JOINT MODELLING OF LONGITUDINAL AND COMPETING RISKS DATA IN CLINICAL RESEARCH

Authors: Laetitia Teixeira
 – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto
CINTESIS - Centro de Investigação em Tecnologias e Serviços de Saúde,
EPIUnit, Instituto de Saúde Pública, Universidade do Porto, Portugal
lcteixeira@icbas.up.pt

Inês Sousa
 – Departamento de Matemática e Aplicações, Universidade do Minho
Centro de Biologia Molecular e Ambiental (CBMA), Portugal
isousa@math.uminho.pt

Anabela Rodrigues
 – Departamento de Nefrologia, Centro Hospital do Porto, Hospital Geral
de Santo António, Instituto de Ciências Biomédicas Abel Salazar,
Universidade do Porto, Portugal
rodrigues.anabela2016@gmail.com

Denisa Mendonça
 – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto
EPIUnit, Instituto de Saúde Pública, Universidade do Porto, Portugal
dvmendon@icbas.up.pt

Abstract:

• Joint modelling of longitudinal and survival data has received much attention in the recent years and is becoming increasingly used in clinical studies. When the longitudinal outcome and survival endpoints are associated, the many well-established models with different specifications proposed to analyse separately longitudinal and time-to-event outcomes are not suitable to analyse such data and a joint modelling approach is required. Although some joint models were adapted in order to allow for competing endpoints, this methodology has not been widely disseminated. The present study has as main objective to model jointly longitudinal and survival data in a competing risk context, discussing the different parameterisations of systematic implementations of these models in the R, using a real data set as an example for the comparison between the different model approaches. The relevance of this issue is associated with the need to draw attention of the users of this statistical software to the different interpretations of model parameters when fitting these models. To reinforce the relevance of these models in clinical research, we give an example of a data set on peritoneal dialysis that was analysed in this context, where death/transfer to haemodialysis was the event of interest and renal transplant was the competing event. Joint modelling results were also compared to separate analysis for these data.

Key-Words:

- *Competing risks; joint modelling; longitudinal data; peritoneal dialysis; time-to-event data.*

AMS Subject Classification:

- 62N, 62P10.

1.   INTRODUCTION

In many clinical research studies, it is relevant to simultaneously analyse information on a longitudinal repeatedly registered biomarker and on the time to a specific outcome event. Furthermore, more than one outcome event may occur. In these situations, when longitudinal and time-to-event outcomes are associated, a joint modelling approach taking competing risks into account is required to correctly analyse such data [23, 25].

Although, in cross-sectional clinical studies only one measure of each clinical parameter (often the baseline) is used to guide medical decisions, the use of additional information on repeated measures of clinical parameters allows a better understanding of the disease progression or treatment benefits [1]. In this type of longitudinal studies, the analysis of repeated measures of clinical parameters may be supplemented with information about the time at which an event of interest has occurred, that is, survival data (also designated by time-to-event data).

With the purpose of analysing separately longitudinal and survival data, methods such as linear mixed model [4] and Cox proportional hazard model [3], respectively, are well-established. However, when longitudinal measurements are correlated with time-to-event (i.e., in the presence of informative censoring - when the reason for censure is related to the study outcomes), when repeated measures are measured with error and/or when some missing values are present a joint modelling approach is required [23]. These aspects that realistically characterize observed data lead to biased inferences if naive separate methods are applied [5, 7, 9, 12, 20, 23].

Therefore, in joint modelling methodology several objectives may be formulated, according to the main focus of the analysis [5, 10]: (i) to analyse the time-to-event outcome, taking into account the effect of a longitudinal outcome as endogenous time-dependent covariate measured with error, (ii) to analyse the longitudinal outcome in the presence of informative (non-random) dropout time and (iii) to analyse effects of covariates of interest on both type of outcomes (longitudinal and time-to-event) simultaneously.

Despite joint modelling of longitudinal and survival data is becoming increasingly popular [2, 18, 24], joint modelling in competing risk framework has not been widely used in medical context. Given the complexity of the joint modelling approach in the presence of competing risks, several limitations can be enumerated, namely the small number of models implemented in statistical software and the restriction associated to the number of shared random effects to be integrated out in the likelihood function due to computational limitations [14].

Several authors have suggested extensions of joint models so that they could be applied in a competing risks problem, such as Elashoff et al. [6, 7], Williamson et al. [25], Li et al. [13] and Rizopoulos [19]. The approaches differ

according to parameterisation, joint likelihood function and estimation method considered. Up to now there are only two statistical packages, available in the CRAN repository, that implement systematically two different parameterisations, the JM package [19] and the joineR package [25].

A very recent review of several implementations of joint modelling was published [11], which summarize four published models, which have software available for model estimation. Each model features a different hazard function, latent association structure between the longitudinal and survival submodels, estimation approach and software implementation. The models described were applied to a trial of anti-epileptic drugs. However, in this work we further discuss the packages joineR and JM, namely the different interpretations of the model coefficients and the application in another clinical area.

