REVSTAT – Statistical Journal
Volume 9, Number 1, March 2011, 57–81

A REVIEW ON JOINT MODELLING OF LONGITUDINAL MEASUREMENTS AND TIME-TO-EVENT

Authors: Inês Sousa
– Departamento de Matemática e Aplicações,
Universidade do Minho, Portugal
isousa@math.uminho.pt

Abstract:

- In longitudinal studies subjects are measured for one or more response variable, over time. Although the underlying evolution of such response variables is continuous in time, in practice the measurements are observed at discrete time points. In longitudinal clinical trials it is also common to observe relevant events, generating time-to-event data. If both types of data are available, we might be interested in the association between the two processes, longitudinal and time-to-event. Commonly, when death is considered the event, the observation sequence of longitudinal measurements is terminated by the event process. When the two observed processes are related, the analysis of the data set should be suited to the specific objectives. We distinguish three situations: if the interest is to analyse the longitudinal outcome response variable with drop-out at the time-to-event; to analyse time-to-event, whilst exploiting correlation with a noisy version of a time-varying risk factor; or to analyse the relationship between the two processes. Joint models assume a full distribution for the joint distribution of longitudinal and time-to-event processes, which includes a description of the relation between the two processes.

Key-Words:

- *longitudinal; time-to-event; survival; Gaussian; correlation structure.*

AMS Subject Classification:

- 62F12.

1. INTRODUCTION

Longitudinal studies are characterised by observation of repeated measurements on a number of subjects at a series of time points. In this work we will only consider continuous response variables. It is of interest, particularly in longitudinal clinical trials, to test significant differences between the underlying processes of the same response variables for different treatment groups.

Time to event data are a set of times on individuals, induced by multiple or single events. We will for this work only consider single events. In clinical trials patients are usually assigned to different treatment groups, or in different age or gender groups. Therefore, the aim of time-to-event analysis is to identify differences in the time-to-event distributions of different groups.

In medical studies it is common to have data on repeated measurements jointly with time-to-event. The interest on data analysis is sometimes on the analysis of time-to-event, allowing for correlation with a time dependent variable, or on the analysis of longitudinal outcome with potentially informative missing data. Individual longitudinal and survival models might be considered. However, the notion of joint modelling is motivated in a setting of dependent longitudinal and time-to-event data.

If the interest of inference is on the association between the response variable and the survival mechanism, the two processes have to be modelled jointly, including parameters that represent their correlation. The proposal goes to the so called joint models for longitudinal and time-to-event. These models are based on a joint distribution for the two processes, longitudinal and failure time.

2. LONGITUDINAL DATA ANALYSIS

A longitudinal data set is characterised by repeated measurements of one or more response variables on a number of subjects at a series of time points. We first introduce linear models for repeated measurements with focus on general linear mixed effects models. For the analysis of repeated measurements it is common to assume independence between subjects, to have the replication across subjects for the analysis of time trajectory. However, this assumption is not adequate for measurements within the same subject, as measurements in time from a same person tend to be correlated. Moreover, measurements from different subjects and within a same individual are also subject to measurement error.

2.1. Notation

In the context of repeated measurements of a response variable, we let Y_{ij} be a response variable measured on subject $i = 1, ..., n$ at time point t_{ij}, with $j = 1, ..., m_i$. We include a set of p explanatory variables given by the vector $\boldsymbol{x}_{ij}$ with dimension p, which can be time dependent or only measured at baseline. The full set of repeated measurements for subject i is represented by the vector $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{im_i})$, with mean $\mathrm{E}[\boldsymbol{Y}_i] = \boldsymbol{\mu}_i$, and variance covariance matrix $\mathrm{Var}(\boldsymbol{Y}_i) = V_i$ of dimension $(m_i \times m_i)$, where each element (j, k) of this matrix is the covariance $\mathrm{Cov}(Y_{ij}, Y_{ik}) = v_{ijk}$, and $\mathrm{Var}(Y_{ij}) = v_{ij}$, for $j = k$.

The most common model-based approach for longitudinal repeated measurements assumes independence between subjects i, where each measurement is a realisation of a Gaussian random variable. The linear model is based on the regression of explanatory variables:

$$Y_{ij} \;=\; \mu_i(t_{ij}) + \epsilon_{ij} \;. \tag{2.1}$$

Different models for longitudinal data differ on the correlation structure for the errors ϵ_{ij}. For the entire data set of $N = \sum_{i=1}^{n} m_i$ longitudinal measurements, we use the notation $\boldsymbol{Y} = (\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n)$ as the random variable of all measurements for all subjects, with the linear model for longitudinal measurements as

$$\boldsymbol{Y} \sim \; MVN\big(\boldsymbol{X\beta}, \boldsymbol{V}(\psi)\big) \,,$$

where $\boldsymbol{X}$ is the $(N \times p)$ design matrix of explanatory variables. The matrix $\boldsymbol{V}$, with dimension $(N \times N)$ and parameters ψ, is a block diagonal matrix, because we assume independence between subjects, with each diagonal matrix $\boldsymbol{V}_i$ representing the variance covariance matrix for subject i.

2.2. General linear mixed models

We will be using linear longitudinal models as defined previously, with ideas from [1] and [2]. The general idea of linear mixed effects models is to assume a structure for the ϵ_{ij}'s as in (2.1), separating pure measurement error from variability between and within individuals. The general linear mixed effects model is defined as

$$Y_{ij} \;=\; \mu_i(t_{ij}) + \Omega_i(t_{ij}) + Z_{ij} \;, \tag{2.2}$$

where $\Omega_i(t_{ij})$ is an unobserved random process, and Z_{ij} are independent realisations of a zero-mean Gaussian random variable with variance τ^2, representing

pure measurement error. Diggle *et al.* [2] propose to decompose the unobserved random process $\Omega_i(t_{ij})$ into two components in an additive way,

$$\Omega_i(t_{ij}) \;=\; \boldsymbol{d}'_{ij}\,\boldsymbol{U}_i + W_i(t_{ij}) \;,$$

where $\boldsymbol{U}_i$ are n independent realisations of a r-dimension multivariate Gaussian random variable with mean zero and variance covariance matrix G, and $\boldsymbol{d}_{ij}$ are r-dimension vectors of explanatory variables for the random process $\boldsymbol{U}_i$. The $W_i(t_{ij})$ are n independent realisations of a stationary Gaussian process with mean zero, variance σ^2 and correlation function $\rho(u)$, with u being time lag. The processes $\boldsymbol{U}_i$ and $W_i(t_{ij})$ are in [2] terminology random effects and serial correlation components, interpreted as the variability between and within individuals, respectively.

