanding, then increasing the number of parameters or the dimensions of the sub-models will lead to method of estimation more computationally demanding. The two-stage approach permits to obtain very fast and with desirable proprieties estimations.

6.1 Model definitions and estimations

Xu and Zeger [63] presented one of the first article that concerns joint model with multivariate longitudinal sub-model. The paper proposed a latent variable model for the joint analysis of a time to event and repeated measures on multiple surrogate marker processes. The authors proposed two complementary approaches to answer the question whether using multiple surrogate processes is better than using only one. A Markov chain Monte Carlo (MCMC) algorithm is used to estimate parameters in the model extending the Xu and Zeger [64] model and the Faucett and Thomas [15] model. In addition some assumptions are made: the time to clinical event T and vector of repeatedly measured biomarkers Y are conditionally independent given η, the treatment X can affect T either through η or directly, and X only affects Y through its influence on η, where η is the latent process.
In case of multivariate longitudinal model all the elements that variate as a function of two components must be considered, the subject i and the biomarker or the longitudinal variable k. Then $\eta_{ik}(t)$ is the latent stochastic

process that represents the true value of the biomarker k of the patient i, $Y_{ik}(t)$ is an imprecise measure of $\eta_{ik}(t)$, $X_{ik}(t)$ is the collection of predictors including a treatment indicator for the same marker process and the same patient, where $k = 1, 2, ..., K$, $i = 1, 2, ..., n$, and T_i represents the time event for the subject i. The authors supposed that given $\eta_{ik}(t)$, $Y_{ik}(t)$ is an independent observation from a generalized linear model with linear predictor $\eta_{ik}(t)$ such that:

$$g_k\{E[Y_{ik}(t)|\eta_{ik}(t)]\} = \eta_{ik}(t) \tag{6.1}$$

where g_k is an arbitrary but known link function. Another assumption supposed that $\eta_{ik}(t)$ follows a multivariate extension of the standard Gaussian random effects model:

$$\eta_{ik}(t) = X_{ik}(t)\beta_k + D_{ik}(t)U_{ik}(t) \tag{6.2}$$

where β_k is an $m_k \times 1$ vector of regression coefficients and U_{ik} is an $r_k \times 1$ vector of random effects corresponding to the biomarker k. At first $U_i = (U_{i1}, U_{i2}, ..., U_{iK})$ is supposed to be a realization of a Gaussian random vector with mean zero and $r \times r$ covariance matrix G, where $r = r_1 + r_2 + \cdots + r_K$. Concerning the survival function, a conditional hazard model is proposed with the form:

$$\begin{aligned} h(t|\eta_i(t), X_i) =& h_0(t; \alpha_0) \\ & \exp\{\alpha_{11}\eta_{i1}(t) + \alpha_{12}\eta_{i2}(t) + \cdots + \alpha_{1K}\eta_{iK}(t) + \alpha_2 X_i\} \end{aligned} \tag{6.3}$$

An analyse of the relative benefits of multiple versus one biormaker evaluating the gain in precision comparing the length of the predictive intervals is done.
An application of this model to the Schizophrenia trial data comparing the risperidone and placebo for the treatment of schizophrenia is presented, comparing the model with only one biomarker and the one with three biomarker and showing that in this situation there is a gain in precision also if it is not so big, then the gain is not so big to warrant the additional risk of bias from the more complex model. The authors used the Markov chain Monte Carlo algorithm to estimate the parameters.
Another paper concerning the joint analysis of time-to-event and multiple longitudinal variables was proposed by Lin, McCulloch and Mayne [32]. The authors extended the model presented by Wulfsohn and Tsiatis [62]. The model allows a direct dependence of the event process on the multiple longitudinal covariates simultaneously and also accommodates the correlation among the longitudinal covariates. A one-step-late EM algorithm is used to handle the direct dependence of the event process on the modelled longitudinal variables along with the presence of other fixed covariates in both processes.
The model formulation is:

$$y_{ij}(t) = X'_{ij}(t)\beta_j + Z'_{ij}(t)b_{ij} + \varepsilon_{ij} \equiv y^b_{ij}(t) + \varepsilon_{ij} \tag{6.4}$$

where $i = 1, ..., n$ is the subject index, $j = 1, ..., J$ is the longitudinal covariate index, $y_{ij}(t)$ denotes the longitudinal response j for subject i measured at time t, $X_{ij}(t)$ are subject i specific covariate vectors for fixed effects at time t in the longitudinal sub-model j with coefficient vector β_j, and $Z_{ij}(t)$ are subject i specific covariate vectors for random effects at time t in the longitudinal sub-model j with coefficient vector b_{ij}. The vector of random effect b_{ij} was taken to be multi-normally distributed with mean B_j and covariance D_{jj}. The measurement error ε_{ij} is assumed uncorrelated with b_{ij}, normally distributed with mean 0 and variance σ_j^2.
In addition the hazard model was given as:

$$\lambda(t|b_{ij}, \omega_i) = \omega_i \lambda_0(t) \exp\left\{ x_i'(t)\gamma_x + \sum_{j=1}^{J} \left[y_{ij}^b(t)\gamma_{yj} + \sum_{l=1}^{p_x} x_{il} y_{ij}^b(t)\gamma_{yjl} \right] \right\} \tag{6.5}$$

where λ_0 is an unspecified conditional baseline hazard. The vector of fixed covariates in x_i is associated with the coefficient vector γ_x while the term $\sum_{l=1}^{p_x} x_{il} y_{ij}^b(t)\gamma_{yjl}$ denotes the interaction between $y_{ij}^b(t)$ and the elements l in covariate vector x_i. It is often supposed to be zero for some l. The frailty distribution of ω_i is assumed to be $Gamma(\frac{1}{\theta}; \theta)$ with mean one and variance θ. Lastly the counting process notation must be considered, where the event process for subject i is written as $(N_i(t); Y_i(t))$ with $N_i(t)$ counting the number of events for subject i by time t and $Y_i(t)$ being a left continuous at risk process with $Y_i(t) = 1$ if subject i is at risk at time t and 0 otherwise. In the survival context, the counting process $N_i(t)$ for each subject i remains at zero unless and until death, when it jumps to one.
For using the maximum likelihood the authors considered some vectors:

$$\begin{aligned} y_i &= (y_{i1}', \ldots, y_{iJ}')' \\ \beta &= (\beta_1', \ldots, \beta_J')' \\ b_i &= (b_{i1}', \ldots, b_{iJ}')' \\ \gamma &= (\gamma_x', \gamma_{y_1}, \ldots, \gamma_{y_J}, \gamma_{y,inter}')' \end{aligned}$$

and some matrices:

$$X_i = \begin{pmatrix} X_{i1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & X_{iJ} \end{pmatrix}$$

$$D = \begin{pmatrix} D_{11} & \ldots & D_{1J} \\ \vdots & \ddots & \vdots \\ D_{J1} & \ldots & D_{JJ} \end{pmatrix}$$

$$\Sigma_i = \begin{pmatrix} \sigma_1^2 I_{n_i 1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_J^2 I_{n_i J} \end{pmatrix}$$

where the row t of X_{ij} is a p-vector of covariates measured at time t. In add from this it is possible to obtain that: $y_i|b_i \sim N(X_i\beta + Z_i b_i, \Sigma_i)$ and

$b_i \sim N(B, D)$. The log-likelihood can be written as:

$$\sum_{i=1}^{n}[l(y_i|b_i) + l(N_i, Y_i|b_i, \omega_i) + l(b_i) + l(\omega_i)] \tag{6.6}$$

where

$$l(N_i, Y_i|b_i, \omega_i) = N_i(\tau)\log\omega_i + \int_0^\tau [\log\lambda_0(t) + W_i(t, \gamma, \beta, b)]dN_i(t) + \\ - \omega_i \int_0^\tau \tau Y_i(t)\exp(W_i(t, \gamma, \beta, b))d\Lambda_0(t)$$

where τ is the maximum potential follow-up time and:

$$W_i(t, \gamma, \beta, b) = x_i'(t)\gamma_x + \sum_{j=1}^{J}\left[y_{ij}^b(t)\gamma_{yj} + \sum_{l=1}^{p_x} x_{i1}y_{ij}^b(t)\gamma_{yjl}\right]$$

Subsequently the authors applied the Monte Carlo E-step to the log-likelihood and a one-step-late M-step with Newton-Raphson iteration from which it is possible to obtain the estimation of all the parameters involved: $\hat{\lambda}_0, \hat{\sigma}_j^2, \hat{B}_j, \hat{D}$ and $\hat{\beta}_j$. In addition the other two parameters, γ and θ, are updated through one-step Newton-Raphson iteration. The authors applied this new method to the data set of the beta-carotene trial showing the benefits of joint modelling the longitudinal and time-to-event variables.
Subsequently Song, Davidian and Tstiatis [50] generalized the model presented by Tsiatis and Davidian [53], thus a generalisation of semi-parametric conditional score estimation for the parameters is presented. For each subject i , $i = 1, ..., n$, let T_i denote the failure time and C_i denote censoring time. The observed survival data are $V_i = min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. Let $X_{ik}(u)$, $k = 1, ..., K$, denote K time-dependent covariates at time u, and let the r-dimensional vector Z_i denote r time-independent covariates for subject i. Assuming that each covariate process $X_{ik}(u)$ satisfies:

$$X_{ik}(u) = \alpha_{ik}'f_k(u) \tag{6.7}$$

where $f_k(u)$ is a $(q_k \times 1)$ vector of functions of u, α_{ik} is a $(q_k \times 1)$ random effect, and f_k and α_{ik} may be different for each k. The covariate processes $X_{ik}(u)$ are not observed directly; rather, longitudinal measurements $W_{ik}(t_{ikj})$ on the covariate k are taken at times t_{ikj}, $j = 1, 2, ..., m_{ik}$, for each i, where:

$$W_{ik}(t_{ikj}) = X_{ik}(t_{ikj}) + e_{ikj} \tag{6.8}$$

and e_{ikj} are normally distributed mean-zero errors with variance σ_{kk} that may reflect both biological variation and measurement errors.
In addition let $e_i = (e_{i1}', ..., e_{iK}')'$, where $e_{ik} = (e_{ik1}', ..., e_{ikm_{ik}}')'$; $t_{ik} = (t_{ik1}', ..., t_{ikm_{ik}}')'$ be the ordered times for subject i, covariate k, and $t_i = (t_{i1}', ..., t_{iK}')'$ be the set of time points where observations on all K covariates are available; $m_i = (m_{i1}', ..., m_{iK}')'$; and $\alpha_i = (\alpha_{i1}', ..., \alpha_{iK}')'$ is $(q \times 1)$, where $q = \sum_k q_k$. It is useful to assume that the conditional distribution

of e_i given $(T_i, C_i, \alpha_i, Z_i, t_i, m_i)$ is normal with covariance matrix depending only on m_i and on the parameters of covariance between errors of two different covariates.
A proportional hazards regression model is assumed for the relationship between the hazard of failure and the covariates, accordingly the hazard for subject i becomes:

$$\begin{aligned}\lambda_i(u) = &\lim_{du\to 0} du^{-1} p\{u \le T_i < u + du | T_i \ge u, \alpha_i, Z_i, C_i, e_i(u), t_i(u)\} = \\ =&\lambda_0 \exp\{\gamma' G(u, \alpha_i) + \eta' Z_i\}\end{aligned} \tag{6.9}$$

where $\lambda_0(u)$ is an unspecified baseline hazard function; $G(u, \alpha_i)$ is a $(s \times 1)$ vector whose elements are functions of u and α_i ; γ and η are $(s \times 1)$ and $(r \times 1)$ parameter vectors; $t_i(u) = \{t_{ikj} \le u; k = 1, ..., K\}$ denotes the observation times up to and including u; and $e_i(u) = \{e_{ikj} : t_{ikj} \le u; k = 1, ..., K\}$. The vector $G(u, \alpha_i)$ allows flexibility in modelling the hazard relationship. The authors used the conditional score estimation for the parameters, generalized the method proposed by Tsiatis and Davidian [53], assuming that $G(u, \alpha_i) = G(u)\alpha_i$.
The applications concern the data set composed by the aids clinical trials group, analysing the time trajectories of CD4 and CD8 and the effect of zidovudine alone or with three others therapies (zidovudine plus didanosine, zidovudine plus zalcitabine, or didanosine alone). The results suggested that this method of estimation is good, confirmed also from a simulation study.
Subsequently Ibrahim, Chen, and Sinha [26] presented a joint model for studying the effect of a vaccine on the melanoma through the analysis of two antibodies.
Let $X_i(t)$ denote the true, unobservable antibody level, and let $Y_i(t) = (Y_{i1}(t), Y_{i2}(t))'$ denote the (2×1) vector of observed antibodies immunological titer measures for subject i at risk at time t, $i = 1, ..., n$. The model can be written as:

$$\begin{cases} Y_{i1}(t) = X_i(t) + \varepsilon_{i1}(t) \\ Y_{i2}(t) = \alpha_0 + \alpha_1 X_i(t) + \varepsilon_{i2}(t) \\ \varepsilon(t) \sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right) \end{cases} \tag{6.10}$$

where $\varepsilon_i(t) = (\varepsilon_{i1}(t), \varepsilon_{i2}(t))'$ is independent of $X_i(t)$.
The survival component of the model is taken to have a proportional hazards structure. For the subject i at risk at time t, let $\chi_i(t)$ denote the history of $X_i(\cdot)$ up to time t, $\mathbf{Y_i(t)}$ denotes the history of the observable covariates $(Y_{i1}(\cdot), Y_{i2}(\cdot))$ up to t, and let z_i denote a $(p \times 1)$ vector of baseline covariates for subject i, such as treatment, gender, age, and so forth. The hazard function for the subject i becomes:

$$h(t|\chi_i(t), \mathbf{Yi(t)}, \mathbf{z_i}) = \mathbf{h_0(t)} \exp(\beta_\mathbf{1} \mathbf{X_i(t)} + \mathbf{z_i'} \beta_\mathbf{2}) \tag{6.11}$$

where $h_0(t)$ is the baseline hazard function, β_1 is a scalar regression coefficient for the longitudinal covariate process, and β_2 is a $(p \times 1)$ vector of

regression coefficients for the baseline covariates.
The method of estimation implemented is the maximisation of likelihood function using the Gibbs sampling. The applications showed that the vaccines are associated with the event.
Brown, Ibrahim, and DeGruttola [4] proposed a joint longitudinal and survival model that has a non-parametric model for the longitudinal markers using a cuBIC B-splines to specify the longitudinal model and a proportional hazard model to link the longitudinal measures to the hazard.
Analysing the longitudinal cuBIC B-spline model is possible to focus on the multivariate case assuming that p is the number of longitudinal outcomes:

$$Y_{ij} = \psi_{\alpha,\beta}(t_{ij}) + \varepsilon_{ij} \tag{6.12}$$

such as:

$$Y_{ij} = \begin{pmatrix} Y_{ij1} \\ \vdots \\ Y_{ijp} \end{pmatrix} \tag{6.13}$$

where Y_{ij} indicates the set of biomarkers of the the subject i observed at time t_{ij}.

$$\psi_{\alpha,\beta}(t_{ij}) = \begin{pmatrix} \psi_{\alpha,\beta,1}(t_{ij}) \\ \vdots \\ \psi_{\alpha,\beta,p}(t_{ij}) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{q} \beta_{ik1} B_k(t_{ij}) + x_i' \alpha_1 \\ \vdots \\ \sum_{k=1}^{q} \beta_{ikp} B_k(t_{ij}) + x_i' \alpha_p \end{pmatrix} \tag{6.14}$$

where $\beta_{ik} = (\beta_{ik1}, ..., \beta_{ikp})' \sim N(b_{0k}, V_{0k})$ and α_i is a vector of parameters linking the vector of baseline covariates x_i to the longitudinal outcome.
Finally:

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijp} \end{pmatrix} \tag{6.15}$$

where $\varepsilon_{ij} \sim N_p(0, \Sigma)$.
Subsequently the authors assumed that $\psi_\beta = \psi_{\alpha=0,\beta}(t_{ij})$ and accordingly defined the hazard function given the longitudinal measures as:

$$h(t|Y) = \lambda(t) \exp(\gamma \psi_\beta(t) + z'\zeta) \tag{6.16}$$

where $\gamma = (\gamma_1, ..., \gamma_p)'$ is a vector of parameters linking the trajectory to the hazard function, $\lambda(t)$ is the baseline hazard, and ζ is a parameter vector linking a vector z of baseline covariates to the failure time. After posing some priors and using some rules of approximations, the authors implemented the Gibbs sampling method for the estimation of the parameters involved in the model. A simulation study and an application on AIDS data set is also made in order to find the efficiency of the model and the cuBIC B-spline model provides a good fit to the longitudinal data that could not be obtained with simple parametric models.
Fieuws, Verbeke, Maes, and Vanrenterghem [16] proposed a multivariate

mixed model specifying a joint distribution for the random effects, thus the univariate mixed models are combined into a multivariate mixed model by specifying a joint distribution for all the random effects. In order to obtain the estimations a pairwise modelling strategy is used, where all possible pairs of bivariate mixed models are fitted using in addition the pattern-mixture approach, the pseudo-likelihood theory, and the Monte Carlo integration. The authors applied this model to analyse the renal graft failure with the study of some biomarkers.

Ghisletta [18] applied a joint multivariate longitudinal survival analysis to the cognitive data of the Swiss Interdisciplinary Longitudinal Study on the Oldest Old. The author simultaneously estimated a multivariate multilevel longitudinal model and a Weibull survival model to test whether individual performance and change in speed and fluency predict survival, controlling for retest effects and sensory functioning.

Albert and Shih [1] proposed a regression calibration approach for jointly modelling multiple longitudinal measurements and discrete time-to-event data, using a regression calibration approach which appropriately accounts for informative drop-out.

The model is composed by T_i which indicates the discrete event-time which can take on discrete values t_j, $j = 1, 2, ..., J$, and Y_{ij} to be a binary indicator of whether the patient i is dead at time t_j. Then $J_i = \sum_{j=1}^{J}(1-Y_{ij}) = J - Y_{i\cdot}$ where $Y_{i\cdot} = \sum_{j=1}^{J} Y_{ij}$ indicates the number of follow-up measurements before the event or the end of follow-up at time t_J, longitudinal measurements are measured at times $t_1, t_2, ..., t_{J_i}$. Denote $X_{1i} = (X_{1i1}, X_{1i2}, ..., X_{1iJ_i})'$, $X_{2i} = (X_{2i1}, X_{2i2}, ..., X_{2iJ_i})'$, ..., $X_{Pi} = (X_{Pi1}, X_{Pi2}, ..., X_{PiJ_i})'$ as the P biomarkers measured repeatedly at $j = 1, 2, ..., J_i$ time points. Further, define $X^*_{pi} = (X^*_{pi1}, X^*_{pi2}, ..., X^*_{piJ_i})'$ as the longitudinal measurements without measurement error for the biomarker p and $X^*_i = (X^*_{1i}, X^*_{2i}, ..., X^*_{Pi})'$. The authors considered a joint model for multivariate longitudinal and discrete time-to-event data in which the discrete event time distribution is modelled as a linear function of previous true values of the biomarkers without measurement error on the probit scale. Specifically:

$$P(Y_{ij} = 1 | Y_{i(j-1)} = 0; X^*_i) = \Phi(\alpha_{0j} + \sum_{p=1}^{P} \alpha_p X^*_{pi(j-1)}) \tag{6.17}$$

where $i = 1, 2, ..., I$, $j = 2, 3, ..., J_i$, Y_{i1} is taken as 0, α_{0j} governs the baseline discrete event time distribution, and α_p measures the effect of the biomarker p ($p = 1, 2, ..., P$) at time t_{j-1} on survival at time t_j. Particularly this formulation allows for examining the effect of multiple true biomarker values at time t_{j-1} on the probability of an event between the time point t_j and t_{j-1}. The longitudinal data is modelled assuming that the fixed and random effect trajectories are linear. Specifically, the multivariate longitudinal biomarkers can be modelled as:

$$X_{pij} = X^*_{pij} + \varepsilon_{pij} \tag{6.18}$$

where

$$X^*_{pij} = \beta_{p0} + \beta_{p1} t_j + \gamma_{pi0} + \gamma_{pi1} t_j \tag{6.19}$$

where β_{p0} and β_{p1} are the fixed effect intercept and slope for the biomarker p, and γ_{pi0} and γ_{pi1} are the random effect intercept and slope for the biomarker p on the individual i. Denote $\beta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, ..., \beta_{P0}, \beta_{P1})'$ and $\gamma_i = (\gamma_{1i0}, \gamma_{1i1}, \gamma_{2i0}, \gamma_{2i1}, ..., \gamma_{Pi0}, \gamma_{Pi1})$, assuming that γ_i is normally distributed with mean 0 and variance Σ_γ, where Σ_γ is $(2P \times 2P)$ variance matrix, and ϵ_{pij} are independent error terms which are assumed to be normally distributed with mean 0 and variance $\sigma^2_{p\varepsilon}$. Alternately the event-time process could be adapted to depend on the random effects of the multivariate longitudinal process (e.g., γ_{pij} can replace $X^*_{pi(j-1)}$).
Conceptually, model can be estimated by maximizing the likelihood:

$$L = \prod_{i=1}^{I} \int_{\gamma_i} ... \int \left\{ \prod_{p=1}^{P} h(X_{pi}|\gamma_{pi0}, \gamma_{pi1}) \right\} \times \left\{ \prod_{j=2}^{J_i} (1 - r_{ij}) \right\} [r^{J_i<J}_{i(J_i+1)} f(\gamma_i) d\gamma_i \tag{6.20}$$

where $r_{ij} = P(Y_{ij} = 1|Y_{i(j-1)} = 0)$, $h(X_{pi}|\gamma_{pi0}, \gamma_{pi1})$ is the product of J_i univariate normal density functions each with mean X^*_{pij} and variance $\sigma^2_{p\varepsilon}$, and $f(\gamma)$ is a multivariate normal density with mean zero and variance Σ_γ. When $P = 1$, the likelihood can be maximized by numerical integration techniques or alternately through Monte Carlo methods, but these methods do not perform well for even moderately high dimensional random effects. Thus the authors proposed a two-stage regression calibration approach for estimation, which is based on two-stage. In the first stage, multivariate linear mixed models can be used to model the longitudinal data. In the second stage, the time-to-event model is estimated by replacing the random effects with corresponding empirical Bayes estimates, where the discrete event time distribution is modelled as a linear function of previous true values of the biomarkers without measurement error on the probit scale. The benefit of the models are shown with a simulation study and with an application made in order to examine the effect of multiple longitudinal biomarkers on the short-term prognosis for patients with primary biliary cirrhosis.
Rizopoulos and Ghosh [44] proposed a new semiparametric multivariate joint model that relates multiple longitudinal outcomes to a time-to-event. In particular, for the subject-specific longitudinal evolutions a spline-based approach is used, the baseline risk function is assumed piecewise constant, and the distribution of the latent terms is modelled using a Dirichlet Process prior formulation.
Let $Y_i = (y'_{i1}, ..., y'_{ik}, ..., y'_{iK})$, $k = 1, ..., K$ denote the K-variate response vector for the subject i $(i = 1, ..., n)$, where y_{ik} is an $n_{ik} \times 1$ vector of longitudinal responses for outcome k taken at some time points $t_{ij,k}$. This formulation allows that the longitudinal responses may be collected at different time points for each outcome. For the time-to-event outcome, let T_i denote the observed event time, taken as the minimum of the true event time T^*_i and the censoring time C_i . Furthermore, the event indicator is defined as $\delta_i = I(T^*_i \leq C_i)$, where $I(\cdot)$ is the indicator function. The condi-

tional distribution of y_{ik} given a vector of random effects b_{ik} is assumed to be a member of the exponential family, with linear predictor given by:

$$g_k[E(y_{ik}(t)|b_{ik}] = f_{ik}(t) \tag{6.21}$$

where $g_k(\cdot)$ denotes a known one-to-one monotonic link function, and $y_{ik}(t)$ denotes the value of the longitudinal outcome k for the subject i at time point t. The unknown function $f_{ik}(\cdot)$ is assumed to describe the true, possibly non-linear, longitudinal profile for the outcome k. To allow for flexible shapes for the subject-specific evolutions for each outcome, a spline-based approach is proposed to approximate the function. Specifically, let $\lambda_k = \lambda_{lk}; l = 1, ..., L_k$ denote an increasing sequence of knot positions, then $f_{ik}(\cdot)$ is assumed to have the form:

$$f_{ik} \approx B_{ik}(\beta_k^{(1)}, \beta_k^{(2)}, b_{ik}^{(1)}) + H_{ik}(t; \beta_k^{(3)}, \beta_k^{(4)}, b_{ik}^{(2)}, \lambda_k) \tag{6.22}$$