Peritoneal dialysis is one of the main renal replacement therapy. The progression of end-stage renal disease patients included in a peritoneal dialysis program is monitored with regular control visits where several clinical parameters are recorded, as well as the time until the occurrence of relevant endpoints. Then, as in many other clinical research areas, in addition to the baseline characteristics, peritoneal dialysis patient data present two different types of outcomes: (i) longitudinal outcome, composed by clinical parameters measured at several time points (such as albumin), and (ii) time-to-event outcome, composed by the follow-up time until the occurrence of an event of interest. In the specific case of peritoneal dialysis patients, it is only possible to observe the first outcome event (and consequently the first time-to-event) from a set of possible competing events: death/transfer to haemodialysis and renal transplant. For this reason, we are in a competing risks framework [22].

As referred above, the focus of the present study is on the two approaches to joint model, which are the only ones implemented in common statistical software (R) for systematic use by any users, (1) JM package by Rizopoulos [19] and (2) joineR package by Williamson et al. [25]. In practice these two implementations of joint models correspond to different parameterisations with different parameter interpretations. With this work we emphasize that it is important to discuss at this stage the differences between the two joint models, since interpretation of model parameters are different, and confusing interpretations may occur. Notice that, using a real data set as an example, we want to analyse the differences of the results when using the two model approaches and make interpretations on the results. It is not our purpose to go further about the performance of the two approaches. Additionally, the implementation of these approaches allows us to illustrate the relevance of the joint modelling methodology in the evaluation of a peritoneal dialysis program.

The objective of this present study is threefold: i) to jointly model longitudinal and survival data in a competing risks framework; ii) to discuss different parameterisations of systematic implementations of these models in the available R statistical software; iii) to analyse data on peritoneal dialysis program under

a joint modelling approach with competing risks and compare results with those from separated longitudinal and survival analysis.

In the next section we review the theory of joint modelling, with focus on the competing risks approach. The third section presents the results of the analysis of the peritoneal dialysis dataset. Finally, a discussion and conclusion compose the last section.

2.   JOINT MODELLING IN THE PRESENCE OF COMPETING RISKS

The joint modelling approach takes into account the association between the survival and longitudinal process, determining simultaneously the parameter estimates for both processes [14]. Different models can be considered, differing on the decomposition of the joint likelihood of the longitudinal and survival processes and on the submodels formulation for each outcome. The models most commonly used are selection models, pattern-mixture models and random effects models, and each model providing different information [21]. The two parameterisations considered in this work are classified as random effects models, where the survival process is assumed to be associated with the longitudinal process through shared random effects. In the presence of competing risks, the survival submodel needs to take into account the presence of several possible endpoints. In order to model jointly a longitudinal and a time-to-event outcome in the presence of competing risks some approaches are presented below.

According to the focus of the analysis, different specifications of the joint model might be considered, which corresponds to different parameterisations of the model, taking us to different interpretations of the model parameters.

When the focus is on the survival process and the interest is to analyse the effect of a endogenous time-dependent covariate (for example a clinical parameter such as albumin measured along time) on the time until an event of interest (for example, death), the time-dependent cause-specific hazard regression model usually used in competing risk survival analysis is not appropriate. Results obtained from this model may be substantially biased since longitudinal measures are measured with error [5, 6]. In these situations, the fundamental idea is to construct a suitable model to describe the evolution in time for the longitudinal outcome, and then to use this estimated evolution as time-dependent covariate in the survival model, considering a jointly estimation [1].

Alternatively, when the focus is on the longitudinal process (for example, of some clinical parameter such as albumin), the joint modelling approach is required when missing observations of the longitudinal outcome may be related with the endpoint observed (i.e. in the presence of informative censoring). The use of a joint modelling approach reduces the bias in the estimates [14].

Additionally, if the focus is on both processes, the aim of the model is on inference regarding the strength of the link between the two processes [14, 16].

Let $y_i(t)$ be the observed value of a longitudinal response for the subject i at time point t, measured with error. Let T_i and C_i be the failure and non-informative censoring times and k the event observed of a set of K possible events ($k = 1, ..., K$). The event indicator is given by $\delta_i = \{I(T_i \leq C_i), k\}$, where $\delta_i = 0$ if non-informative censoring occurs.

2.1. JM package

The JM package that implements the parameterisation proposed by Rizopoulos [19] was adapted to a competing risks problem [19]. This approach considers a linear mixed effects submodel for the longitudinal outcome and a relative risk submodel for each possible competing event. This model allows to quantify the effect of a longitudinal covariate in the time-to-event outcome, particularly when the longitudinal covariate is measured with error [14].

Consider $m_i(t)$ the true and unobserved value of the longitudinal outcome $y_i(t)$ at time t. In order to measure the effect of an endogenous covariate on the risk for an event, $m_i(t)$ needs to be estimated. Furthermore, the complete history of the true unobserved longitudinal process up to time point t, $M_i(t) = \{m_i(s), 0 \leq s < t\}$, is successfully reconstructed using the available measurements $y_i = \{y_i(t), t = 1, ..., n_i\}$ of each subject (where n_i represents the number of longitudinal measurements for each subject i) and a set of modelling assumptions. A linear mixed effects model is considered to describe the subject-specific longitudinal evolutions and it is defined as:

(2.1) $y_i(t|x_{1i}, W_{1i}) = x_{1i}(t)^T \beta_1 + W_{1i}(t) + \varepsilon_i(t) = m_i(t) + \varepsilon_i(t)$

where β_1 denotes the vector of the unknown fixed effects parameters, $x_{1i}(t)$ denotes row vectors of the design matrix for the fixed effects and $\varepsilon_i(t)$ is the measurement error term with variance σ^2 ($\varepsilon_i(t) \sim N(0, \sigma^2)$). $W_{1i}(t)$ is the value at time t of an unobserved zero-mean Gaussian random process.