Notice that decomposing ϵ_{ij} in the previous additive way implies that

$$\mathrm{Var}(\epsilon_{ij}) \;=\; D_i G D'_i + \sigma^2 H_i + \tau^2 I_i \;,$$

where H_i is a matrix with (j,k) element $h_{ijk} = \rho(|t_{ik} - t_{ij}|)$.

For estimation of model parameters we will use likelihood-based methods. The full likelihood is easily available for the entire data set.

2.3. Modelling missing process in longitudinal analysis

In this section we will be referring to balanced longitudinal study designs. This meaning that the study specifies that all subjects are observed at the same equally spaced time points, the same number of times. It is common that not all subjects provide the complete set of measurements for the study, originating missing values. Therefore, we review longitudinal models that cope with potentially informative missing data. In particular, we consider longitudinal models where the event is "drop-out of the study". This will lead us to distinguish different reasons for missing values, and how they can be associated with the repeated measurement processes.

Missing values in longitudinal studies occur in two different ways. They can be missing at intermittent times in the sequence, which means that other measurements are observed following missing values; for example, when a patient does not feel well for the visit, or just forgets the appointment. The other type of missing values appear when all other values after this are also missing, and the patient is said to have dropped-out of the study (measurement sequence terminates prematurely [3]). There might be several reasons for a patient to drop-out a study such as death, feeling the treatment is not helpful to them or just moving house.

The main concern of longitudinal analysis with missing data arises when there is an association between the longitudinal profile and the missing process. For example, if a patient drops-out the study because he/she believes that the treatment is not being effective, the missing values should not be dissociated from the measurement process. Therefore, it is necessary to distinguish between different reasons for missing values, to be possible to conjecture on possible association. Little and Rubin [4] classified the nature of missing data mechanism as:

> **MCAR** — Missing Completely At Random: when the probability of missing does not depend on either the observed or unobserved measurements. For example, when a patient forgets to attend the appointment.
>
> **MAR** — Missing At Random: when the probability of missing depends on the observed data, but not on the unobserved measurements. Conditional on the observed measurements, missing process and data are independent. For example, the patients leaves the study on doctors advice based on previous observed longitudinal measurements.
>
> **MNAR** — Missing Not At Random: when the probability of missing depends on observed and unobserved data. For example, when a patient leaves the study because he/she feels ill on the day of their appointment, and the illness is related with all the longitudinal profile, including those measurements that would have been observed if they would have kept on going to the appointments.

In a setting of time-to-event, it is reasonable to consider missing values as events, and the design times at which the missing values occur as the set of possible event times. The events associated with intermittent missingness are multiple events in a same subject. However, it is commonly assumed that this type of missing data is missing completely at random, because other measurements are observed after in time. Hence, intermittent missing values are treated as ignorable and inferences can me made using likelihood based methods.

The drop-out missing value originates a single event, identified as the time that terminates the longitudinal sequence. It is usual in clinical trials to record the cause of the patient's drop-out. This information helps to identify the nature of the missing data.

Let $\boldsymbol{Y}$ be the random variable associate with the complete data vector for a single subject, that can be decompose as $\boldsymbol{Y} = (\boldsymbol{Y}_{\text{obs}}, \boldsymbol{Y}_{\text{mis}})$ with observed and missing measurements, respectively. Also, $\boldsymbol{D}$ be the missing data indicator (0/1) for the same subject, for observed and missing measurements, respectively. The model for the complete data requires the specification of the joint distribution $[\boldsymbol{Y}, \boldsymbol{D}]$, where $[\cdot]$ represents the density distribution. Using this notation, Little [5] contrasts different models for the drop-out mechanism that come in parallel with

the nature of the missing data as before. These are:

Covariate-Dependent Drop-out — when the drop-out mechanism does not depend on any longitudinal values, but is allowed to depend on the covariates:

$$[\boldsymbol{D}|\boldsymbol{Y}] = [\boldsymbol{D}] \quad \equiv \quad \textbf{MCAR} \; .$$

Missing-at-Random Drop-out — when the drop-out mechanism depends only on observed data:

$$[\boldsymbol{D}|\boldsymbol{Y}] = [\boldsymbol{D}|\boldsymbol{Y}_{\text{obs}}] \quad \equiv \quad \textbf{MAR} \; .$$

Nonignorable Outcome-Based Drop-out — when the drop-out may depend on missing components of $\boldsymbol{Y}$:

$$[\boldsymbol{D}|\boldsymbol{Y}] = [\boldsymbol{D}|\boldsymbol{Y}_{\text{obs}}, \boldsymbol{Y}_{\text{mis}}] \quad \equiv \quad \textbf{MNAR} \; .$$

It is proved that likelihood-based inferences on the model parameters are unbiased when ignoring the missing values of the data [6], if the data is believed to be MCAR. The standard procedure for testing for MCAR is to compare the empirical distributions of complete observed variables for respondents and non-respondents subjects, using t-tests [7].

If the likelihood function can be factorised into two independent parts, one corresponding to the response parameters and the other corresponding to missing parameters, the missing process is considered to be at least MAR with respect to the response process. Under the MAR assumption Rubin [6] shows that if the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, on the distributions $f(\boldsymbol{y}|\boldsymbol{\theta})$ and $f(\boldsymbol{d}|\boldsymbol{y}, \boldsymbol{\psi})$, are distinct, then likelihood inference is possible, by integrating out the density of $\boldsymbol{y}_{\text{mis}}$. If the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ do not have common components it is possible to factorise

$$f(\boldsymbol{y}_{\text{obs}}, \boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\psi}) \; = \int f(\boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}}|\boldsymbol{\theta}) \, f(\boldsymbol{d}|\boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}}, \boldsymbol{\psi}) \; d\boldsymbol{y}_{\text{mis}} \; ,$$

and under the MAR assumption $f(\boldsymbol{d}|\boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}}, \boldsymbol{\psi}) = f(\boldsymbol{d}|\boldsymbol{y}_{\text{obs}}, \boldsymbol{\psi})$, so

$$f(\boldsymbol{y}_{\text{obs}}, \boldsymbol{d}|\boldsymbol{\theta}, \boldsymbol{\psi}) \; = \; f(\boldsymbol{d}|\boldsymbol{y}_{\text{obs}}, \boldsymbol{\psi}) \, f(\boldsymbol{y}_{\text{obs}}|\boldsymbol{\theta}) \; .$$

Therefore, maximisation of the likelihood for model parameters, requires the maximisation of two independent terms, that do not share common parameters. However a theorem proved by Molenberghs *et al.* [8] implies that MAR is untestable without additional assumptions no matter how much data are available. Also, Molenberghs and colleagues [9] derive the bias on parameter estimates when data is MCAR and MAR and simple methods like last observation carried forward and complete case analysis are used, and show that likelihood-base methods provide consistent estimators.