where the approximation to $f_{ik}(t)$ consists of two parts, the time-independent and the time-dependent parts. The time-independent part $B_{ik}(\cdot)$ includes a set of baseline covariates located in the vectors $x_{ik}^{(1)}$ and $x_{ik}^{(2)}$, with corresponding vectors of fixed effects $\beta_k^{(1)}$ and $\beta_k^{(2)}$ and random effects $b_{ik}^{(1)}$. For the time-dependent part $H_{ik}(\cdot)$ the authors used a natural cuBIC spline basis functions with knots at λ_{lk}. The covariate vectors $x_{ik}^{(3)}$ and $x_{ik}^{(4)}$, with corresponding fixed effects $\beta_{lk}^{(3)}$ and $\beta_{lk}^{(4)}$, and random effects $b_{ilk}^{(2)}$, are used to include possible interactions of baseline covariates with the time-dependent part.
The effects of the longitudinal outcomes and of baseline covariates on the survival times are captured via a relative risk model of the form:

$$\begin{aligned} h_i(t|F_i^H(t), \omega_i) &= \lim_{dt \to 0} \frac{P[t \leq T_i^* < t + dt | T_i^* \geq t, F_i^H(t), \omega_i]}{dt} \\ &= h_0(t) \exp\left\{ \omega_i' \gamma + \sum_{k=1}^{K} m_{ik}[f_{ik}(t), r(\beta_k, b_{ik}), \phi_i, \alpha_k] \right\} \end{aligned} \tag{6.23}$$

where $F_i^H(t) = \{f_{ik}(s), 0 \leq s < t, 1 \leq k \leq K\}$ denotes the history of the true and unobserved longitudinal process up to time t, ω_i denotes a vector of baseline covariates with corresponding regression coefficients γ, and function $m_{ik}(\cdot)$ specifies which components of the longitudinal process for outcome k are related to the survival times, where ϕ_i denotes a frailty term and α_k denotes a parameter vector measuring the effect of the longitudinal outcome k to the time-to-event. To complete the specification of the survival model, the baseline risk function is assumed of the form:

$$h_0(t) = \sum_{q=1}^{Q} \xi_q I(\nu_{q-1} < t \leq \nu_q)$$

where $0 = \nu_0 < \nu_1 < \cdots < \nu_Q$ denotes a split of the time scale, with ν_Q being larger than the largest observed time, and ξ_q denotes the value of the

hazard in the interval $(\nu_{q-1}, \nu_q]$. The latent terms of the multivariate joint model consist of the random effects in the longitudinal process and possibly a frailty term in the survival process.
The authors proposed three parameterisations of the function m_{ik} and compared them in a simulation study and in an application to the analysis of renal graft failure using a Bayesian formulation for the semi-parametric multivariate joint model, and deriving the posterior inferences using a Markov chain Monte Carlo algorithm.
Choi, Anderson, Richards, and Thompson [5] implemented a joint model for mixed multivariate longitudinal measurements, applying it to the prediction of time until lung transplant or death in idiopathic pulmonary fibrosis. Specifically, the authors formulated a unified Bayesian joint model for the mixed longitudinal responses and time-to-event outcomes. For the longitudinal model of continuous and binary responses, multivariate generalized linear mixed models using shared random effects is investigated.
For subject i, $i = 1, ..., n$, let y_{1ij} and y_{2ij} denote the outcome j at time point t_{ij} consisting of continuous and binary components, respectively. Furthermore, let $y_i = (y'_{1i}, y'_{2i})'$ denote the bivariate longitudinal outcome vector for subject i, where $y_{hi} = (y_{hi1}, ..., y_{hij})'$, $h = 1, 2$, $j = 1, ..., n_{hi}$, is an n_{hi}-dimensional column vector giving the longitudinal outcome h for subject i. For the longitudinal bivariate response vector, y_i, with different data types, generalized linear mixed effects model is assumed that:

$$E(y_{hi}|b_{hi}) = g_h(X_{hi}\beta_h + Z_{hi}b_{hi}) \tag{6.24}$$

where $g_h(\cdot)$ denotes a known bijective link function that differs across data types, X_{hi} and β_h denote an $(n_{hi} \times p_h)$ design matrix of covariate values and a p_h-dimensional vector of fixed effects, respectively, and Z_{hi} and b_{hi} denote the $(n_{hi} \times q_1)$ design matrix of covariates and a q_h-dimensional vector of normally distributed random effects with a zero mean and covariance matrix Σ, respectively.
The generalized linear mixed effects model can be written with an identity link for the continuous response and the logit link for the binary response:

$$\begin{aligned} E(y_{1i}|b_{1i}) &= X_{1i} + \beta_1 + Z_{1i}b_{1i} \\ logit(P(y_{2i} = 1)|b_{2i}) &= X_{2i}\beta_2 + Z_{2i}b_{2i} \end{aligned} \tag{6.25}$$

assuming that b_{1i} follows a normal distribution with a mean vector of zeros and variance-covariance matrix Σ and that b_{2i} is proportional to b_{1i}, i.e. $b_{2i} = A_0 b_{1i}$, where A_0 is a diagonal matrix of unknown constants. Considering the survival sub-model, let T_i denote the true event time for subject i, C_i be the censoring time, and $\delta_i = I(T_i \leq C_i)$ be the event indicator. Let $T_i^* = min(T_i, C_i)$ be the observed event time for subject i. The proportional hazard model is given by:

$$\lambda_i(t|x_{3i}, U_{3i}) = \lambda_0(t) \exp[x'_{3i}\beta_3 + U_{3i}] \tag{6.26}$$

where x_{3i} is a p_3-dimensional vector of covariates with regression coefficients β_3, and $\lambda_0(t)$ is the baseline hazard function, which can be assumed to be of

a parametric form or left unspecified. To express the effects of longitudinal outcomes on the time-to-event outcome, the shared parameters, U_{3i}, are associated with the random effects of longitudinal outcomes, b_{1i} and b_{2i}. The joint model connects the longitudinal response sub-models and the event time outcome sub-model:

$$U_{3i} = a_1' b_{1i} + a_2' b_{2i} + b_{3i} \tag{6.27}$$

where $a = (a_1', a_2')'$ is a set of unknown constants and b_{3i} is a normally distributed frailty term with mean zero and variance σ_3^2, independent of the $b_i = (b_{1i}', b_{2i}')'$.
Then the authors built the log-likelihood for the observed data and implemented a Bayesian approach for parameter inferences, using a Gibbs sampling algorithm. This method of estimation was applied to a simulation study and to mortality in idiopathic pulmonary fibrosis outcomes study focusing on the survival function. Simulation studies indicate good performance for the models and the two longitudinal responses jointly contribute nearly significantly to the prediction of failure times.
He and Lou [20] developed a joint model that consists of a multilevel item response theory model for the multiple longitudinal outcomes, and a Cox proportional hazard model with piecewise constant baseline hazards for the event time data. Shared random effects are used to link together the two models. The model inference is conducted using a Bayesian framework via Markov Chain Monte Carlo simulation implemented in BUGS language. This model is applied to analyse the Parkison's disease.
Proust-Lima, Joly, Dartigues, and Jacqmin-Gadda [39] proposed another way to solve the problem related to the multivariate longitudinal data. These authors proposed a multivariate joint model in which the longitudinal outcomes are considered as realizations of a single latent process which is defined in continuous time and represents the common unobserved factor that drives the observed longitudinal trajectories. In addition different error terms for each longitudinal outcome are considered which account for the extra correlation in the repeated measurement of the longitudinal outcome not captured by the random effect.

6.2 The two-stage approach

After analysing the different models and methods of estimation already proposed, we choice the two-stage approach as an extension of the univariate method of estimation ([36, 37, 51, 53, 55, 65]). This option was chosen as a first possible method to solve the computational problem. As already said, the joint model is computationally demanding, then increasing the number of parameters or the dimensions of the sub-models will lead to method of estimation more computationally demanding. The two-stage approach permits to obtain very fast and with desirable proprieties estimations.
In this thesis the survival sub-model is a proportional hazard model which

is defined as a function of the $m_{iq}(t_{ij})$ that denotes the true and unobserved value of the longitudinal covariate q for subject i:

$$h_i(t|M_i(t),\omega_i) = h_0(t)\exp\left[\gamma'\omega_i + \sum_q \alpha_q m_{iq}(t)\right] \tag{6.28}$$

In (6.28):

- $M_i(t) = \{m_{iq}(s), 0 \leq s < t, \forall q = 1,\ldots,Q\}$ indicates the history of the true unobserved longitudinal processes up to time t
- α_q quantifies the effect of the longitudinal outcome q onto the risk of an event
- $h_0(t)$ indicates the baseline hazard function
- ω_i are the covariates that influence the risk of the event with coefficient γ.

The biggest difference from the survival sub-model with univariate longitudinal sub-model concerns the parameter α_q, thus a parameter for each longitudinal covariate considered is introduced.
This model can be simplified using a vectorial form for the parameter α_q for each $q = 1,...,Q$, thus the equation (6.28) becomes:

$$h_i(t|M_i(t),\omega_i) = h_0(t)\exp\left[\gamma'\omega_i + \boldsymbol{\alpha}'\mathbf{m}_i(t)\right] \tag{6.29}$$

where the vector $\boldsymbol{\alpha}$ and $\mathbf{m}_i(t)$ are composed by stacking the single elements of α_q and $m_{iq}(t)$ respectively.
Concerning the longitudinal sub-model a linear multivariate mixed model is proposed:

$$\begin{cases} y_{iq}(t_{ij}) = m_{iq}(t_{ij}) + \epsilon_{iq}(t_{ij}) \\ m_{iq}(t_{ij}) = x'_{iq}(t_{ij})\beta_q + z'_{iq}(t_{ij})b_{iq} \\ \epsilon_{iq}(t_{ij}) \sim N(0,\sigma^2) \\ b_{iq} \sim N(0,\Sigma) \\ b_{1q},...,b_{Nq},\epsilon_{1q},...,\epsilon_{Nq} \quad \textit{independent} \end{cases} \tag{6.30}$$

where q is the longitudinal variable index, $y_{iq}(t_{ij})$ is composed by the $m_{iq}(t_{ij})$ and by a random error term $\epsilon_{iq}(t_{ij})$, and β_q are the fixed effects for $x_{iq}(t_{ij})$, while b_{iq} are the random effects for $z_{iq}(t_{ij})$.
As in the survival sub-model, also in the longitudinal sub-model it is possible to simplified the model introducing the matrix formulation:

$$\mathbf{m}_{iq} = X_{iq}\beta_q + Z_{iq}b_{iq} + \epsilon_{iq} \tag{6.31}$$

where X_{iq} and Z_{iq} are the design matrices (with corresponding row vectors $x'_{iq}(t_{ij})$ and $z'_{iq}(t_{ij})$), while the vectors $\mathbf{m}_{iq}$ is composed by stacking the single elements $m_{iq}(t_{ij})$.
The two-stage approach is based on two steps. In the first step the fixed

effects and random effects for all the longitudinal sub-models are estimated using a maximum likelihood approach. Subsequently in the second step these estimates are used to impute appropriate values of $m_{iq}(t_{ij})$ that are substituted in the classical partial likelihood of the Cox model.
Thus in the first step it is possible to obtain through the maximum likelihood method the estimation of the parameters for the longitudinal sub-model, where the likelihood function can be expressed as:

$$L_{ML}(\theta_q) = \prod_{i=1}^{n} \left\{ (2\pi)^{-n_i/2} |V_{iq}|^{1/2} \exp\left[-\frac{1}{2} (\mathbf{y}_{iq} - X_{iq}\beta_q)' V_{iq}^{-1} (\mathbf{y}_{iq} - X_{iq}\beta_q) \right] \right\} \tag{6.32}$$

where $V_{iq} = Z_{iq}\Sigma Z'_{iq} + \sigma^2 I_J$ and θ_q indicates the parameters that must be estimated for each longitudinal covariate q. The likelihood (6.32) is related to the fact that:

$$Y_{iq} \sim N(X_{iq}\beta_q; Z_{iq}\Sigma Z'_{iq} + \sigma^2 I_J)$$

because the density function of the covariates are given by:

$$f(Y_{iq}) = \int f(Y_{iq}|b_{iq}) f(b_{iq}) db_{iq}$$

Then considering the likelihood function for each subject i it is possible to obtain the equation (6.32).
The estimations for the fixed effects obtained by the maximisation of the function above (6.32) are:

$$\hat{\beta}_q = \left[\sum_{i=1}^{n} X'_{iq} V_{iq}^{-1} X_{iq} \right]^{-1} \sum_{i=1}^{n} X'_{iq} V_{iq}^{-1} \mathbf{y}_{iq} \tag{6.33}$$

And the estimations for the random effects are:

$$\hat{b}_{iq}(\theta) = \Sigma Z'_{iq} W_{iq} (\mathbf{y}_{iq} - X_{iq}\hat{\beta}_q) \tag{6.34}$$