To quantify the effect of $m_i(t)$ on the risk for an event, λ_i, the authors proposed the use of a relative risk model:

(2.2) $\lambda_i(t|M_i(t), x_{2i}) = \lambda_0(t) \exp\{x_{2i}^T \beta_2' + \alpha m_i(t)\}$

where $\lambda_0(t)$ denotes the baseline risk function and x_2 is a vector of baseline covariates with a corresponding vector of regression coefficients β_2'. Parameter α quantifies the effect of the underlying longitudinal outcome on the risk for an event: $\exp(\alpha)$ denotes the relative increase in the risk for an event at time t that results from one unit increase in $m_i(t)$ at the same time point, adjusting for the remaining exploratory variables in the model.

In the presence of competing risks the notation for the survival submodel needs to be adapted. Then, for the event k, the standard relative risk model can be defined as:

$$(2.3) \qquad \lambda_{ki}(t|M_i(t), x_{2i}) = \lambda_{0k}(t)\exp\{x_{2i}^T\beta_{2k}' + (\alpha_1 + \alpha_k)m_i(t)\}$$

where $k = 1$ represents the event of interest and $k = 2, ..., K$ the competing events, $\lambda_{0k}(t)$ denotes the baseline risk function and x_2 is a vector of baseline covariates with a corresponding vector of regression coefficients β_2'. α_1 quantifies the effect of the underlying longitudinal outcome on the risk for the event of interest and $\alpha_2, ..., \alpha_K$ quantifies the additional effect of the underlying longitudinal outcome on the risk for the respective competing event. In this model, each of β_{2k}' is interpreted as the effect of each explanatory variable on the relative risk of event k after adjusting for the effect of the longitudinal response, which might also include the effect of the same explanatory variable. Then, the overall effect of a covariate on the hazard might be decomposed into the direct effect (survival submodel) and the indirect effect (longitudinal submodel) [11].

The estimation method proposed in this approach is the maximum likelihood considering a joint distribution of the observed outcomes $\{T_i, \delta_i, y_i\}$. This joint distribution is defined assuming that the vector of time-independent random effects W_{1i} underlies both the longitudinal and survival processes (the random effects account for both the association between the longitudinal and event outcomes, and the correlation between the repeated measurements in the longitudinal process). The likelihood function is given by

$$(2.4) \qquad p(T_i, \delta_i, y_i|W_{1i}; \theta) = p(T_i, \delta_i|W_{1i}; \theta)p(y_i|W_{1i}; \theta)$$

and

$$(2.5) \qquad p(y_i|W_{1i}; \theta) = \prod_{j=1}^{n_i} p\{y_i(t_{ij})|W_{1i}; \theta\}$$

where $\theta = (\theta_t^T, \theta_y^T, \theta_{W_1}^T)$ denotes the full parameter vector, with θ_t denoting the parameters for the event time outcome, θ_y the parameters for the longitudinal outcome and θ_{W_1} the parameters of the random-effects covariance matrix. Additionally, it is assumed that given the observed history, the censoring mechanism and the visiting process are independent of the true event times and future longitudinal measurements.

In the presence of competing risks, the likelihood part for the event process takes the form:

$$
\begin{aligned}
p(T_i, \delta_i|W_{1i}; \theta_t, \beta_1) = &\prod_{k=1}^{K}[\lambda_{0k}(T_i)\exp\{x_{2i}^T\beta_{2k}' + \alpha_k m_i(T_i)\}]^{I(\delta_i=k)} \\
&\times \exp(-\sum_{k=1}^{K}\int_0^{T_i}\lambda_{0k}(s)\exp\{x_{2i}^T\beta_{2k}' + \alpha_k m_i(s)\}ds)
\end{aligned}
$$

(2.6)

The Expectation-Maximization (EM) algorithm was used to maximized the log-likelihood function $l(\theta) = \sum_{i=1}^{n} \log p(T_i, \delta_i, y_i; \theta)$. The aim is to find the parameter values $\hat{\theta}$ that maximize the observed data log-likelihood $l(\theta)$, but by maximizing instead the expected value of the complete data log-likelihood (treating random effects as missing data). Additional information of this approach can be founded in Rizopoulos [19].

2.2. joineR package

Williamson et al. [25] proposed a competing risks random-effects joint model fitting a cause-specific hazard submodel (allowing for competing risks) with a separate latent association between longitudinal measurements and each event [25]. The idea behind this model is to analyse data arising from competing survival and longitudinal processes simultaneously exploiting dependencies between the components. Given that the main focus of this approach is the link between longitudinal and survival processes, the association between these two processes is represented through shared latent random effects. For example, for a shared latent random effect model, this association is achieved through the inclusion of the longitudinal random intercept and/or random slope terms into the survival process model [14].