Missing values which are MCAR or MAR are known in the literature as ignorable, because longitudinal analysis can be performed ignoring them. However, missing values originated by MNAR are said to be informative or non-ignorable [10].

3. TIME-TO-EVENT DATA ANALYSIS

Time-to-event data is generated by observing several subjects until a single or multiple event occur, and the data is the waiting time. For example, in a medical context a single time-to-event is the time to recurrence of a health condition, time of response to a treatment or time to death from a certain cause. To determine time-to-event correctly, it is necessary to choose an appropriate time origin, which has to be easily identified and common for all patients. Usually, time from randomisation, time from diagnosis or time from beginning of medication is chosen. From now we will refer to failure time with the same meaning as time-to-event.

The special difficulty with time-to-event data, is that the event will not occur for some subjects during the follow-up period of the study. The only information available for these patients is a maximum time, up to which it is known not to have observed the event. For example, in a clinical trial where failure time is time to death, not all patients will die during the study. For these patients we observe a right-censored time, which in the maximum is the follow-up time of the study. The set of failure and censored times we call survival data, or sometimes observed lifetime. Therefore, the analysis of time-to-event data is also commonly called survival data analysis.

The observed censored times can represent subjects still alive when the study is finished, or subjects who drop-out of the study. We consider drop-out time, the time when a subject drops out of the study, and we use it analogously to time-to-event.

3.1. Notation

Let the random variable F denote the time-to-event and let $f_1, ..., f_n$ be a random sample from F on $i = 1, ..., n$ subjects. However, the event is not always observed and every subject i has associated a censored time, coming from a random variable C, where $c_1, ..., c_n$ is the random sample from C for the same subjects. Therefore, the observed survival data is the realisation $s_i = \min\{f_i, c_i\}$, $i = 1, ..., n$, of a random variable S.

A common assumption in survival analysis is non-informative censoring, meaning that random variables C and F are independent. Therefore, if $F \leq C$, $S = F$ and a failure time is observed, if $C < F$, $S = C$ and a censoring time is observed. The observed data are realisations (s_i, δ_i), where s_i is defined as before, and δ_i is a subject indicator (1/0), for failure or censored time, respectively.

To describe the distribution of failure time f, it is more appropriate to use the survival and hazard functions. The survival function $S(t)$ is defined as the probability of failure time being beyond some point t, $S(t) = \mathrm{P}(F > t)$. The hazard function is the probability of failure time occur in the next short period of time, given that failure time did not occur up to that time and all the past history,

$$\lambda(t) \;=\; \lim_{\Delta t \to 0^+} \frac{\mathrm{P}\big(t \leq F < t + \Delta t \,|\, F \geq t\big)}{\Delta t} \;,$$

and is defined as the instantaneous death rate for an individual surviving to time t. It is possible to combine the two definitions and get the relation

$$\lambda(t) \;=\; \frac{\mathtt{f}(t)}{S(t)} \;, \tag{3.1}$$

where $\mathtt{f}(t)$ is the density function of F.

For observed survival data (s_i, δ_i) on subjects $i = 1, ..., n$, the likelihood function of model parameter is the product of probabilities given the observed data, for all subjects i. Usually the censoring mechanism is ignored [11] and the likelihood of interest is

$$L(\theta; \boldsymbol{s}, \boldsymbol{\delta}) \;\propto\; \prod_i \mathtt{f}(s_i)^{\delta_i} \times S(s_i)^{1-\delta_i} \;, \tag{3.2}$$

where each failure time contributes with the density function and each censored time contributes with the survival function.

3.2. Survival models

When modelling survival data, the most common non-parametric method is the product-limit estimator [12], sometimes called the Kaplan–Meier estimator. Consider the ordered subset of $k \leq n$ unique observed failure times from the observed survival times, $s_{(1)} < ... < s_{(k)}$. Let d_i be the number of failures which occur at t_i and n_i the number of individuals who are at risk just before time t_i, making up the risk set $R(t_i)$, say. Notice that n_i represents the number of subjects that survive at least until time t_i. Therefore, d_i/n_i is an estimate of the probability of failure at time t_i, conditional on surviving up to t_i. The product-limit

estimator is then defined as

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < s_{(1)} \ , \\ \prod_{s_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) & \text{if } t \geq s_{(1)} \ . \end{cases} \tag{3.3}$$

The most common way to model survival data is through the hazard function, including a set of q explanatory variables $\boldsymbol{W}$ measured at baseline to predict failure time. The most widely use semi-parametric model is the so called Cox proportional hazards model [13],

$$\lambda(t|\boldsymbol{W}) = \lambda_0(t) \exp(\boldsymbol{W}'\boldsymbol{\alpha}) \ , \tag{3.4}$$

where $\boldsymbol{\alpha}$ is a $(q \times 1)$ vector of parameters to estimate, and $\lambda_0(t)$ is the unknown hazard function at the baseline variables $\boldsymbol{W} = \boldsymbol{0}$. This is a semi-parametric model, because the baseline covariates are modelled parametrically whereas the baseline hazard function is modelled non-parametrically with no specific form. The function $\lambda_0(t)$ is considered a nuisance parameter in the Cox proportional hazards analysis. Therefore, when writing the likelihood function for this model as in (3.2), it is not possible to estimate simultaneously the baseline hazard function and parameters of interest $\boldsymbol{\alpha}$. Consequently, Cox [14] suggests an estimation method based on conditional probabilities at the set of failure times, which is based on maximising the partial likelihood. The main advantage of the partial likelihood is that it does not depend on the baseline proportional hazard function $\lambda_0(\cdot)$, and only parameters of explanatory variables are estimated. For complete details on parameter estimates in the partial likelihood and score function vector see [15].

3.3. Time dependent covariates in time-to-event analysis

When the interest is on inference for the model parameters of a time-to-event process, we allow for survival data analysis, which deals with censored event times. The most popular model is the Cox proportional hazards model, where the hazard of an individual with some covariates is proportional to a baseline function of time [13], as discussed before. This model allows for fixed covariates that do not change over time [16], and parameters are estimated by maximising the partial-likelihood [14]. However, it is often the case that time-dependent covariates are available and these also want to be included in the survival model.