These estimations are expressed as a function of the design matrices and of the covariance matrices.
Subsequently it is possible to estimate the proportional hazard parameters through the partial log-likelihood:

$$pl(\gamma, \alpha) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ R_i(t) r(t,i) - \log \left[\sum_j R_j(t) \exp[r(t,j)] \right] \right\} dN_i(t) \tag{6.35}$$

where $R_i(t)$ assumes value 1 if the subject i is at risk 0 otherwise, $N_i(t)$ indicates the number of observed events for subject i by time t, and $r(t,i) = \gamma'\omega_i + \boldsymbol{\alpha}'\hat{\mathbf{m}}_i(t)$, where $\hat{\mathbf{m}}_i(t)$ is composed by staking the single elements $\hat{m}_{iq} = x'_{iq}(t)\hat{\beta}_q + z'_{iq}(t)\hat{b}_{iq}$ that are the values imputed in the partial log-likelihood obtained by the estimations of the linear mixed model.
This method is based on the partial likelihood function, expressed as:

$$L(\gamma, \boldsymbol{\alpha}) = \prod_{i=1}^{n} \prod_{t} \left\{ \frac{R_i(t) \exp[r(t,i)]}{\sum_j R_j(t) \exp[r(t,j)]} \right\}^{dN_i(t)} \tag{6.36}$$

applying the logarithm function it is possible to obtain the former equation (6.35).
The partial likelihood function for each subject i is evaluated as the probability to fail for a subject i at a fixed time t given the set of subject at risk at the fixed time t, which becomes in formula:

$$\begin{aligned} L_i(\gamma, \boldsymbol{\alpha}, t) =& p(individual\, i\, fails | set\, of\, individual\, at\, risk\, at\, time\, t) \\ =& \frac{p(individual\, i\, fails | at\, risk\, at\, time\, t)}{\sum_{l\, subject\, at\, risk} p(individual\, l\, fails | at\, risk\, at\, time\, t)} \\ =& \frac{h_i(t|M_i(t), \omega_i)}{\sum_l h_l(t|M_l(t), \omega_l)} \\ =& \frac{h_0(t) \exp\left[\gamma' \omega_i + \boldsymbol{\alpha}' \boldsymbol{m}_i(t)\right]}{\sum_l h_0(t) \exp\left[\gamma' \omega_l + \boldsymbol{\alpha}' \boldsymbol{m}_l(t)\right]} \end{aligned}$$

where l are the subjects at risk. In order to consider all the subject an indicator $R_i(t)$ can be introduces which assumes value 1 if the subject i is at risk 0 otherwise. Then the previous function becomes:

$$L_i(\gamma, \boldsymbol{\alpha}, t) = \frac{h_0(t) R_i(t) \exp\left[\gamma' \omega_i + \boldsymbol{\alpha}' \boldsymbol{m}_i(t)\right]}{\sum_j h_0(t) R_j(t) \exp\left[\gamma' \omega_j + \boldsymbol{\alpha}' \boldsymbol{m}_j(t)\right]}$$

where the index of summation j refers to all the subjects, not only the ones at risk.
In a situation in which there are time dependent covariates a new element must be considered: $N_i(t)$, that is the number of observed events for subject i by time t. Then the partial likelihood becomes:

$$L_i(\gamma, \boldsymbol{\alpha}, t) = \left\{ \frac{R_i(t) \exp\left[\gamma' \omega_i + \boldsymbol{\alpha}' \boldsymbol{m}_i(t)\right]}{\sum_j R_j(t) \exp\left[\gamma' \omega_j + \boldsymbol{\alpha}' \boldsymbol{m}_j(t)\right]} \right\}^{dN_i(t)}$$

Subsequently, for obtaining the partial likelihood function, the likelihood function for each subject i and for each instant time t must be considered, from which it is possible to obtain the equation (6.36).
The estimations for the two parameter vector $\boldsymbol{\alpha}$ and γ can be obtained by differentiating the partial log-likelihood function with respect to the parameter and posing it equal to zero.

$$\frac{\partial pl(\gamma, \boldsymbol{\alpha})}{\partial \gamma'} = \sum_{i=1}^{n} \int_0^{\infty} \left\{ R_i(t) \omega_i - \frac{\sum_j R_j(t) \omega_j \exp[r(t,j)]}{\sum_j R_j(t) \exp[r(t,j)]} \right\} dN_i(t) = 0$$

$$\frac{\partial pl(\gamma, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}'} = \sum_{i=1}^{n} \int_0^{\infty} \left\{ R_i(t) \boldsymbol{m}_i(t) - \frac{\sum_j R_j(t) \boldsymbol{m}_j(t) \exp[r(t,j)]}{\sum_j R_j(t) \exp[r(t,j)]} \right\} dN_i(t) = 0$$

Trough the numerical method of Newton-Raphson it is possible to solve these equations and obtain the estimations: $\hat{\gamma}$ and $\hat{\boldsymbol{\alpha}}$.
The exponential of each element of the vector estimated $\hat{\alpha}$ expressed as $\exp(\hat{\alpha}_q)$ denotes the relative increase in the risk for an event at time t that

results from one unit increase in $\hat{m}_{iq}(t)$ at the same time point, and $\exp(\gamma_j)$ denotes the ratio of hazards for one unit change in ω_{ij} at time t.
Ye et al. [65] made inference for the risk coefficient estimators, using the standard errors calculated based on the induced partial likelihood as if all the true covariate values were known. This method does not take into account the uncertainty of the estimated time-varying covariates. Therefore, the estimated standard errors for the risk coefficients are likely to be biased and tend to be smaller than the true variance of these risk coefficient estimates. The authors proved, through a simulation studies, that this assumption permits to obtain standard errors that are very close to the empirical ones.
Then supposing that all the true covariate values were known, the estimated parameter $\hat{\gamma}$ is:

- consistent
- asymptotically normally distributed with mean γ, the true parameter vector, and variance $E(I(\gamma))^{-1}$, the inverse of the expected information matrix $I(\gamma) = \frac{\partial^2 pl(\gamma,\alpha)}{\partial\gamma'\partial\gamma}$

Using the same assumptions the estimated parameter $\hat{\boldsymbol{\alpha}}$ is:

- consistent
- asymptotically normally distributed with mean $\boldsymbol{\alpha}$, the true parameter vector, and variance $E(I(\boldsymbol{\alpha}))^{-1}$, the inverse of the expected information matrix $I(\boldsymbol{\alpha}) = \frac{\partial^2 pl(\gamma,\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}'\partial\boldsymbol{\alpha}}$

Chapter 7

Applications

As already said the joint model is very useful in medical area because in clinical trail it is very interesting to analyse the two subgroups, for example placebo and treated, in order to study the longitudinal covariates that could influence the survival or the effect of a new drug. In this thesis we decide to focus on a different application, in fact for the first time the joint models are used to study the length of stay before graduation of university students. Thus the new method of estimation is applied to investigate the effect of some longitudinal covariates on an event that had never been investigated in joint models, the graduation.

7.1 The data set

Before presenting the results and the characteristics of the data set, we have to recall that the University system in Italy is different than in other European countries, where registration for the subsequent academic year is possible only if all credits in the syllabus of the previous year have been acquired. In Italy there are no limits for previously accumulate credits, each university student can proceed with his own trajectory of passed exams, free to depart from formal progression, with, at most, constraints on the sequence of examinations in the same subject, and can stay in the system for as long as he likes. Then this could lead to situations in which the student stay too much time in the system. In addition, the undergraduates in the Italian University system can obtain a degree only after three year of study and after having passed all the exams in the syllabus.
The data set that we analysed consists of the cohort of 1215 undergraduate students enrolled in the faculty of Economics for the academic year 2005-2006 at the University of Milano-BICocca, in the Lombardy region. Every student was followed for 5 academic years from the enrolment (the time is expressed in days), and every exam passed by the student was recorded. Table 7.1 reports the characteristics of the students. Females made up 55.31% of the students. Students aged 19 or under reached a total of 69.38%, meaning that most of them proceeded regularly through high school and enrolled immediately after it. In addition the number of students that obtained at

Table 7.1: Characteristic of data set

Character	Value	Frequency	Relative Frequency
Gender	Male	543	0.4469
	Female	672	0.5531
Age	≤ 19	843	0.6938
	>19	372	0.3062
High School Mark	≥ 90	324	0.2667
	<90	891	0.7333
Area	Milan	198	0.1630
	province Milan	373	0.3070
	Lombardy	530	0.4362
	Other	114	0.0938
Status	Drop-out	389	0.3202
	Degree	578	0.4757
	Still enrolled	248	0.2041

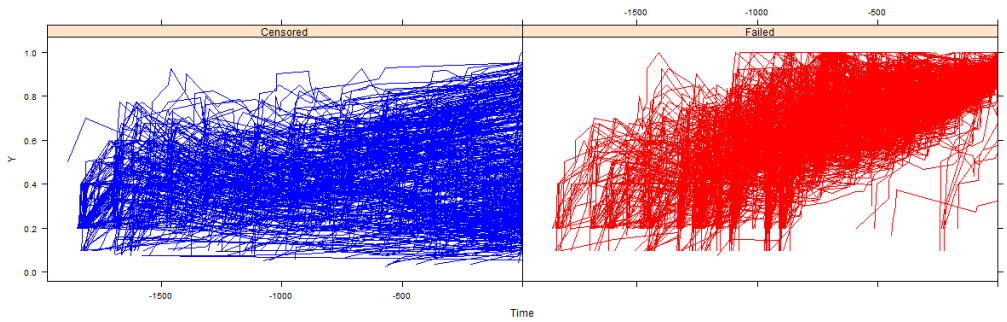

Figure 7.1: Trajectories of *cfu percentage*

least 90 out of 100 at high school final exam is low, only 26.67%. Concerning the living area, most students live in the province of Milan or in Lombardy. Lastly, 32.02% of students left the system before graduation (drop-out), while 47.57% obtained a degree and 20.41% were still enrolled at the end of the observation period of 5 academic yeas.
As every student was followed for 5 academic years, it was possible to observe their paths. After each exam the student can have a different level of *credit formative unit (cfu) or European University Credit (EUC) percentage*, and a different *average grade* and move to a different situation from the previous one, where the *cfu percentage* is a function of the maximum number of possible credits. Figure 7.1 shows the trajectories of *cfu percentage* for every student, where the censored units in the graphs are the students that leave the university (drop-out) or the students that are still enrolled at the university at the end of the fifth academic year, while failed units are the students that obtain the degree. As shown in Figure 7.1 the *cfu percentage* for the censored students is very low at the beginning and is lower compared to the failed students level, where the level at the beginning is low but increases more quickly. Analysing Figure 7.2 for the trajectories of

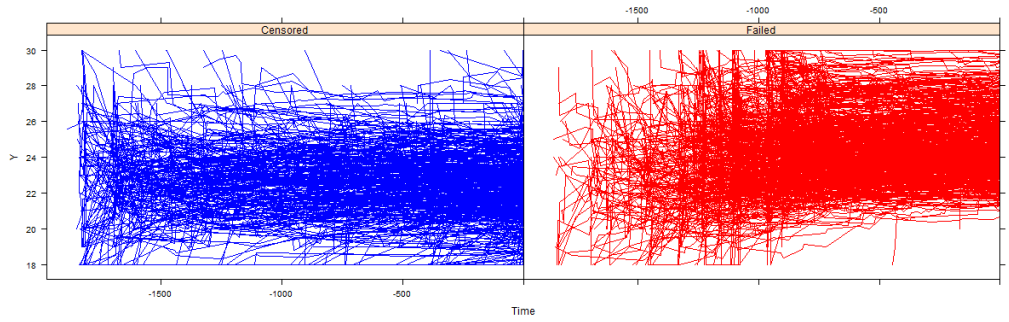

Figure 7.2: Trajectories of *average grade*

the *average grade*, we notice that the *average grade* for the failed students is slightly higher than that of the censored students.