A Gaussian linear model is assumed for longitudinal response $y(t)$ at time t (longitudinal submodel):

$$(2.7) \qquad y_i(t|x_{1i}, W_{1i}) = x_{1i}(t)^T \beta_1 + W_{1i}(t) + \varepsilon_i(t)$$

where β_1 denotes the vector of the unknown fixed effects parameters, $x_{1i}(t)$ denotes row vectors of the design matrix for the fixed effects, $W_{1i}(t)$ the value at time t of an unobserved zero-mean Gaussian random process and $\varepsilon_i(t)$ denotes zero-mean Gaussian measurement error with variance σ^2.

The difference between the JM and joineR approaches is in the survival submodel. Survival time is associated with the longitudinal response through a second zero-mean latent Gaussian process $W_{2i}(t)$, correlated with $W_{1i}(t)$. A semi-parameteric proportional hazards model is assumed conditioned to $W_{2i}(t)$, with hazard λ_i defined as:

$$(2.8) \qquad \lambda_i(t|x_{2i}, W_{2i}) = \lambda_0(t) \exp\{x_{2i}^T \beta_2 + W_{2i}(t)\}$$

where $\lambda_0(t)$ is an unspecified baseline hazard and x_2 is a vector of baseline covariates with a corresponding vector of regression coefficients β_2. The longitudinal and survival processes are assumed to be conditionally independent given W_1 and W_2, usually considered as a linear combination of Gaussian random effects [25]. If the two processes W_1 and W_2 were independent, we would be in the presence of two separate analyses (longitudinal and survival). Though, being W_1 and W_2 related with each other, their correlation will drive the association between longitudinal and survival processes.

This model was extended in order to include competing risks. In this case, a cause-specific hazard submodel with a separate latent association between longitudinal measurements and each possible event was considered. The longitudinal submodel remains the same type of model considered in the joint model without competing risks. On the other hand, one survival submodel for each competing risks is considered. Thus the survival submodel for cause k is defined as:

$$(2.9) \qquad \lambda_{ki}(t|x_{2i}, W_{2i}) = \lambda_{0k}(t) \exp\{x_{2i}^T \beta_{2k} + W_{2ki}(t)\}$$

where $\lambda_{0k}(t)$, $k = 1, 2, ..., K$, are unspecified baseline hazard functions, x_2 is a vector of baseline covariates and $W_{2k}(t)$, $k = 1, 2, ..., K$, are zero-mean latent Gaussian processes. In this case, it is assumed that $W_{2k}(t) = \gamma_k W_1(t)$, i.e., W_1 and W_2 are proportional. The parameter γ_k indicates the level of association between the two components, i.e, quantify the effect of the unobserved stochastic process W_1 on the risk for the event k. Longitudinal responses and competing risks survival times are assumed to be conditionally independent given W_1 and W_2. In this parameterisation of the joint model the coefficient β_{2k} corresponds to the total effect of each explanatory variable on the relative risk of event k, after adjusting for an unobserved Gaussian process that do not include fixed effects. This different interpretation can be contrasted with the one previously given to β'_{2k} in the JM package.

The likelihood function for observed data is factorized as the product of the marginal distribution of y and the conditional distributions of competing events $\eta \in (1, ..., K)$ given the observed values of y.

Considering θ the combined vector of unknown parameters and $L_y(y, \theta)$ the standard likelihood corresponding to the marginal multivariate normal distribution of y. Conditional on latent processes $W_{2k}(t)$, the competing risks are independent of themselves and of the measurements y. The likelihood function is given by:

$$(2.10) \qquad L(y, \theta, \eta) = L_y(y, \theta) \prod_{k=1}^{K} L_{\eta|y,k}$$

where

$$(2.11) \qquad L_{\eta|y,k} = E_{W_{2k}|y}\{L_{\eta|W_{2k}}(\theta, \eta = k|W_{2k})\}$$

in which the conditional likelihood for each competing event, $L_{\eta|W_{2k}}(\theta, \eta = k|W_{2k})$ captures any likelihood contribution arising from the number of longitudinal measurements observed before the kth competing event. In this model parameterisation, it is assumed that there is an unobserved process W_1 that drives both y and risk for event, λ_k. The effect of covariates in hazard is both direct and overall [11].

In order to maximize the likelihood of the observed data and estimate the parameters of interest, EM algorithm is used, similarly as the JM approach. More details of this approach can be founded in Williamson et al. [25], Diggle et al. [5], Henderson et al. [10] and in Philipson et al. [17].

3.   PERITONEAL DIALYSIS DATA

In order to compare the two model specifications presented above and discuss the interpretation of model parameters, the methods, JM and joineR, were used to analyse peritoneal dialysis data. For the joineR, the extension to accommodate competing risks was requested directly to the author since our analyse started before the formal library to perform this analysis became available at September 2017.