The Cox proportional hazards model can be extended to incorporate the observed time-dependent covariates [17], with the partial likelihood evaluated at

each event time in the form

$$\prod_{t_i:\ \text{event time}} \frac{\exp\{\boldsymbol{W}_i'(t_i)\,\boldsymbol{\alpha}\}}{\sum_{j\in R(t_i)}\exp\left\{\boldsymbol{W}_j'(t_i)\,\boldsymbol{\alpha}\right\}}\ , \tag{3.5}$$

where $R(t_i)$ represents the set of subjects at risk at event time t_i. This model is described in [16] Chapter 8, and the efficiency of the parameter estimates using partial likelihood is compared with those obtained from a fully parametric model. Moreover, Hougaard [11] in section 2.4.4 argues that time dependent covariates have to be predicted, which means the trajectory of the covariate has to be known at every time points.

There are also ways for handling with missing time dependent covariates, as in longitudinal models mention in the previous section. Lin and Ying [18] estimate from the subjects with complete measurements, the conditional expectation of missing covariates at all time points. Thus, in the parameter estimating equations this is subtracted to the observed covariate. They claim this method is generally more efficient than using only subjects with complete data. However, it is stated that the validity of the method *"depends critically on the MCAR assumption"*.

Paik and Tsai [19] suggest a very similar estimator, with the advantage that is consistent under the missingness mechanism. But also in this work, the authors conclude that when the missing probability depends on unobserved values of the covariate, their estimator is biased.

If we consider longitudinal measurements as time dependent covariates, it is ignored that these are measured with error, and the observe measurements are a noisy version of true process. A drawback of the previous methods is that they do not account for measurement error in the repeated measurements. Prentice [20] shows that regression coefficients on the partial likelihood are asymptotically biased when it accommodates covariates measured with error, and he suggests a modified partial likelihood, using conditional expectations on the relative observed hazard. Altman and DeStavola [21] review the different problems of including time dependent covariates measured with error in survival data analysis. Following this, models for the joint distribution of time-to-event and longitudinal response variables have been proposed, included in the so called area of joint analysis of longitudinal and time-to-event outcomes.

4. JOINT MODELLING

In the context of joint modelling it is necessary to establish a clear framework to distinguish terminology from longitudinal and time-to-event processes.

Two processes are considered, the longitudinal $\boldsymbol{Y}$ and time-to-event $\boldsymbol{F}$ processes with possible association, which we are interested in. Another common issue in any data set is the missing process generating missing data. Therefore, the missing data can be missing of longitudinal measurements or missing to observe the event. When the event is not observed the missing process is called censoring $\boldsymbol{C}$, and missing of longitudinal measurements is called missing data $\boldsymbol{D}$. The censoring process $\boldsymbol{C}$ is usually assumed non-informative, in the sense that is considered independent of the time-to-event and longitudinal processes.

The missing of longitudinal measurements can be intermittent or terminating the sequence of longitudinal measurements, as discussed in section 2. In the case of intermittent missingness we assume these MCAR and it is known that these can be ignored in the likelihood function. If a missing longitudinal measurement terminates the sequence of longitudinal measurements, we call it a drop-out time from a drop-out process $\boldsymbol{D}$, as corresponds to a subject drop-out of the study. Moreover, the drop-out process cannot be ignored in most of the cases and it is considered to be MNAR.

We then have four processes, longitudinal process $\boldsymbol{Y}$, drop-out process $\boldsymbol{D}$, time-to-event process $\boldsymbol{F}$ and censoring process $\boldsymbol{C}$, and assumptions on possible associations on these processes is necessary. For example, we might assume that the event of interest is drop-out time, and so processes $\boldsymbol{D}$ and $\boldsymbol{F}$ are the same. This is an assumption of many clinical trials, as there is no record of an actual time-to-event, and the time of the last observation is considered the failure time. Furthermore, the missing longitudinal measurements caused by drop-out time are allowed to be associated with the event time, that is MNAR.

Different associations are possible between the four processes. We will consider the situation that time-to-event is available in the data set and the event generates missing data. Therefore, the time-to-event process completely determines the drop-out process. Figure 1 represents graphically this situation.

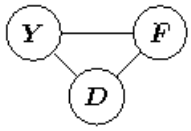

Figure 1: Graphical representation of possible associations between longitudinal, time-to-event and drop-out processes.

Another situation would be the longitudinal process associated with both processes time-to-event and drop-out, but these not associated with each other.

4.1. Model based joint models

We review full likelihood methods for exact estimation of model parameters in the joint distribution of repeated measurements and time-to-event. These are what we will call *joint models*, which model the joint distribution $[\boldsymbol{Y}, \boldsymbol{F}]$, for $\boldsymbol{Y}$ and $\boldsymbol{F}$ the random variables of repeated measurements and failure time, respectively. Inference on model parameters is done through the decomposition of the full likelihood. Nevertheless, it is not clear a joint distribution for the two random variables. Therefore, in joint models the joint distribution is factorised using Bayes rule.

The two different factorisations of the joint distribution generates different model strategies that contrast model interpretations, and consequently their suitability for individual problems. These are pattern-mixture and selection models [4, 22] that factorise:

pattern-mixture models	selection models
$[\boldsymbol{Y}, \boldsymbol{F}] = [\boldsymbol{F}]\,[\boldsymbol{Y}\vert\boldsymbol{F}]$	$[\boldsymbol{Y}, \boldsymbol{F}] = [\boldsymbol{Y}]\,[\boldsymbol{F}\vert\boldsymbol{Y}]$.

The parameters involved in each of the model components have different interpretations, in one model they are the parameters of the conditional distribution in the other they are the parameters of the marginal distribution. Depending on the context, the parameters of interest for inference will also be different. Notice that, if event is drop-out, $\boldsymbol{F} = \boldsymbol{D}$ in the terminology of section 2, and if the missing process is MCAR the two model strategies are equivalent, as the two processes are independent.

The model strategy depends mostly on the nature of the statistical problem and the scientific questions to be answered. Although mathematically the models describe exactly the same joint distribution, they have different statistical interpretations. Selection models are mainly used when inference is on time-to-event model parameters, improving the inference by allowing for correlation in the longitudinal measurements. In opposition, when primary interest is on the longitudinal trajectory, which might be associated with an event pattern, the pattern-mixture models are more commonly used. Therefore, the two different approaches lead to different understanding and inferences of the model parameters, together with different views on how to store the data.

Pattern-mixture models stratify regression models by missing pattern cohort, then model the marginal distribution of the response as a mixture of distributions over the patterns. These models are useful as an exploratory tool to check on longitudinal profile differences between drop-out groups. Selection models assume a model for the complete longitudinal data and then multiply by the

probability of observing the event given the complete data, though the observed data does not match the complete data.

Selection models can be seen as an alternative to pattern-mixture models for data with many complex missing patterns. The terminology of these models is clear for pattern-mixture models, they model a mixture of conditional distributions each for each missing pattern data. For selection models they model the selection of drop-outs condition on the measurement history.