7.2 Study of length of stay at the university with one longitudinal covariate

Several models were estimated considering the available covariates. The models were then chosen on the basis of significant parameters and the AIC and BIC values. Two joint models were analysed. In the first one, the longitudinal sub-model concerns the *cfu percentage* obtained by the student after each exam, and the survival sub-model is related to graduation. In the second joint model, the longitudinal sub-model concerns the *average grade* of all passed exams, and the survival sub-model is related to graduation. In formula:

$$\begin{aligned} y_i(t) &= \beta_0 + \beta_1 t_i + b_{i0} + b_{i1} t_i + \epsilon_i(t) \\ h_i(t) &= h_0(t) \exp[\gamma_1 gender + \gamma_2 age + \gamma_3 hsmark + \gamma_4 area + \alpha m_i(t)] \end{aligned} \tag{7.1}$$

where $m_i(t)$ is the *cfu percentage* or the *average grade* depending on the joint model considered, *gender* is coded as 0 for male and 1 for female, *age* is the age of the undergraduates at enrolment, *hsmark* is the mark obtained at high school final exam, and *area* indicates where the student lives at enrolment (1 = Milan, and 0 = Outside Milan).
The JM [40] package of R is used for the estimations. Table 7.2 reports the course in Management, Statistics and IT (MSIT) composed of 30 students, the longitudinal sub-model is a linear mixed-effects model for the *cfu percentage* while the survival sub-model is a relative risk model with two possible baseline risk functions, namely Weibull or the Piecewise-constant. For the piecewise-constant baseline hazard, the value of α is 25.3305, with $\exp(\alpha) = \exp(25.3305 \cdot 0.01) = 1.2883$, which means that an increase of 0.01 in the *cfu percentage*, increases the risk of graduation 1.2883 fold. Every covariate's parameter is significant at a level of 0.0001. As an example an increase of one year in the *age* at enrolment decreases the risk of graduation. This means that younger students are more likely to obtain a degree.

Table 7.2: Coefficients of joint model *cfu percentage*

	Weibull			Piecewise-constant		
	Value	Std. Err	p-value	Value	Std. Err	p-value
Longitudinal						
(Intercept)	0.4889	0.0381	<0.0001	0.4853	0.0381	<0.0001
timelog	0.0002	0.0001	0.0002	0.0002	0.0000	<0.0001
Event						
(Intercept)	-119.4774	39.3711	0.0024			
age	-0.4165	0.9692	0.6674	-0.7216	0.0195	<0.0001
gender	3.7927	1.6349	0.0204	3.0751	0.0030	<0.0001
hsmark	3.2476	2.1822	0.1367	2.3691	0.0015	<0.0001
area	1.5083	1.1857	0.2033	1.1180	0.0085	<0.0001
α	29.9640	9.8399	0.0023	25.3305	0.0011	<0.0001
log(shape)	2.5452	0.3307	<0.0001			
AIC	-474.1383			-454.6892		
BIC	-455.9227			-429.4677		

Female students are likely to obtain degree 21.65 times more than the male students. The *hsmark* and the *area* of living also have a positive effect on the risk of graduation. A similar effect that an increase in the *cfu percentage* has on the risk of the graduation is found in the model with a Weibull baseline hazard where the α parameter is positive, but in this model the other parameters are not significant, only the *gender* is significant at a level of 0.05 with a positive effect on the risk of graduation.
Concerning the model with piecewise-constant baseline hazard, Figure 7.3 shows some diagnostic graphs.

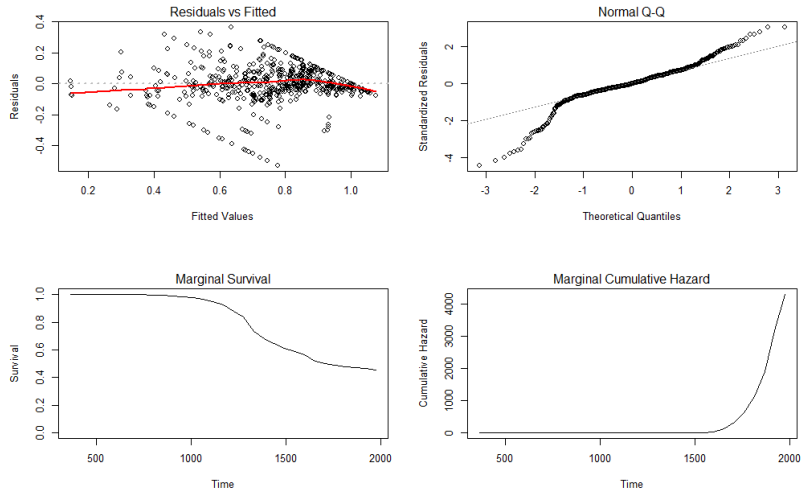

Figure 7.3: Diagnostic plots for the joint model of *cfu percentage* with piecewise-constant baseline

The graphs show that the residuals versus the fitted have few tendency, the data are normal except for the tail, and the estimated marginal and

marginal cumulative survival show that the survival function shrinks in the first period after three year and then it decreases slowly.
Figure 7.4 displays some diagnostic graphs for the Weibull baseline hazard model.

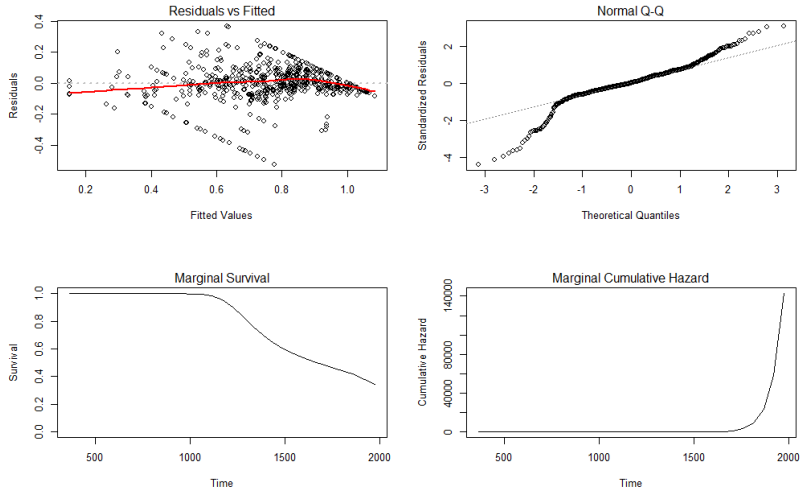

Figure 7.4: Diagnostic plots for the joint model of *cfu percentage* with Weibull baseline

The graphs shows that the residuals versus the fitted have few tendency, the data are normal except for the tail, and the estimated marginal and marginal cumulative survival show that after three years the survival function decreases steadily.
Table 7.3 considers the course in Management Accounting composed of 258 students. The longitudinal sub-model is a linear mixed-effects model of the *average grade*, while the survival sub-model is a relative risk model with Weibull as baseline risk function.

Table 7.3: Joint model *average grade*: coefficients

	Value	Std. Err	p-value
Longitudinal			
(Intercept)	22.8557	0.1814	<0.0001
timelog	0.0007	0.0001	<0.0001
Event			
(Intercept)	-42.0691	3.0021	<0.0001
age	-0.0873	0.0443	0.0486
gender	0.0090	0.1878	0.9619
area	0.2752	0.2853	0.3348
hsmark	0.0938	0.7809	0.9044
α	0.4155	0.0577	<0.0001
log(shape)	1.4891	0.0706	<0.0001

The parameter α is positive, which means that an increase of one unit in

the *average grade* increases the risk of graduation 1.5274 fold, as $\exp(\alpha) = \exp(0.4236) = 1.5274$. In this model only the enrolment *age* parameter is significant with a level of 0.05. An increase of *age* at enrolment decreases the risk of graduation. Figure 7.5 contains some diagnostic graphs.

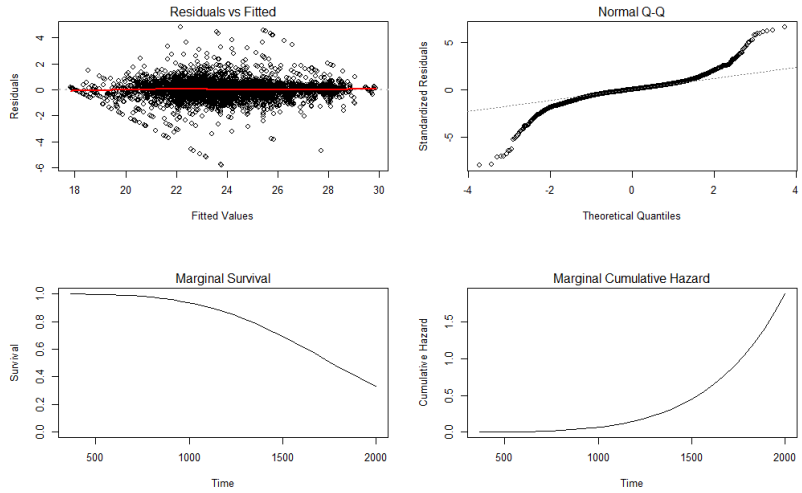

Figure 7.5: Diagnostic plots for the joint model of *average grade*

The graphs show that the residuals versus the fitted have no tendency, the data are normal except for the tail, and the estimated marginal and marginal cumulative survival show that after three years the survival function decreases stably.
Table 7.4 considers the course in Economics and Tourism composed of 124 students, the longitudinal sub-model is a linear mixed-effects model of *cfu percentage* while the survival sub-model is the Weibull accelerated failure time model. The accelerated failure time (AFT) model can be considered

Table 7.4: Joint model *cfu percentage* AFT: coefficients

	Value	Std. Err	p-value
Longitudinal			
(Intercept)	0.4195	0.0106	<0.0001
timelog	0.0002	0.0000	<0.0001
Event			
(Intercept)	7.9795	0.4335	<0.0001
age	0.0306	0.0174	0.0781
gender	0.0952	0.0679	0.1607
area	0.1346	0.0695	0.0530
hsmark	0.0647	0.2961	0.8271
α	-1.8420	0.2843	<0.0001
log(shape)	1.7025	0.1109	<0.0001

for the *cfu percentage* supposing that the *cfu percentage* has an accelerative effect on graduation. The α parameter is negative, that means that an

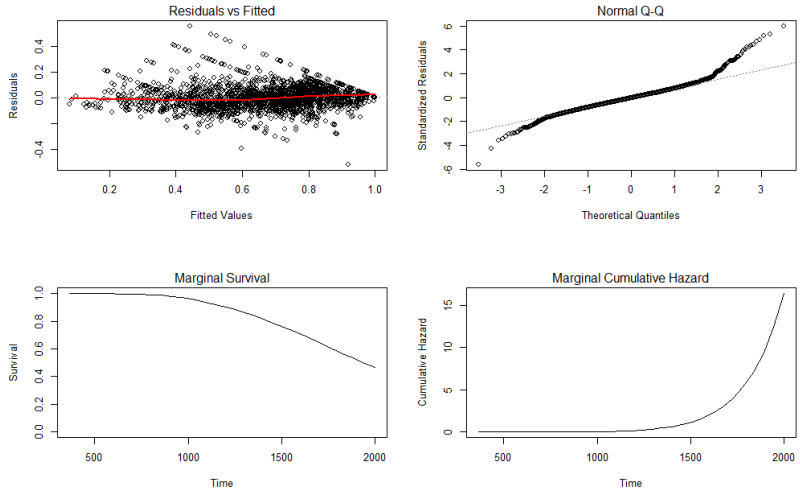

Figure 7.6: Diagnostic plots for the joint model of *cfu percentage* with AFT

increase of *cfu percentage* increases the hazard function. Figure 7.6 reports some diagnostic graphs. The graphs show that the residuals versus the fitted have no tendency, the data are normal except for the tail, and the estimated marginal and marginal cumulative survival show that after three years the survival function has a stable decrease.
The relation differs among the courses, supposedly related to the different undergraduate features in each course.
As there are two longitudinal covariates that are separately analysed, it would be interesting to analyse these two longitudinal covariates jointly in order to evaluate the jointly effect of both covariates on the risk of the event, as we suppose that the graduation is related to the *cfu percentage* but also the *average grade* may influence the event. Then we proposed a joint model in which the longitudinal sub-model is multivariate, or better in this case bivariate.

7.3 Study of length of stay at the university with two longitudinal covariates

Concerning the application using a bivariate longitudinal sub-model, the longitudinal sub-model concerns the *percentage of the cfu* obtained by the student after each exam and the *average grade* of all passed exams while the survival sub-model regards graduation. In formula:

$$\begin{aligned} y_{i1}(t) &= \beta_{01} + \beta_{11} t_i + b_{i01} + b_{i11} t_i + \epsilon_{i1}(t) \\ y_{i2}(t) &= \beta_{02} + \beta_{12} t_i + b_{i02} + b_{i12} t_i + \epsilon_{i2}(t) \\ h_i(t) &= h_0(t) \exp[\gamma_1 gender + \gamma_2 age + \gamma_3 hsmark + \gamma_4 hsclass + \\ &\quad \alpha_1 m_{i1}(t) + \alpha_2 m_{i2}(t)] \end{aligned} \tag{7.2}$$

where $m_{i1}(t)$ is the *cfu percentage*, $m_{i2}(t)$ indicates the *average grade*, and *hsclass* indicated the class (type) of high school attained (1 = Lyceum, and 0=Not Lyceum).
As for the univariate case, also for the multivariate case several models were estimated considering the available covariates. The models were then chosen on the basis of significant parameters and the AIC and BIC values.
Firstly the small data set composed by all the 30 students of Management, Statistics and IT course (MSIT) is analysed. The estimations are reported in Table 7.5.