Along the permanence in peritoneal dialysis program, different types of information concerning the patients and their health condition are collected. Firstly, information about baseline characteristics of the patients such as sex and age is considered. During the follow-up, albumin is usually recorded in each control visit (usually one per month). Finally, the event that forced the patient to abandon the treatment program (death, transfer to haemodialysis and renal transplantation) and the respective follow-up time are also reported given their clinical relevance. Then, due to the diversity of information resulting of the motorization of these patients, efficient and powerful regression models, such as joint models for longitudinal and time-to-event outcomes are required to analyse such data.

The sample of this study comprises patients included in the peritoneal dialysis program of the Peritoneal Dialysis Unit, Nephrology Department, Hospital Geral de Santo António, Centro Hospitalar do Porto, Porto, Portugal. The sample is composed by 160 patients who started peritoneal dialysis therapy between October 1999 and February 2013. Sex and age were considered as baseline covariates. Serum albumin level is an important clinical parameter for end-stage renal patients and it is used to assess the health status of patients in dialysis [15]. Low albumin level is associated with kidney failure. The number of measures and the time between measures differed for each patient. Combined survival, characterized by the combined event death/transfer to haemodialysis, represents an important indicator for the evaluation of a peritoneal dialysis program. Then, in this application, this combined event was considered as the event of interest and renal transplantation as the competing risk event. Registry data collection and analysis was submitted to ethical appreciation and approved by the National Commission of Data Protection, which is the national supervisory authority for personal data control.

Females represented 51.9% (n=83) of the total sample (n=160), which has an overall mean age of 47.9 years (sd=14.4 years). Thirty patients (18.8%) had diabetes. The median of follow-up time was 27.4 months (IQR: 12.8-49.0 months). Considering the longitudinal outcome, the number of measures of albumin varied among patients, with a minimum of 1 observation and a maximum of 60 observations. The median of observations per patient was 13 (IQR: 6-23 observations). The mean score of albumin was 3.7 g/dL (sd=0.4 g/dL) for a total of 3129 observations. Considering the time-to-event outcome, 53 (33.1%) patients experienced

the event of interest (death or transfer to haemodialysis) and 41 (25.7%) the competing risk (renal transplant). Survival times were censored for 66 (41.2%) patients who were still active on the peritoneal dialysis program at the end of the study.

3.1. Exploratory analysis

A *spaghetti plot* showing the albumin individual progressions (grey lines) of the longitudinal response for the different competing events is presented in Figure 1. The black lines in Figure 1 represent a smooth spline of all observation points in the same plot.
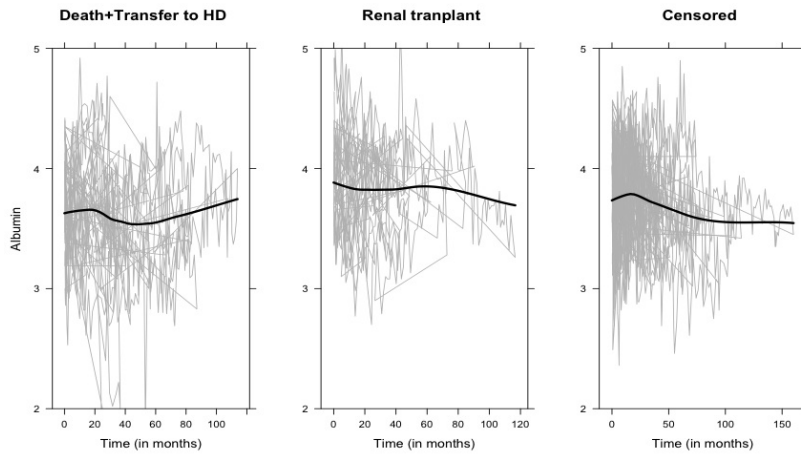

Figure 1: Smooth spline empirical mean of albumin evolution for the three subset of events: death/transfer to haemodialysis, renal transplant and censored.

Considering Figure 1, we verify that the mean of albumin score differs slightly according to the final event observed, showing a possible association between longitudinal albumin evolution and survival endpoint. Then, the analysis requires a joint modelling approach.

An estimate of the empirical variogram $\gamma(u)$ is presented in Figure 2. The diagram shows both the basic quantities (u_{ijk}, v_{ijk}), where $v_{ijk} = \frac{1}{2}(r_{ij} - r_{ik})^2$ is calculated from observed half-squared differences between pairs of residuals, of an ordinary least squares model (considering albumin as dependent variables and gender, age and time as independent variables), and $u_{ijk} = t_{ij} - t_{jk}$ the corresponding time-differences, and the kernel smooth estimate of $\gamma(u)$. To accentuate the shape of the smooth estimate, the vertical axis was truncated at 0.2. The variogram smoothly increases with lag corresponding to a decreasing correlation as

observations are separated in time. The horizontal line represents the variogram-based estimate of the process variance, which is substantially large than the value of the sample variogram, indicating that the positive correlation remains at arbitrarily large time separations. The empirical variogram of residuals, after fitting the data for an ordinary least squares model, allows us to understand the correlation structure of the longitudinal data. From Figure 2, we can see that the total variance in the data can be decomposed into three variance components, variance between and within subjects and measurement error. Therefore, a longitudinal approach shows to be adequate for these data.
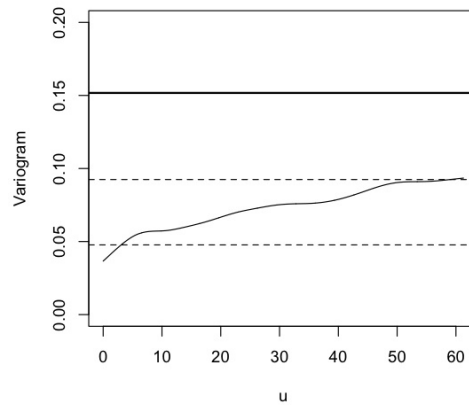

Figure 2: Empirical Variogram.