The models above can be extended to incorporate random effects, in this case they are called random pattern-mixture models and random selection models. The individual unobserved random effects in the selection models are included in the marginal longitudinal model, whereas in the pattern-mixture models these come in the marginal distribution of the event times. Therefore, when jointly modelling repeated measurements $\boldsymbol{Y}$, event times $\boldsymbol{F}$ and random effects $\boldsymbol{U}$, the joint distributions are:

random pattern-mixture models	random selection models
$[\boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{U}] = [\boldsymbol{U}]\ [\boldsymbol{F}\vert\boldsymbol{U}]\ [\boldsymbol{Y}\vert\boldsymbol{F}]$	$[\boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{U}] = [\boldsymbol{U}]\ [\boldsymbol{Y}\vert\boldsymbol{U}]\ [\boldsymbol{F}\vert\boldsymbol{Y}]\ .$

Diggle [23] defines one different class of joint models, these as *random effects models*. Random effects joint models assume that both repeated measurements and event time depend on a unobserved random effect, these specified through a certain bivariate distribution. The random effects joint model is described by assuming conditional independence between $\boldsymbol{Y}$ and $\boldsymbol{F}$ given the random effects $\boldsymbol{U} = (U_1, U_2),$ as

random effects model

$$[\boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{U}] = [\boldsymbol{U}]\ [\boldsymbol{Y}\vert U_1]\ [\boldsymbol{F}\vert U_2]\ .$$

In random effects joint models the association between longitudinal measurements and time-to-event is completely determine by the correlation structure between the two random effects U_1 and U_2. The three different strategies to model the joint distribution, can be distinguish visually by diagrams presented by Diggle [23] and shown here in Figure 2.

The diagrams in Figure 2 represent conditional independence graphs for the three random variables. The absence of an edge indicates conditional independence between the two vertices of the edge, given the third vertice involved in the graph. In Figure 2(a) it is represented the saturated model, where all the associations are possible. Figure 2(b) represents selection models, where longitudinal measurements are influenced by their individual random effects, and it is the realisation of the measurement process that will influence the event, and not the random effect. On the contrary, in pattern-mixture models in Figure 2(c)

the individual random effects will determine the time of event, which after being predefined develops the individual longitudinal profile with some error. Regarding random effects joint models, Figure 2(d) suggests that both processes are a joint response to an unobserved individual specific process, and conditional on the responses being independent of each other.

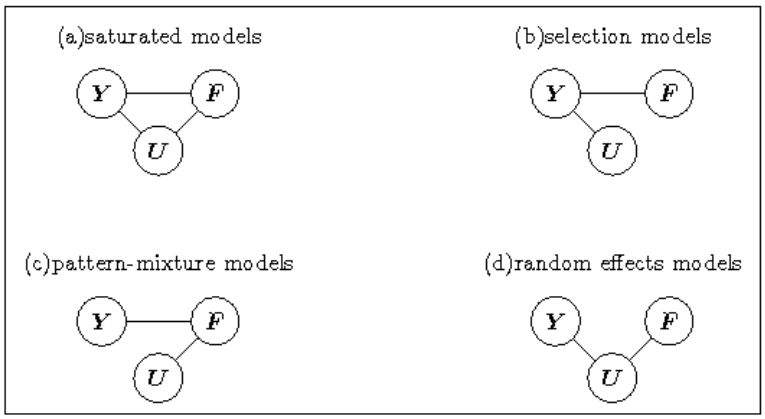

Figure 2: Graphical representation of saturated, selection, pattern-mixture and random effects models as in [23].

Little [5] produces a detailed review on selection and pattern-mixture models, when different missing mechanisms are present, in parallel with examples of data sets. Hogan and Laird [24] give a good comparison between pattern-mixture and selections models, and refers to random effect pattern-mixture and random effect selection models, but not with the same definition as we introduce here. We refer to this below, when giving examples of random effects models.

4.1.1. Pattern-mixture models

In mixture models it is necessary to specify a model for the marginal distribution of the event times $[\boldsymbol{F}]$ and the conditional distribution of $[\boldsymbol{Y}|\boldsymbol{F}]$. For the former, standard distributions would be the multinomial, or through modelling the hazard function with a Cox model, additive model or accelerated life model. The latter distribution is not always established, as the sample space of drop-out patterns can be integrated out of the conditional distribution, and inferences are made directly on the marginal distribution of $[\boldsymbol{Y}]$. The main goal of pattern-mixture models is to adjust the inference about $\boldsymbol{Y}$ for the effects of drop-out, with the convenience of not having to specify the event time marginal distribution.

One of the first pattern-mixture models was proposed by Wu and Bailey [25] whose aim was to compare rates of change of a continuous variable under informative missingness, for k different treatment groups. A conditional linear random effects model is proposed for the continuous variable, where the random effects $\boldsymbol{U}_i = (U_{1i}, U_{2i})$ are conditional on the event time. Especially the random slope is a polynomial function of some degree, on the event time,

$$(U_{2i}|\boldsymbol{F}_i = f_i) \;=\; \sum_{l=0}^{L} \gamma_{lk}\, s_i^l \;.$$

Two alternative estimation methods based on simple linear regression are proposed for the expected values of the random slope in the k group, namely linear minimum variance unbiased estimator and linear minimum mean squared errors estimator. These are compared by simulation studies with three other estimates and found to be more efficient. Testing for MNAR in this model, corresponds to test for the alternative hypothesis $H_A\colon \gamma_l \neq 0$, for all l, when the null hypothesis is $H_0\colon \gamma_l = 0$, for all l. Wu and Bailey [26] consider the particular case of $l = 1$.

Little [22] proposes a pattern-mixture model, where the drop-out patterns are considered realisations of a multinomial model, such that model parameters are the probabilities of having each of the drop-out patterns. For each drop-out pattern, the conditional distribution of the longitudinal measurements are assumed to be multivariate Gaussian. Usually, pattern-mixture models have a large number of parameters due to possible high number of drop-out patterns, which can cause identifiability problems. Therefore, it is mentioned the need to have at least one more longitudinal complete case than number of response variables to obtain consistent estimates for each pattern.

For a saturated model, where each multivariate distribution has distinct parameters, not all the parameters are possible to estimate. For example, the model parameters of missing patterns with no observed measurements, will not come out in the likelihood. The work of Little proposes parameter restrictions to the conditional models, which would reflect a certain missing process. The same approach is extended to categorical response variables, where the multinomial distribution is defined through a contingency table. However, it is noticed that the methodology can be inefficient because it requires a reasonable large number of complete cases.