Table 7.5: Results Management, Statistics and IT course data set

Models	1	2	3	4	5	6
fitmark	**0,5695 .**	0,7887 *	1,305 *	1,309 *	1,318 *	1,342 *
fitcfu	**19,54 ***	21,75 *	21,91 *	21,94 *	21,99 *	21,63*
hsmark		-3,653	-6,395 .	-6,188	-6,294	-6,259
area			1,371	1,357	1,365	1,311
age				0,2519	0,2534	0,2756
hsclass					0,03837	-0,06133
gender						-0,1375
Pseudo R-square	**0,045**	0,047	0,05	0,051	0,051	0,051
Likelihood ratio	**27,73**	28,96	30,92	30,99	30,99	31,01
Wald test	**9,6**	9,75	9,16	9,51	9,52	9,36
Score (logrank)	**14,31**	14,34	14,38	16,33	16,4	16,78
df	**2**	3	4	5	6	7
AIC	**42,25989**	43,03122	43,06887	44,99776	46,9953	48,97411
BIC	**51,04707**	56,21199	60,64323	66,96572	73,35685	79,72924

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Studying the AIC, the BIC and the pseudo R-square values, the best model is that one that considers in the hazard function only the two longitudinal covariates, $fitmark$ and $fitcfu$ ($\hat{m}_{i1}$ and $\hat{m}_{i2}$). The estimations for the parameter $\alpha_1 = 0,5695$ and $\alpha_2 = 19.54$ mean that an increase of one unit in the *mean mark* increases the risk of the event of 1.7674 fold ($\exp(0.5695) = 1.7674$), while an increase of 0.01 in the *cfu percentage* increases the risk of the event of 1.2158 fold ($\exp(19.54 \cdot 0.01) = 1.2158$).
Comparing the survival function estimated through the Kaplan-Meier estimation without considering any covariate and the survival function obtained from the model just analysed (Figure 7.7), it is possible to see that the two curves are very close.
Some diagnostic graphs can be moreover analysed, studying for example the residuals which are all very close to 0 (Figure 7.8).
If the proportional hazard model is a good model the survival distribution of the Cox-Snell Residuals must be near the exponential function [7, 24]. In this case it is possible to observe that the two distributions are very close (Figure 7.9).
As the data set composed by all the students of the Management, Statistics and IT course is small, we decided to analyse the data set composed by all the male undergraduates of the Economic Faculty (543 students). The estimations of the parameter are shown in Table 7.6.

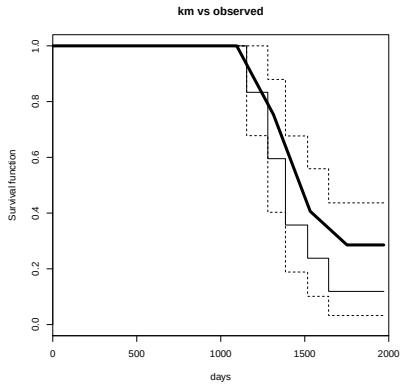

Figure 7.7: KM and survival function MSIT

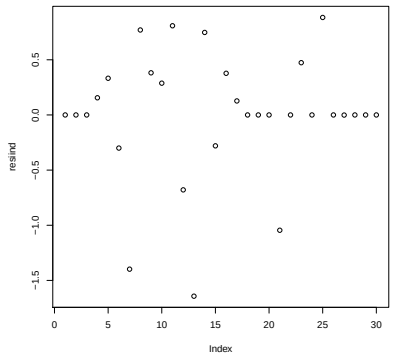

Figure 7.8: Residuals MSIT

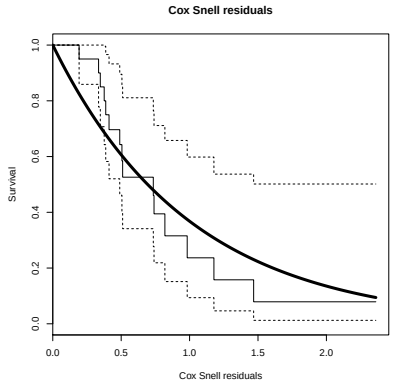

Figure 7.9: Cox-Snell Residuals MSIT

Studying the AIC, the BIC and the pseudo R-square values, the best model is that one that considers in the hazard function the two longitudinal covariates, and the variables *hsmark* and *area*. The estimations for the parameter $\alpha_1 = 0,0166$ and $\alpha_2 = 24,29$ mean that an increase of one unit in the *mean mark* increases the risk of the event of 1.1258 fold ($\exp(0.1185) = 1.1258$), while an increase of 0.01 in the *cfu percentage* increases the risk of the event

Table 7.6: Results Male undergraduates (543 students)

Models	1	2	3	4	5
fitmark	0,0166	0,1112 *	**0,1185 ***	0,1090 *	0,11 *
fitcfu	24,29 ***	26,93 ***	**27,25 *** **	27,38 ***	27,38 ***
hsmark		-2,656 ***	**-2,687 *** **	-2,506 ***	-2,589 ***
area			**-0,35 ***	-0,3432 .	-0,3541 *
hsclass				-0,1647	-0,1474
age					-0,03042
Pseudo R-square	0,077	0,079	**0,08**	0,08	0,08
Likelihood ratio	803,9	822,8	**826,4**	827,8	827,9
Wald test	281,3	232,3	**229,2**	227,8	227,8
Score (logrank)	466,8	467	**467**	467	467,4
df	2	3	**4**	5	6
AIC	1600,324	1583,367	**1581,792**	1582,433	1584,272
BIC	1614,739	1604,991	**1610,623**	1618,471	1627,518

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

of 1.3132 fold ($\exp(27.25 \cdot 0.01) = 1.3132$). In addition the parameter for the variable *hsmark* means that an increase of the *hsmark* decreases the risk of the event, while concerning the variable *area* passing from area 0 (not Milan) to area 1 (Milan) the risk of the event decreases.
Comparing the survival function estimated through the Kaplan-Meier estimation without considering any covariate and the survival function obtained from the model just analysed (Figure 7.10), it is possible to see that the two curves are very close, where the survival function obtained by the model suggested is the thicker one.

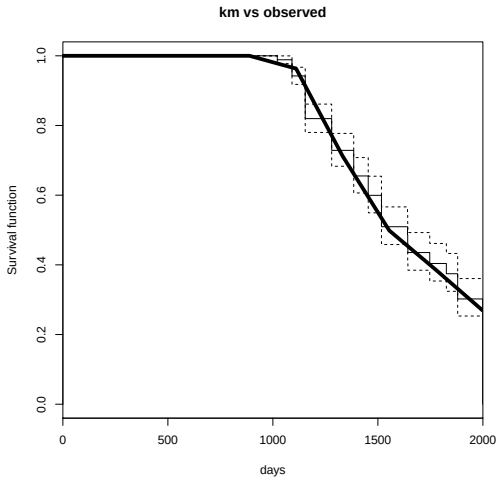

Figure 7.10: KM and survival function male sub-population

Some diagnostic graphs can be moreover analysed, studying for example the residuals which are all very close to 0 (Figure 7.11), except for few outliers. The survival distribution of the Cox-Snell Residuals is close to the expo-

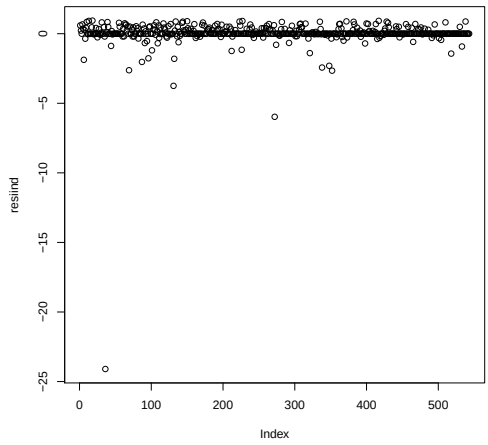

Figure 7.11: Residuals male sub-population

nential function (Figure 7.12), except for a little central area.

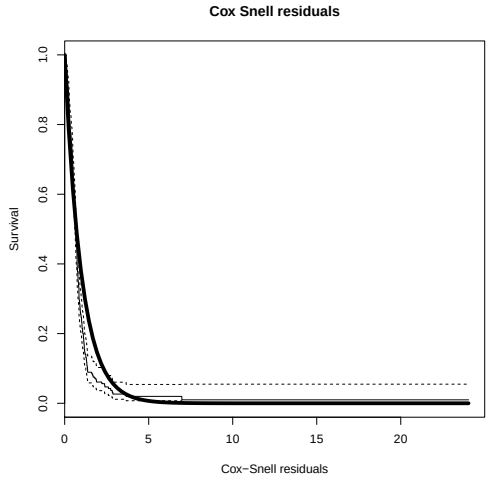

Figure 7.12: Cox-Snell Residuals male sub-population

Subsequently it is interesting to check if the multivariate joint models give more information concerning the survival function than the univariate joint models. We decided to create an index that is able to summarise the situation, considering the differences between the survival function:

$$\delta(t) = \frac{\hat{S}_{KM}(t) - \hat{S}_k(t)}{\hat{S}_{KM}(t) - \hat{S}_1(t)} \tag{7.3}$$

where $\hat{S}_k(t)$ indicates the estimation of survival function at time t for the joint models with k longitudinal covariates, while $\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} \frac{r_i - d_i}{r_i}$ indicates the Kaplan and Meier estimation of the survival function at time t. Thus if the value of this index is lower than one this indicates that the survival function of the multivariate joint models is nearer to the the survival

function estimated through the Kaplan-Meier estimation than the survival function obtained by the univariate joint model, as the numerator of the index is slower than the denominator.
In Figure 7.13 there are the values obtained for this index for the male sub-population comparing the univariate joint model that consider only the $cfupercentage$ with the bivariate joint model. It is possible to assert that the introduction of a new covariate increases the estimation of the survival function as, except for few time instants, the value of the index is always lower than 1.

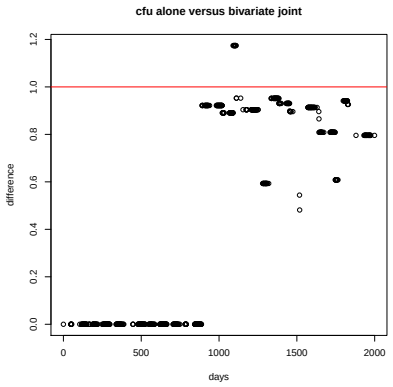

Figure 7.13: Univariate versus bivariate male

In addition it is interesting to comparing the survival function of two different profiles, changing the value of a covariate. For example in Figure 7.14 there are two survival functions, one for a male that obtained 95 as high school mark (the solid one) and one for a male obtained 60 as high school mark (the dashed one). The survival function for the first student is higher than the survival function for the second student, then the second student has a better situation, as in this case the event is a positive thing: the graduation.

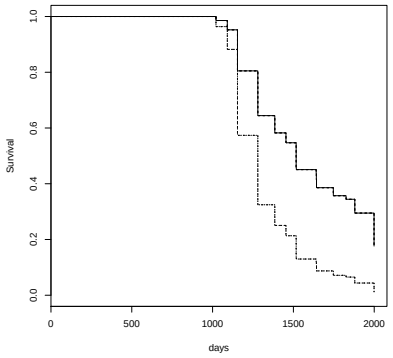

Figure 7.14: Difference survival function for student obtaining high school mark 95 (solid) and 60 (dashed)

As the variables analysed are two, it is interesting to observe also the dif-

ferences in two survival function changing the value in *area*. For example in Figure 7.15 there are two survival function, one for a male that lives in Milan (the dashed one) and one for a male that lives out of Milan (the solid one). The survival function for the first student is higher than the survival function for the second student, then the second student has a better situation, as in this case the event is a positive thing: the graduation.

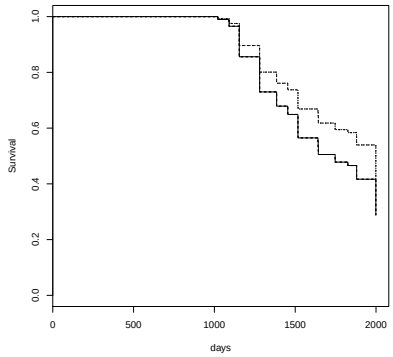

Figure 7.15: Difference survival function for student living in Milan (dashed) or not (solid)

After analysing the situation for the male sub-population, it is interesting studying the model for the female sub-population (672 students). The estimations are shown in Table 7.7.