The cumulative incidence curves [8] give a global idea about the survival process. Figure 3 summarizes the cumulative incidence estimates for the two possible events taking competing risks into account (the time axis were halted at 60 months because the proportion of patients free of an event, but still in follow-up, becomes small). The probability of death/transfer to haemodialysis is always higher than the probability of renal transplantation. For example, the probabilities of death/transfer to haemodialysis by 1, 2 and 3 years after starting peritoneal dialysis were 0.08, 0.16 and 0.26 respectively and by the same time points the probabilities of renal transplantation were 0.05, 0.14 and 0.20.

3.2. Joint modelling

With the purpose of evaluating the relationship between longitudinal albumin scores and death/transfer to haemodialysis, in the presence of the competing risk renal transplantation, two joint model specifications implemented in the
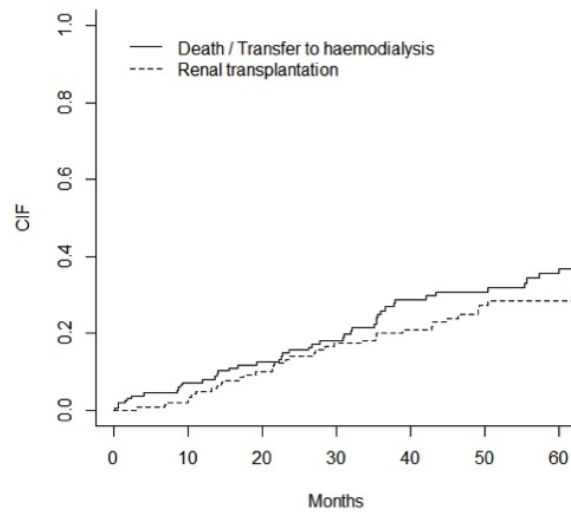
Figure 3: Cumulative incidence curves for death/transfer to haemodialysis
(solid line) and for renal transplantation (dotted line).

software R were analysed: the JM package proposed by Rizopoulos [19] and the
joineR package proposed by Williamson et al. [25], both adapted to a competing
risk situation. Furthermore, the parameters estimates and their standard errors
using the joint modelling specifications were compared to those obtained with
the independent models, a linear mixed model for the longitudinal outcome and
a time-dependent Cox model with competing risks for the survival outcome.

For the two joint model specifications discussed above, a linear mixed-
effects model was assumed for the longitudinal albumin outcome, with evolution
in time for each patient with different average effects per sex and age. For no-
tation simplification the individual index i and time index j were dropped. The
longitudinal submodel used was defined as (see equations (2.1) and (2.7)):

$$(3.1) \quad y(t) = m(t) + \varepsilon(t) = \beta_0 + \beta_{11} Sex + \beta_{12} Age + \beta_{13} time + b_0 + b_1 time + \varepsilon(t)$$

where y represents the albumin score and β_{11}, β_{12} and β_{13} represent the pa-
rameters of the fixed-effects part composed by the main effect of sex, age and
time, respectively. The unobserved zero-mean Gaussian random process $W_1(t)$
as in Equation (2.1) and (2.7) is, in this case, a linear combination of a ran-
dom intercept b_0 and a random slope b_1. That is, (b_0, b_1) has bivariate Gaussian
distribution with variances $\sigma^2(b_0)$ and $\sigma^2(b_1)$, respectively, and correlation ρ.

Notice that, from Figure 2 the empirical variogram indicates the need to
include a random effect at subject level (b_0, included) but also a possible Gaussian
stochastic process with a time correlation structure, as well as a random noise
($\varepsilon(t)$, included). However, we have fitted a model without a Gaussian stochastic
process because none of the systematic implementations, JM and joineR, allow
to include such a term in the model. This is due to the computational implemen-

tation involved to estimate parameters of such a model. In particular, we would have to integrate out a continuous Gaussian process in any time points observed.

In practice, this is one of the arguments of the importance of having systematic implementations of any statistical model to be fitted by any users. Though, it implies correct interpretation and use of these statistical models. Therefore, a random intercept effect and a random slope effect were included in the model for peritoneal dialysis data, because this is the possible way to incorporate a time-dependent correlation structure within patient, which was indicated by the variogram, in any of the model implementations.

For the event process, the two approaches presented in this study have different formulation.