The model proposed by Little does not allow for observed censored survival times, because it specifically models the probability of observing an event. Nevertheless, this can be extended to different multinomial distributions for each subsets of subjects with failure and censored times. This is a reasonable model when we want to assume independence between censoring and failure processes.

The work by Hogan and Laird [27] is motivated by limitations of the previous models that assume fully observed events, and parametric models for the drop-out process. Accordingly, the distribution $[\boldsymbol{F}]$ is specified non-parametrically by estimating multinomial probabilities with incomplete data, and the conditional distribution $[\boldsymbol{Y}|\boldsymbol{F}]$ is assumed a linear regression with individual random effects $\boldsymbol{b}_i$.

An advantage of using a Gaussian model for the conditional distribution is that the unconditional distribution is a mixture of conditional normal distributions. When the event times are completely observed the maximum likelihood estimates are easily obtained by maximising the likelihood.

The work of Li and Schluchter [28] examines different conditional models for the random effects $\boldsymbol{b}_i|\boldsymbol{F}_i$. Firstly, they consider a conditional quadratic and linear models, where random effects follow quadratic or linear regression curves on survival times. Secondly, a pattern-mixture model is described in the general form of the mixed effects model, using a single parameter for each missing pattern. These models differ on the design matrix $\boldsymbol{d}_i$, and many other models can be defined depending on the observed survival time allowing for censoring. For example, for a non-parametric model it is possible to use a piecewise linear or spline model.

In the opinion of Hogan and Laird [24] pattern-mixture models appeared to approximate selection models, as these are difficult to fit, have problems with identifiability and sensitivity to parametric assumptions. In addition, they considered pattern-mixture models not very appealing, because they mainly focus on the stratification of the sample by time of drop-out. However, they consider their main advantage over selection models, to be able to integrate out the cumulative distribution of $\boldsymbol{F}$. Therefore, it is possible to make inference on the marginal longitudinal parameters, without specifying a model for the drop-out process. When specifying a model for the drop-out, non-parametric estimators are usually use, like Kaplan–Meier as in [27].

All pattern-mixture models presented here make the assumption that cannot be verified of $f_{Y_{\text{obs}}|F} = f_{Y|F}$, which is not equivalent to $f_{Y_{\text{obs}}} = f_Y$ [27]. Other authors discuss more carefully about identification problems of pattern-mixture models, generated by unverifiable assumptions between the distribution of the complete data and only the observed measurements.

Thijs and co-authors [29] look at sensitivity analysis for pattern-mixture models, and propose three different strategies to fit pattern-mixture models under identified restrictions. The model strategies allow extrapolation beyond the time of drop-out, and inference on the distribution of the unobserved outcomes given the observed ones is possible. When restrictions are made it is plausible to perform a sensitivity analysis as the model assumptions are well identified. Birmingham *et al.* [30] present three class of restrictions that identify marginal distributions of the outcome, and are comparable to restrictions in selection models.

4.1.2. Selection models

In selection models the marginal distribution of longitudinal measurements is modelled, whereas the model for event time is conditional on the response variable. The most common approach for the distribution of repeated measurements is a linear mixed effects model, usually only random intercept and slope. Generally, choices for the conditional time-to-event distribution are logistic linear regression, probit regression, where probabilities are modelled as function of some longitudinal measurements. However, in some works the Cox proportional hazards model and accelerated life model are proposed.

One of the earliest proposals on selection models, as defined here, is [31] where the event time depends directly on the repeated measurements. The probability of drop-out at any time t_k is a parametric logistic linear model, with regression parameters on all the history of the observed measurements $(y_1, ..., y_{k-1})$ and on the unobserved measurement at drop-out time y_k, that is,

$$\text{logit}\big\{P(\boldsymbol{F} = k|\boldsymbol{y})\big\} \;=\; \beta_0 + \beta_1 y_k + \sum_{j=2}^{k} \beta_j \, y_{k+1-j} \;.$$

If $\beta_1 = \beta_2 = ... = \beta_k = 0$ the missing process is completely at random, and if only $\beta_1 = 0$ and all other different from 0 the missing values are missing at random.

This same model is used in [23], and likelihood ratio tests are performed on the β parameters to test for random drop-out and informative drop-out. A drawback of this model is the restriction on a monotone drop-out. This model does not deal with censoring times, but the model can be extended to accommodate for that.

Scharfstein and colleagues [32] refer to a general logistic regression model, in the context of a selection model, for the probability of drop-out given the complete vector of measurements,

$$\text{logit}\big\{P(\text{drop-out}|\boldsymbol{Y})\big\} \;=\; \beta_0 + q(\boldsymbol{Y}) \;,$$

where $q(\cdot)$ can be any function. For the particular application, they consider the class of functions $\mathcal{Q} = \{\alpha \log(\boldsymbol{Y}) \colon \alpha \in \mathbb{R}\}$, where α is a selection bias parameter. That is, it is possible to test for the value of α to be zero to understand the missing process, as before for the values of β.

In this paper the advantage of having a methodology that depends on a general function $q(\cdot)$ is discussed. The flexibility of function $q(\cdot)$ quantifies the influence of the response on the probability to drop-out, which allows a straightforward sensitivity analysis, where different assumptions can be tested.

Most of the joint models described here model repeated measurements data with the popular parametric linear mixed effects model. However, in many applications the data may not fit well by linear models, or it is of interest to model the response non-parametrically. Brown *et al.* [33] propose a cubic B-spline to model the longitudinal data, so that there is no parametric assumption on the trajectory of subject's longitudinal profile. This approach is relevant when inference on the effect of the longitudinal measures on the time-to-event is of interest, but not on the longitudinal process or its trajectory over time. Therefore, this approach allows a much more flexible modelling of the longitudinal data. Bayesian methods are used for the estimation, and the B-spline is extended to accommodates estimation on multiple response variables.

4.1.3. Random effects models

There are many models called selections models, that we include in our classification of random effects models. Although, the conditional distribution of time-to-event is modelled, this is conditional on a latent process, and the longitudinal and time-to-event processes are assumed independent conditional on the latent process. Therefore, we include these models in the class of random effects joint models. These models are also called shared parameter models, because the longitudinal response and missing mechanisms are modelled by sharing random effects.

In random effects joint models we assume both event time and longitudinal process dependent on a underlying disease or illness progression, defined by a random effect, rather than to the actual outcome. Moreover, the two processes are independent conditional on the unobserved random effects. For example, in [24] the joint distribution is defined as

$$[\boldsymbol{Y}, \boldsymbol{F}] = \int_{\boldsymbol{U}} [\boldsymbol{U}]\ [\boldsymbol{Y}|\boldsymbol{U}]\ [\boldsymbol{F}|\boldsymbol{U}]\ d\boldsymbol{U}\,.$$

Wu and Carroll [34] propose a random effects model called an "informative right censoring" model. In this model it is assumed that the repeated measurements follow a linear mixed effect model, with subject specific random effects, in particular random intercept and random slope. They further assume a general density function $M(t)$ for the failure process, conditional on the subject specific random effects. In particular they use a probit drop-out model for $M(t)$, and estimation of model parameters is obtained by maximising a pseudo-likelihood.