Table 7.7: Results Female undergraduates (672 students)

Models	1	2	3	4	5
fitmark	0,1007 **	**0,1122 *** **	0,1165***	0,1229 ***	0,1236 ***
fitcfu	46,39 ***	**47,83 *** **	48,23 ***	48,45 ***	48,44 ***
area		**0,4491 ***	0,4446 *	0,442 *	0,4447 *
hsclass			0,1686	0,19	0,1917
hsmark				-0,2733	-0,2707
age					0,006524
Pseudo R-square	0,091	**0,092**	0,092	0,092	0,092
Likelihood ratio	1311	**1318**	1320	1320	1320
Wald test	176,1	**164,9**	163,5	161,6	161,6
Score (logrank)	555,3	**555,5**	556,1	556,4	557,3
df	2	**3**	4	5	6
AIC	2438,889	**2434,334**	2434,263	2435,937	2437,899
BIC	2453,942	**2456,915**	2464,37	2473,571	3483,06

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Studying the AIC, the BIC and the pseudo R-square values, the best model is that one that considers in the hazard function the two longitudinal covariates, and the variable *area*. The estimations for the parameter $\alpha_1 = 0.1007$ and $\alpha_2 = 46.39$ mean that an increase of one unit in the *mean mark* increases the risk of the event of 1.1187 fold ($\exp(0.1122) = 1.1187$), while

an increase of 0.01 in the *cfu percentage* increases the risk of the event of 1.6133 fold ($\exp(47.83 \cdot 0.01) = 1.6133$). In addition the parameter for the variable *area* means that passing from area 0 (not Milan) to area 1 (Milan) the risk of the event increases.
Comparing the survival function estimated through the Kaplan-Meier estimation without considering any covariate and the survival function obtained from the model just analysed (Figure 7.16), it is possible to see that the two curves are very close, where the survival function obtained by the model suggested is the thicker one.

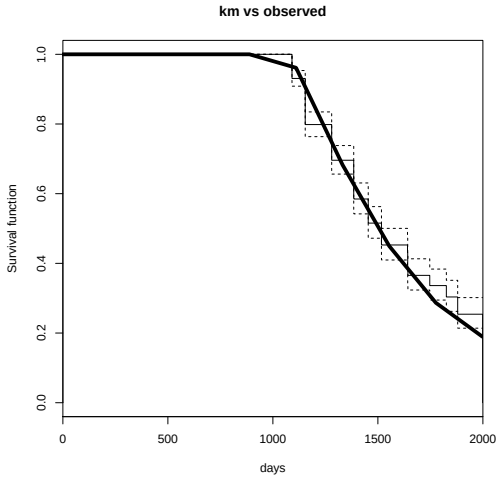

Figure 7.16: KM and survival function female sub-population

Some diagnostic graphs can be moreover analysed, studying for example the residuals which are all very close to 0 (Figure 7.17).

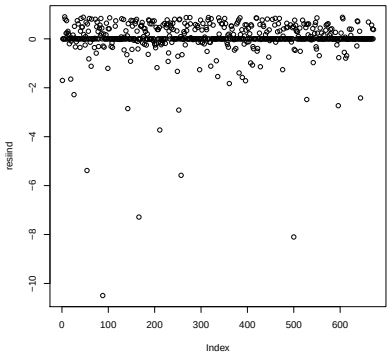

Figure 7.17: Residuals female sub-population

The survival distribution of the Cox-Snell Residuals in close to the exponential function (Figure 7.18).
Also for the female sub-population it is interesting to analyse the gain in survival function estimation with the introduction of a new longitudinal covariate. Then the values obtained for the index previously introduced for the

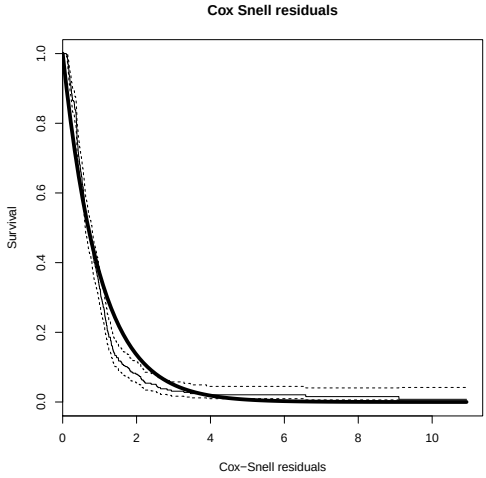

Figure 7.18: Cox-Snell Residuals

female sub-population are evaluated comparing the univariate joint model that consider only the $cfupercentage$ with the bivariate joint model. The results are showed in Figure 7.19. It is possible to assert that the introduction of a new covariate increases the estimation of the survival function as, except for few time instants, the value of the index is always lower than 1.

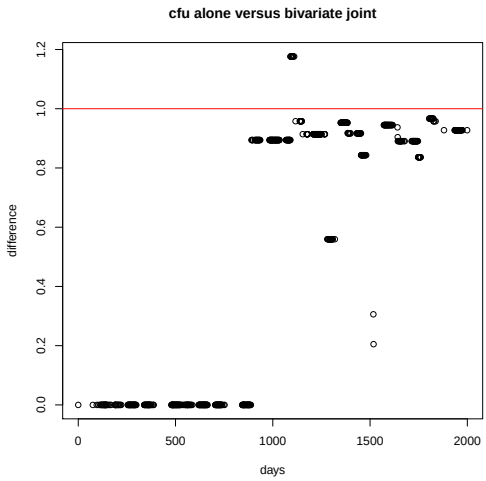

Figure 7.19: Univariate versus multivariate

As for the male sub-population, also for the female sub-population it is interesting to comparing the survival function of two different profiles, changing the value of a covariate. For example in Figure 7.20 there are two survival functions, one for a female that lives in Milan (the dashed one) and one for a female that lives outside Milan (the solid one). The survival function for the second student is higher than the survival function for the first student, then the first student has a better situation, as in this case the event is a positive thing: the graduation.

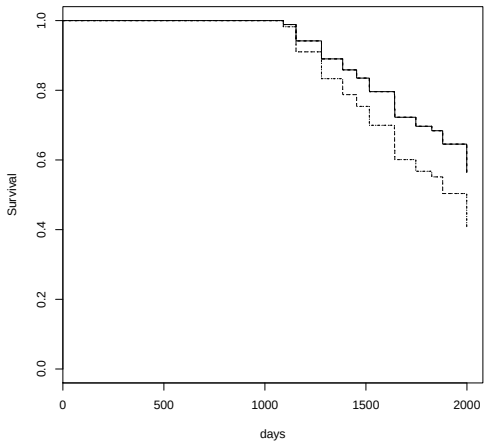

Figure 7.20: Difference survival function for student living in Milan (dashed) or not (solid)

7.4 Conclusion

In this chapter, it is proposed the use of joint models in a new type of application for analysing of the timing of student graduation. In particular the relation between *cfu percentage* and the *average grade* of the exams with graduation was investigated using a joint model of longitudinal and time-to-event data. As shown, the parameters that indicates the association are significant and indicates the effect of the longitudinal covariates on the risk of the graduation.
In the univariate applications the relation between the single longitudinal covariates and the graduation differs among the courses, supposedly related to the different undergraduate features in each course.
In the bivariate applications the results obtained permit to assert that the joint models allow to estimate the influence of two or more longitudinal covariates on the graduation. In these applications the two longitudinal covariates considered are the *cfu percentage* and the *average grade* of the exams. The estimations are significant and permit to assert that an increase of the two longitudinal covariate considerably increases the risk of graduation. In addition, in some of the models presented, is also possible to consider the effect of some exogenous covariates, such as *area* and *hsmark*.
The fact that the introduction of a new covariate has a positive effect is confirmed by the analysis of the improvement in the estimation of the survival function. This improvement is summarise by the introduction of a new simple index, which, through the graphical representation, is able to give us a fast answer if there is an improvement or not.
The results are encouraging and deal to several ideas of future application work. In fact further application work concerns a competing risks model in order to jointly analyse the graduation and the drop-out in a survival competing risks model, with two possible univariate longitudinal sub-models or

a bivariate longitudinal sub-model for the *cfu percentage* and *average grade* of the exams.
Perhaps, instead of considering the *cfu percentage*, which has value between 0 and 1, it would be possible to transform it into a binary variable considering whether the student has sufficient requirements in terms of cfu (for example greater than 80% for each academic year).

Chapter 8

Conclusions and future work

This thesis concerns the joint models for longitudinal and time-to-event data which is a recent family of models that jointly analyses longitudinal and survival data. The models are composed by two sub-models, the longitudinal and the survival sub-model. Then as a first step we presented a brief review of the longitudinal and survival models used, which are the sub-models of the joint models. Particularly the linear mixed models and the Cox models with proportional hazard function are analysed.
Subsequently we presented a review of the univariate joint models that permit to evaluate the influence of one longitudinal covariate on the hazard function, mentioning the definitions and the methods of estimation presented in literature. These models had been analysed from several researchers, then there is a very extensive literature available. The formulation of the joint models are often similar, with the addition of different assumptions on the distribution of the covariates or with the use of different methods of estimation, both Bayesian or frequentist.
After analysing the classical joint model, it is interesting to analyse the possible extensions of the joint models, that help to deal with different situations. For example several extensions help to deal with heterogeneity in the sample or they could consider relations between the risk of the event and the longitudinal covariates which are not linear.
Another interesting extension concerns the case in which one or both submodel are multivariate. We focused on the extension in which only the longitudinal sub-model is multivariate in order to investigate if more than one longitudinal covariates influenced the risk of the event. We decided to focus on this extension as we think that an increase of information, given by the addition of one or more longitudinal covariates in the model, will deal to better estimations of the survival function and of the level of the longitudinal covariates influence on the risk of the event. We proved that the joint model with multivariate longitudinal sub-model allows to obtain better results than using only one longitudinal covariate, as shown by the comparison of the estimated survival function in the applications.
Subsequently we presented a review of the joint models with multivariate longitudinal sub-model. Several authors defined the longitudinal sub-model as a multivariate linear mixed model and the survival sub-model as an haz-

ard function expressed as a function of the longitudinal covariates jointly considered. Concerning the methods of estimation the authors proposed different solutions, which often are the generalisation of a method of estimation used when both sub-models are univariate, for example the Markov chain Monte Carlo algorithm [63], the one-step-late EM algorithm [32], the semiparametric conditional score estimation [50], the Gibbs sampling algorithm [26]. Alternately the authors used different strategy for the estimation like for example a pairwise modelling strategy [16], or the use of a latent variable [39].

As said the most important problem related to the multivariate situation concerns the computational aspect of the estimation. In fact considering that the univariate case is computational demanding, increasing the number of the parameters or the dimension of the sub-models will lead to higher computational demanding situations. In order to solve this problem sometimes the authors focused on possible assumptions on the distribution of the covariates, on different formulation of longitudinal sub-models, or on the use of Bayesian methods for the estimations.

For solving the computational problem related to the joint models, extended by the extension of the longitudinal sub-model, we instead proposed as a first possible solution a two-stage estimation method that allows to obtain very fast and with desirable proprieties estimations. This method is deeply analysed presenting the sub-models used and the steps for obtaining the estimations, focusing on the parameters that indicate the relation between the hazard function and the longitudinal covariates.

We applied the joint model to study the length of stay before graduation of university students, thus the aim is to quantify the relation between this event, the graduation, with one or more longitudinal covariates. At first we implemented two different joint models with univariate longitudinal sub-models, in order to investigate the influence of the *cfu percentage* and the *mean mark* on the risk of graduation. The results showed a significant positive relation between the risk of the graduation and the longitudinal covariate individually analysed.

We think that both longitudinal covariates can jointly influenced the risk of graduation, as the mark obtained at the exam may influence the time to graduation, not only the amount of cfu already obtained from the exams passed. Then implemented an algorithm in which the longitudinal submodel is multivariate, in our case bivariate, and the hazard function is expressed as a function of both longitudinal covariates jointly considered. The results showed that there is a significant positive relation between the longitudinal covariates and the risk of graduation.

In addition we compared the results obtained by a joint model that considers only one longitudinal covariate with a joint model that considers two longitudinal covariate through comparison of the survival function estimated. In order to summarise the improvement in estimation, we proposed a new index which is based on the distance between the survival function obtained by the Kaplan-Meier estimation and the survival functions estimated by two joint models, one with univariate and the other one with bivariate lon-

gitudinal sub-model. The values of this index suggest that adding a new longitudinal covariate gives a better estimation of the survival function.
At the end of the applications we conclude that this data set could be further investigated, by the implementation of other extensions of the joint models, like for example considering two opposing terminal event, graduation and drop-out, with an univariate or multivariate longitudinal sub-model.
Further work concerns a deeper analysis of the two-stage approach and the implementation of the EM algorithm, extending Rizopoulos [43] model with multivariate longitudinal sub-models. The aim is to make comparison between the two methods of estimation: the two-stage and the joint likelihood approach, comparing the efficiency and the efficacy of the two methods of estimation.
In addition other types of extensions may be considered. For example extending the survival sub-model considering multivariate survival sub-model, like competing risk or recurrent events, with univariate or multivariate longitudinal sub-models.