For the JM joint model, two cause-specific relative risks models were assumed, one for each possible event (see equation (2.3)):

$$(3.2) \quad \begin{cases} \lambda_1(t) = \lambda_{01}(t) \exp\{\beta'_{211} Sex + \beta'_{212} Age + \alpha_1 m(t)\} \\ \\ \lambda_2(t) = \lambda_{02}(t) \exp\{\beta'_{221} Sex + \beta'_{222} Age + (\alpha_1 + \alpha_2) m(t)\} \end{cases}$$

The parameters β'_{211}, β'_{212} and α_1 denote the direct effects of sex, age, and albumin, respectively, on the risk for death/transfer to haemodialysis and the parameters β'_{221} and β'_{222} denote the effects of sex and age, respectively, on the risk for renal transplantation. The parameter α_2 corresponds to the additional effect of the albumin score on the renal transplantation.

Considering the joineR joint model, a semi-parameteric cause-specific hazard model for each event was assumed (see equation (2.9)):

$$(3.3) \quad \begin{cases} \lambda_1(t) = \lambda_{01}(t) \exp\{\beta_{211} Sex + \beta_{212} Age + \gamma_1 W_{21}(t)\} \\ \\ \lambda_2(t) = \lambda_{02}(t) \exp\{\beta_{221} Sex + \beta_{222} Age + \gamma_2 W_{22}(t)\} \end{cases}$$

The parameters β_{211}, β_{212} and γ_1 denote the effects of sex, age, and albumin in the underlying unobserved process W_1, respectively, on the risk for death/transfer to haemodialysis while the parameters β_{221}, β_{221} and γ_2 denote the effects of sex, age, and albumin in the underlying unobserved process W_1, respectively, on the risk for renal transplantation. This approach has as focus the link between the two longitudinal and survival processes. Therefore, the association between these processes is represented through shared latent random effects, achieved through the inclusion of the longitudinal random intercept (b_0) and random slope (b_1) terms into the survival process.

The parameters estimates and respective p-value using joint modelling approaches are presented in Table 1. For both approaches, standard error of the parameter estimates were obtained by refitting the models to 500 bootstrap samples generated using the original data. The bootstrap sampling was performed with replacement.

Among the joint models fitted, despite the formulation of the longitudinal submodel was the same, the association structures and method of estimation used can have different influences on the longitudinal submodel estimates [11]. In this application, similar results were obtained for both approaches, with a decrease in the albumin score along time.

Considering the results obtained for the survival submodel with the JM package [19], we verify an association between albumin and the risk of death/ transfer to haemodialysis ($\hat{\alpha}_1 = -1.24, p = 0.011$), meaning that a unit decrease in the marker corresponds to a $\exp(-(-1.24)) = 3.5$-fold increase in the risk for death/transfer to haemodialysis, controlling for the remaining factors in the model. No association between albumin and the risk of renal transplantation was found $(\hat{\alpha}_1 + \hat{\alpha}_2) = 0.54$ ($se = 0.47$), $p = 0.250$). Younger patients have a statistically significant higher hazard of getting a renal transplant (hazard ratio for one year decrease in age equals $\exp(-(-0.041)) = 1.04$ ($p < 0.001$). The direct effect of age in the hazard must be interpreted by also adjusted for the age-specific effect on albumin (longitudinal submodel). The log-likelihood from this joint model was -730.2515.

Results based on the joineR package [25] show a significantly $\hat{\gamma}_1$ estimate indicating that albumin score is positively associated with time to death/transfer to haemodialysis. However, no evidence of association between albumin and time to renal transplantation was found ($\hat{\gamma}_2 = 0.28$, $p = 0.625$). As expected, the estimates of the association parameters for the two competing events have opposite signs given that these two events have opposite reasons for discontinuation of therapy. Age (direct effect) was identified as statistically significant risk factor for renal transplantation (higher ages present lower hazard of renal transplantation), but not for death/transfer to haemodialysis. The log-likelihood from this joint model was -603.9029.

3.3. Separate analysis

Comparison of the parameters estimated and their standard errors from the joint model with the naive independent approach (independent linear mixed model and cause-specific hazard model), presented in Table 2, shows the differences of approaches. Results obtained for longitudinal outcome were similar. However, different results were obtained for time-to-event outcome. In separate analysis, sex was a significant factor for both events ($HR = 1.41$ ($p < 0.001$) for event death/transfer to haemodialysis and $HR = 1.42$ ($p < 0.001$) for event renal transplantation). Additionally, albumin (considered as time-dependent covariate) was a statistically significant factor for the event renal transplant ($HR = 1.57$, $p < 0.001$).

	JM package		joineR package	
	Coefficient (se)	p	Coefficient (se)	p
Longitudinal model				
Fixed effects				
Intercept	3.88 (0.11)	< 0.001	3.88 (0.12)	< 0.001
Sex (male)	0.24 (0.062)	< 0.001	0.24 (0.078)	0.002
Age	-0.0052 (0.002)	0.014	-0.0051 (0.002)	0.025
Time	-0.0015 (0.0018)	0.400	-0.0014 (0.0007)	0.069
Survival submodel				
Event of interest (D/TH)				
Sex (male)	0.41 (0.33)	0.209	0.12 (0.28)	0.649
Age	-0.012 (0.011)	0.278	-0.007 (0.010)	0.502
Association coefficient	-1.24 (0.49)	0.011	-1.41 (0.50)	0.005
Competing risk (RT)				
Sex (male)	0.51 (0.40)	0.204	0.62 (0.37)	0.091
Age	-0.041 (0.012)	< 0.001	-0.048 (0.013)	< 0.001
Association coefficient	0.54* (0.47)	0.250	0.28 (0.59)	0.625
$\hat{\sigma}(\varepsilon)$	0.0524		0.0524	
$\hat{\sigma}(\hat{b_0})$	0.147		0.159	
$\hat{\sigma}(\hat{b_1})$	0.000117		0.000123	
$\hat{\rho}$	-0.388		-0.346	
Log-likelihood	-730.2515		-603.9029	