Testing for non-informative missingness under this model, is equivalent to test for the regression parameters that relate the conditional probability with the random effects to be zero. A test statistic is proposed for testing for non-informative missing process.

The model propose by Schluchter [35] is also a random effects model, as they use a trivariate Gaussian distribution to model the joint distribution of logarithm transformation of time-to-event, and the random intercept and random slope. These are random effects in a linear mixed effects model for the longitudinal measurements, and it is assumed that the event time is associated with an underlying process that is unobserved. The parameter estimates are acquired with an EM algorithm on the complete log-likelihood of the parameters given the observed data.

Some of the advantages of this model enumerated by the author are, the allowance for unbalanced data due to staggered entries or unequally-time visits, it is possible to use all the data available and possibility to apply likelihood ratio tests on the model parameters. In particular, on the correlation parameters of the trivariate distribution, which represent the association between random effects and event time. However, there is the computational disadvantage of this model, that may require large amounts of data to obtain convergence in the EM algorithm. We think this models simplifies the association structure, by only having two cross-correlation parameters, between the event time and two random effects, initial value and slope.

DeGruttola and Tu [36] propose to extend the two random effects joint model to include a general structure for the random effects. The conditional distribution of any transformation of failure times is modelled as a linear mixed effects model, and longitudinal and time-to-event processes share the random effects. Thus, this model assumes that both processes are measurements with error of the same unobserved latent process that represent health deterioration. The estimation of the model parameters is by an EM algorithm.

In all random effects models mentioned before, the survival time is modelled parametrically. Another very popular approach to the conditional distribution of the event times is by semiparametric survival models, such as the Cox proportional hazards model. Faucett and Thomas [37] propose one of the first random effects models with proportional hazards for the event time. They consider the joint analysis of longitudinal measurements and survival time as the joint distribution of two models, covariate tracking model and disease risk model. The former models the longitudinal response as a linear mixed model with subject specific random effects, intercept and slope, as in a linear growth curve model,

$$Y_{ij} \;=\; U_{1i} + U_{2i}\, t_{ij} + \epsilon_{ij} \;.$$

The latter allows a Cox proportional hazards model for the disease risk, with the same random effects as in the longitudinal model, assuming that these describe the true latent process

$$\lambda(t|\boldsymbol{U}_i) \;=\; \lambda_0(t)\, \exp\big\{\beta(U_{1i} + U_{2i}t)\big\}\,.$$

Wulfsohn and Tsiatis [38] consider the same model as Faucett and Thomas, as an alternative to the two-stage model. In this proposal, they notice that the normality of the random effects is on the overall subjects, and constant over time, which does not imply normality on the random effects of the subjects at risk at a certain time point.

In the two stage model the random effects are estimated in the first stage, by fitting a longitudinal model, and these are input to the Cox proportional hazards model in a second stage. In the model propose by Wulfsohn and Tsiatis [38] the parameters are estimated using all the information available at each time point, by maximising the full likelihood of the joint distribution. Although, the models by Faucett and Thomas and by Wulfsohn and Tsiatis are the same, they use different approaches for the parameter estimation. Faucett and Thomas follow a MCMC approach, with a Gibbs sampling whereas Wulfsohn and Tsiatis use an EM algorithm for the estimation.

Henderson and colleagues [39] propose an extension to the previous model, including a Gaussian stochastic process to each longitudinal response linear model

$$Y_{ij} \;=\; \mu(t_{ij}) + \Omega_{1i}(t_{ij}) + \epsilon_{ij}$$

and event time hazard model

$$\lambda(t|\boldsymbol{\Omega}) \;=\; \lambda_0(t)\,\exp\big\{\alpha(t) + \Omega_{2i}(t)\big\}\,.$$

The stochastic processes are components of a bivariate Gaussian process $\boldsymbol{\Omega}(t) = \{\Omega_1(t), \Omega_2(t)\}$. This is an extension of the previous model as each Gaussian stochastic process is assumed as in [40],

$$\Omega_{1i}(t) \;=\; \boldsymbol{d}_i \boldsymbol{U}_i + W_i(t)\;,$$

where $\boldsymbol{U}_i$ are the associated Gaussian random effects and $W_i(t)$ is a stationary Gaussian process that introduces serial autocorrelation. The last component is not considered in any of the previous models. It is then assumed that both processes are independent given $\boldsymbol{\Omega}(t)$. Therefore, the association between the two processes is interpreted by the correlation between the two latent variables. Moreover, in the absence of association between the two processes, the analysis becomes as two independent longitudinal and survival analyses.

Guo *et al.* [41] propose a model which they call a random pattern-mixture model, that incorporates aspects from both selection and mixture models. This model considers random subject-specific effects on the conditional longitudinal response, as in most of the cases, and a random pattern specific effects $\boldsymbol{V}$. The model implies the factorisation

$$[\boldsymbol{Y}, \boldsymbol{F}, \boldsymbol{U}] \;=\; \int_{\boldsymbol{V}} [\boldsymbol{V}]\;[\boldsymbol{F}|\boldsymbol{V}]\;[\boldsymbol{U}|\boldsymbol{V}]\;[\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V}]\;d\boldsymbol{V}\,.$$

5. DISCUSSION

The approach for the analysis of repeated measurements and time-to-event data, depends on research interests and on the assumptions we are ready to make on the available data. We have seen how these models can incorporate extra information. However, the assumption on association between processes need to be tested and we have seen that, for example MAR is not testable without additional assumptions. If the primary interest is on the time-to-event process, repeated measurements are used as time-dependent covariates in a Cox proportional hazards model. Conversely, if the interest is to make inference on the longitudinal profile, the missing pattern has to be considered.

We reviewed different methods for joint modelling of longitudinal and time-to-event data, based on the full likelihood of the joint distribution of the two processes. Different factorisations of the joint distribution lead to different model interpretations, namely pattern-mixture and selection models. We argue the approach of the analysis depends on scientific questions that need to be answered, and on the nature of association between processes. Cox [42] describes four different types of relation between a longitudinal process and failure times, not only in medical context and shows the implication of these on appropriate analysis in each case.