Appendix

As the most of the applications of the joint models focused on the medical data, we decided to analyse a medical data set which is highly studied, where the event of interest in the death of the patient. The data set is composed by 312 randomised patients with primary biliary cirrhosis, a rare autoimmune liver disease, at Mayo Clinic. The data frame has got 1945 observations and 20 variables. These are the variables analysed:

- id: patients identifier
- years: number of years between registration and the earlier of death, transplantion, or study analysis time.
- status: a factor with levels alive, transplanted and dead.
- drug: a factor with levels placebo and D-penicil.
- age: at registration in years.
- sex: a factor with levels male and female.
- year: number of years between enrollment and this visit date
- serBilir: serum bilirubin in mg/dl.
- albumin: albumin in gm/dl.
- SGOT: SGOT in U/ml.
- prothrombin: prothrombin time in seconds.

The results obtained from the applications of the two-stage estimation are shown in Table 8.1, introducing all the different variables in the model and considering *serBilir* and *albumin* as the longitudinal covariates that characterised the linear mixed model. Analysing the AIC, the BIC and the pseudo R-square values, the best model is that one that considers only four covariates in the hazard function:

$$h_i(t|M_i(t),\omega_i) = h_0(t) \exp\left[\gamma_1 age + \gamma_2 prothrombin + \alpha_1 fitbilirubin + \alpha_2 fitalbumin\right]$$

where $fitbilirubin$ and $fitalbumin$ are the value imputed ($\hat{m}_{i1}$ and $\hat{m}_{i2}$). The most important parameters are $\hat{\alpha}_1 = 0,0756$ and $\hat{\alpha}_2 = -1,4289$. This

Table 8.1: Results pbc data set

Models	1	2	3	4	5	6
Fitbilirubin	0,0733***	0,0813***	**0,0756*** **	0,0756***	0,0691***	0,0688***
Fitalbumin	-1,6190***	-1,5041***	**-1,4289*** **	-1,4356***	-1,3362***	-1,3321***
age		0,0439***	**0,0415*** **	0,0393***	0,0440***	0,0442***
prothrombin			**0,0532 ***	0,0556 *	0,0574 *	0,0574 *
sex				-0,2625	-0,2292	-0,2318
fitsgot					0,0021	0,0021
drug						-0,0343
Pseudo R-square	0,064	0,076	**0,077**	0,078	0,079	0,079
Likelihood ratio	138,8	164,2	**168,1**	169,4	170,9	170,9
Wald test	147,6	169,6	**180,6**	179,9	177,1	177,6
Score(logrank)	188,4	215,8	**229,7**	230,3	230,5	213,1
df	2	3	**4**	5	6	7
AIC	1246,599	1223,237	**1221,291**	1221,985	1222,493	1224,454
BIC	1257,884	1240,164	**1243,861**	1250,198	1256,348	1263,952

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

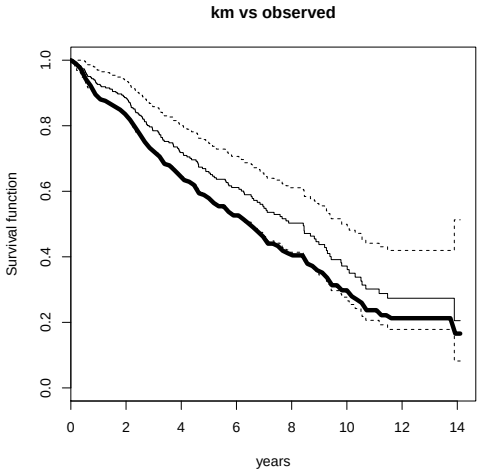

Figure 8.1: Kaplan-Meier (KM) versus Survival function

parameters mean that an increase of the $fitbilirubin$ increases the risk of the event, while an increase of the $fitalbumin$ decreases the risk of the event. In addition the parameter for the covariates *age* and *prothrombin* are positive, thus an increase of the value of these covariates rises the risk of the event.
Subsequently it is interesting to make a comparison between the survival function estimated through the Kaplan-Meier estimation without considering any covariate (the thinner one in Figure 8.1) and the survival function obtained from the model just analysed (the thicker one). As the graph shows the two curves are not so different and distant.

Bibliography

[1] P. Albert and J. Shih. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. *The Annals of Applied Statistics*, 4:1517–1532, 2010.

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

[3] J. Barrett, P. Diggle, R. Henderson, and D. Taylor-Robinson. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society*, 77:131–148, 2015.

[4] E. Brown, J. Ibrahim, and V. DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61:64–73, 2005.

[5] J. Choi, S. Anderson, T. Richards, and W. Thompson. Prediction of transplant-free survival in idiopathic pulmonary fibrosis patients using joint models for event times and mixed multivariate longitudinal data. *Journal of Applied Statistics*, 41:2192–2205, 2014.

[6] D. Cox. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[7] D. Cox and E. Snell. A general definition of residuals. *Journal of the Royal Statistical Society*, 30:248–275, 1968.

[8] K. Das, R. Li, Z. Huang, J. Gai, and R. Wu. A bayesian framework for functional mapping through joint modeling of longitudinal and time-to-event data. *International Journal of Plant Genomics*, pages 1–12, 2012.

[9] V. De Gruttola and X. Tu. Modeling The Relationship Between Progression Of CD4-Lymphocyte Count And Survival Time. In N. Jewell, K. Dietz, and Vernon T. Farewell, editors, *AIDS Epidemiology Methodological Issues*, pages 275–296. Springer Science, 1992.

[10] V. De Gruttola and X. Tu. Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003–1014, 1994.

[11] V. De Gruttola, M. Wulfsohn, M. Fischl, and A. Tsiatis. Modelling the relation between survival and cd4 lymphocytes in patients with aids and aids-related complex. *Journal of the Acquired Immune Deficiency Syndrome*, 6:359–365, 1993.

[12] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

[13] P. Diggle, I. Sousa, and A. Chewynd. Joint modelling of repeated measurements and time-to-event outcomes: the fourth armitage lecture. *Statistics in medicine*, 27:2981–2998, 2008.

[14] A. Dobson and R. Henderson. Diagnostics for joint longitudinal and dropout time modeling. *Biometrics*, 59:741–751, 2003.

[15] C. Faucett and D. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in Medicine*, 15:1663–1685, 1996.

[16] S. Fieuws, G. Verbeke, B. Maes, and Y. Vanrenterghem. Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, 9:419–431, 2008.

[17] D. Follman and M. Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51:151–168, 1995.

[18] P. Ghisletta. Application of a joint multivariate longitudinal–survival analysis to examine the terminal decline hypothesis in the swiss interdisciplinary longitudinal study on the oldest old. *Journal of Gerontology: Psychological Sciences*, 63B:185–192, 2008.

[19] M. Greenwood. The natural duration of cancer. *Reports on Public Health and Medical Subjects*, 33:1–26, 1926.

[20] B. He and S. Luo. Joint modeling of multivariate longitudinal measurements and survival data with applications to parkinson's disease. *Statistical Methods in Medical Research*, 0:1–13, 2013.

[21] D. Hedecker and R. Gibbons. *Longitudinal Data Analysis*. Wiley, Hoboken, 2006.

[22] A. Henderson, V. De Gruttola, and M. Wulfsohn. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480, 2000.

[23] J. Hogan and N. Laird. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, 16:259–272, 1997.

[24] D. Hosmer, S. Lemeshow, and S. May. *Applied Survival Analysis, Regression Modeling of Time-to-Event Data*. John Wiley and Sons, Hoboken, 2008.

[25] F. Hsieh, Y. Tseng, and J. Wang. Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics*, 62:1037–1043, 2006.

[26] J. Ibrahim, M. Chen, and D. Sinha. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, 14:863–883, 2004.

[27] J. Ibrahim, H. Chu, and M. Chen. Basic concepts and methods for joint model of longitudinal ans survival data. *Journal of Clinical Oncology*, 28:2796–2801, 2010.

[28] E. Kaplan and P. Meier. Nonparametric estimation for incomplete observations. *Journal of the American Statistical Association*, 93:457–481, 1958.

[29] N. Laird and J. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

[30] N. Law, J. Taylor, and H. Sandler. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*, 3:547–563, 2002.

[31] N. Li, R. Elashoff, and G. Li. Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biometrical Journal*, 51:19–30, 2009.

[32] H. Lin, C. McCullocj, and S. Mayne. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21:2369–2382, 2001.

[33] R. Little. Modelling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90:1112–1121, 1995.

[34] L. McCrink, A. Marshall, and K. Cairns. Advanced in joint modelling: A review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81:249–269, 2013.

[35] G. Molennberghs and G. Verbeke. *Models for Discrete Longitudinal Data*. Springer-Verlang, New York, 2005.

[36] M. Murawska, D. Rizopoulos, and E. Lesaffre. A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *Journal of Probability and Statistics*, pages 1–18, 2012.

[37] Y. Pawitan and S. Self. Modeling disease market processes in aids. *Journal of the American Statistical Association*, 88:719–726, 1993.

[38] C. Proust-Lima, P. Joly, J. Dartigues, and H. Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics and Data Analysis*, 53:1142–1154, 2009.

[39] C. Proust-Lima, P. Joly, J. Dartigues, and H. Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics and Data Analysis*, 53:1142–1154, 2009.

[40] D. Rizopoulos. Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35, 2010.

[41] D. Rizopoulos. Dynamic predictions and prospective accurancy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829, 2011.

[42] D. Rizopoulos. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics and Data Analysis*, 56:491–501, 2012.

[43] D. Rizopoulos. *Joint model for Longitudinal and Time-to-Event Data with applications in R.* CRC Press, Boca Raton, 2012.

[44] D. Rizopoulos and P. Ghosh. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30:1366–1380, 2011.

[45] D. Rizopoulos, G. Verbeke, and E. Lesaffre. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society*, 71:637–654, 2009.

[46] D. Rizopoulos, G. Verbeke, and G. Molenberghs. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*, 66:20–29, 2010.

[47] D. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976.

[48] S. Self and Y. Pawitan. Modeling a marker of disease progression and onset of disease. In N. Jewell, K. Dietz, and Vernon T. Farewell, editors, *AIDS Epidemiology Methodological Issues*, pages 231–255. Springer Science, 1992.

[49] P. Slasor and N. Laird. Joint model for efficient estimation in proportional hazards regression models. *Stastistics in Medicine*, 22:2137–2148, 2003.

[50] X. Song, M. Davidian, and A. Tsiatis. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, 3:511–528, 2002.

[51] J. Sweeting and S. Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53:750–763, 2011.

[52] Y. Tseng, F. Hsieh, and J. Wang. Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92:587–603, 2005.

[53] A. Tsiatis and M. Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88:447–458, 2001.

[54] A. Tsiatis and M. Davidian. Joint modeling of longitudinal an time-to-event data: An overview. *Statistica Sinica*, 14:809–834, 2004.

[55] A. Tsiatis, V. De Gruttola, and M. Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90:27–37, 1995.

[56] G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer-Verlang, New York, 2000.

[57] Y. Wang and M. Taylor. Jointly modellling longitudinal and event time data with applciation to acquired immunodeficienecy syndrome. *Journal of the American Statistical Association*, 455:895–905, 2001.

[58] L. Wu, X. Hu, and H. Wu. Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics*, 9:308–320, 2008.

[59] L. Wu, W. Liu, and X. Hu. Joint inference on hiv dynamics and immune suppression in presence of measurement errors. *Biometrics*, 66:327–335, 2010.

[60] M. Wu and K. Bailey. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics*, 45:939–955, 1989.

[61] M. Wu and R. Carroll. Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process. *Biometrics*, 44:175–188, 1988.

[62] M. Wulfsohn and A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339, 1997.

[63] J. Xu and S. Zeger. The evaluation of multiple surrogate endpoints. *Biometrics*, 57:81–87, 2001.

[64] J. Xu and S. Zeger. Joint analysis of longitudinal data comprising repeated measures and times to events. *Applies Statistics*, 50:375–387, 2001.

[65] W. Ye, X. Lin, and J. Taylor. Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. *Biometrics*, 64:1238–1246, 2008.

[66] M. Yu, N. Law, J. Taylor, and H. Sandler. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14:835–862, 2004.

[67] Y. Zheng and P. Heagerty. Prospective accuracy for longitudinal markers. *Biometrics*, 63:332–341, 2007.