The model specification of selection models is more intuitive, usually with linear mixed effects model for the marginal distribution of the repeated measurements and a proportional hazards model for the conditional time-to-event distribution. However, these models usually involve intensive computational methods, as numerical integration and convergence difficulties.

The most common model for the longitudinal response variable, in pattern-mixture and selection models, is the general linear mixed effects model. Though, we notice the model proposals mainly differ in the random effects to use. Tsiatis and Davidian [43] give an interesting discussion on the philosophical issues of which of the fixed effects, random effects and stochastic processes should be included to model the longitudinal measurements. Their arguments are mainly related with biological processes that are involved in the specific data sets.

In particular we are not aware of pattern-mixture models that include a stochastic process in the conditional distribution of the longitudinal measurements. This could be related with model restrictions to include a stochastic processes on a conditional distribution which already has a time dependent process.

In this work, the focus is on the informative missingness of longitudinal measurements, due to an event. However, subjects do not always experience the event, and a censoring time is the only information available. The censoring

mechanism is always assumed non-informative and independent of time-to-event and longitudinal processes. In more complex models the censoring mechanism can be considered informative with an associated distribution, which would imply different models.

ACKNOWLEDGMENTS

This work has been supported by the grant number SFRH/BD/10266/2001 from FCT-Portugal, that supported the PhD of the author.

REFERENCES

[1] Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963–974.

[2] Diggle, P.J.; Heagerty, P.J.; Liang, K.-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data*, Oxford University Press, Oxford (second edition).

[3] Diggle, P.J. (1989). Testing for random dropouts in repeated measurement data, *Biometrics*, **45**, 1255–1258.

[4] Little, R.J. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, John Wiley, New York (second edition).

[5] Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies, *Journal of the American Statistical Association*, **90**(431), 1112–1121.

[6] Rubin, D.B. (1976). Inference and missing data, *Biometrika*, **63**(3), 581–592.

[7] Fuchs, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data, *Journal of the American Statistical Association*, **77**(378), 270–280.

[8] Molenberghs, G.; Michiels, B.; Kenward, M.G. and Diggle, P.J. (1998). Monotone missing data and pattern-mixture models, *Statistica Neerlandica*, **52**(2), 153–161.

[9] Molenberghs, G.; Thijs, H.; Jansen, I.; Beunckens, C.; Kenward, M.G.; Mallinckrodt, C. and Carroll, R.J. (2004). Analyzing incomplete longitudinal clinical trial data, *Biostatistics*, **5**(3), 445–464.

[10] Hogan, J.W. and Laird, N.M. (1996). Intention-to-treat Analyses for incomplete repeated measures data, *Biometrics*, **52**, 1002–1017.

[11] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, Statistics for Biology and Health, Springer, London.

[12] Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations, *Journal of the American Statistical Society*, **53**, 457–481.

[13] Cox, D.R. (1972). Regression models and life tables (with Discussion), *Journal of the Royal Statistical Society – series B Methodological*, **34**(2), 187–220.

[14] Cox, D.R. (1975). Partial likelihood, *Biometrika*, **72**(2), 269–276.

[15] Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis – Techniques for Censored and Truncated Data*, Springer, New York.

[16] Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*, Monographs on Statistics and Applied Probability, 21, Chapman and Hall, London.

[17] Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data – Extending the Cox Model*, Statistics for Biology and Health, Springer, London.

[18] Lin, D.Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements, *Journal of the American Statistical Association*, **88**(424), 1341–1349.

[19] Paik, M.C. and Tsai, W.-Y. (1997). On using the Cox proportional hazards model with missing covariates, *Biometrika*, **84**(3), 579–593.

[20] Prentice, R.L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model, *Biometrika*, **69**(2), 331–342.

[21] Altman, D.G. and DeStavola, B.L. (1994). Practical problems in fitting a proportional hazards models to data with updated measurements of the covariates, *Statistics in Medicine*, **13**, 301–341.

[22] Little, R.J.A. (1993). Pattern-Mixture models for multivariate incomplete data, *Journal of the American Statistical Association*, **88**(421), 125–134.

[23] Diggle, P.J. (1998). *Dealing with missing values in longitudinal studies*. In "Recent Advances in the Statistical Analysis of Medical Data" (B.S. Everitt and G. Dunn, Eds.), Arnold, London, 203–228.

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

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

[26] Wu, M.C. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring cause by death and withdrawal, *Statistics in Medicine*, **7**, 337–346.

[27] Hogan, J.W. and Laird, N.M. (1997). Mixture models for the joint distribution of the repeated measurements and event times, *Statistics in Medicine*, **16**, 239–257.

[28] Li, J. and Schluchter, M.D. (2004). Conditional mixed models adjusting for non-ignorable drop-out with administrative censoring in longitudinal studies, *Statistics in Medicine*, **23**, 3489–3503.

[29] Thijs, J.; Molenberghs, G.; Michiels, B.; Verbeke, G. and Curran, D. (2002). Strategies to fit pattern-mixture models, *Biostatistics*, **3**(2), 245–265.

[30] Birmingham, J.; Rotnitzky, A. and Fitzmaurice, G.M. (2003). Pattern-mixture and selection models for analysing longitudinal data with monotone missing patterns, *Journal of the Royal Statistical Society – series B Methodological*, **65**(1), 275–297.

[31] Diggle, P.J. and Kenward, M.G. (1994). Informative drop-out in longitudinal data analysis (with discussion), *Applied Statistics*, **43**(1), 49–93.

[32] Scharfstein, D.O.; Daniels, M. and Robins, J.M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes, *Biostatistics*, **4**(4), 495–512.

[33] Brown, E.R.; Ibrahim, J.G. and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival, *Biometrics*, **61**, 64–73.

[34] Wu, M.C. and Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process, *Biometrics*, **44**, 175–188.

[35] Schluchter, M.D. (1992). Methods for the analysis of informatively censored longitudinal data, *Statistics in Medicine*, **11**, 1861–1870.

[36] DeGruttola, V. and Tu, X.M. (1994). Modelling progression of CD4-Lymphocyte count and its relationship to survival time, *Biometrics*, **50**, 1003–1014.

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

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

[39] Henderson, R.; Diggle, P.J. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data, *Biostatistics*, **1**(4), 465–480.

[40] Diggle, P.J. (1988). An approach to the analysis of repeated measurements, *Biometrics*, **44**, 959–971.

[41] Guo, W.; Ratcliffe, S.J. and Have, T.T. (2004). A random pattern-mixture model for longitudinal data with dropouts, *Journal of the American Statistical Association*, **99**(468), 929–937.

[42] Cox, D.R. (1999). Some remarks on failure-times, surrogate markers, degradation, wear, and the quality of life, *Lifetime Data Analysis*, **5**, 307–314.

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