Table 1: Parameter estimates for joint models fitted to albumin (longitudinal outcome) and time to peritoneal dialysis treatment failure (survival outcome) in the presence of competing risks. * indicates $\alpha_1 + \alpha_2$. joineR package: Williamson et al. [24]; JM package: Rizopoulos [19]. D/TH: Death/Transfer to haemodialysis; RT: Renal transplantation.

4.   DISCUSSION/CONCLUSION

It is very common to find clinical studies with both longitudinal measurements and event times. These measures are recorded on the participant of the study during follow-up time. Joint models are appropriate when interest lies in the association between a longitudinal covariate measured with error in a survival analysis or when accounting for event-dependent dropout in a longitudinal analysis. Several simulation studies have shown that joint model could be substantially more efficient than the separate analysis [6] because these models use

		Separate analysis	
		Coefficient (se)	p
Longitudinal model	Fixed effects		
	Intercept	3.88 (0.11)	< 0.001
	Sex (male)	0.24 (0.062)	< 0.001
	Age	-0.005 (0.002)	0.014
	Time	-0.0013 (0.0011)	0.270
Survival model	Event of interest (D/TH)		
	Sex (male)	0.34 (0.070)	< 0.001
	Age	0.022 (0.0025)	0.371
	Albumin	-0.58 (0.086)	< 0.001
	Competing risk (RT)		
	Sex (male)	0.35 (0.085)	< 0.001
	Age	-0.039 (0.0031)	< 0.001
	Albumin	0.45 (0.10)	< 0.001

Table 2: Parameter estimates longitudinal and survival model fitted separately, considering albumin as longitudinal outcome and time to peritoneal dialysis treatment failure as survival outcome in the presence of competing risks. Longitudinal model: linear mixed model; survival model: cause-specific. Cox proportional hazard model with time-dependent covariate. D/TH: Death/Transfer to haemodialysis; RT: Renal transplantation.

information from both outcomes. The literature about this theme is vast, and some review paper [14, 16, 21, 23] present and discuss different type of joint models focused on a single event with non-informative censoring. However, the majority of these models have only one event for the time-to-event outcome, excluding the possibility of observing competing risks. A very recent review paper described four approaches of joint models of longitudinal and survival data in the presence of competing risks, with application to an epilepsy drug randomized controlled trial. However, despite the recent methodological developments in the field of joint modelling of competing risks and longitudinal data, they remain still limited options for fitting these models in standard statistical software programs [11].

This work represents as far as we know the first study in the peritoneal dialysis area using joint modelling approach of longitudinal and survival data taking competing risks into account. The results obtained with this methodology produced new information about peritoneal dialysis program. Specifically, with this model it is possible to evaluate the association between the two processes, which cannot be obtained with standard survival models, contributing for a better knowledge of peritoneal dialysis program resulting in better management of the treatment program.

The development of different parameterisations with different perspectives and focuses allows to obtain different conclusions, and the choice of the model was related with the main clinical objective defined. Therefore, the three main objectives formulated in the context of joint modelling could be related with three main clinical research questions: (i) to evaluate the impact of albumin level on the combined survival, given that albumin level were recorded with measurement error; (ii) to analyse longitudinal evolution of albumin clinical parameter, given that lower levels of albumin may be associated with higher risk of mortality and morbidity (consequently less fit for renal transplant), i.e., in the presence of informative censoring; (iii) to evaluate the association between the progression of albumin level and combined survival and the identification of factors that influenced both outcomes.

In this paper, two parameterisations of a random shared effects joint model were compared considering an example in peritoneal dialysis. These two approaches are focused in distinct aspects. The parameterisation implemented in JM focuses mainly on the influence of a longitudinal variable measured with error in the estimation of the survival submodel. In this case, it is possible to quantify the effect of the longitudinal outcome in the survival hazard. On the other hand, the parameterisation implemented in joineR focuses mainly on the link between the processes, considering shared latent random effects to represent the correlation between longitudinal and survival process [14]. For this reason, the evaluation of the effect of an unobserved condition, shared between longitudinal and survival, in the hazard is possible using this parameterisation.

The two parameterisations presented provided complementary conclusions, given that they have different focus/objectives. The JM package was used to build a joint model when the focus is on a patient's survival and the inaccuracies in estimating albumin score. The joineR package was used to investigate the effect of a patient's changing albumin levels linking the longitudinal and survival processes through latent random effects. Although the two parameterisations present some differences relatively to the formulation, the modelling method of the baseline function and the survival submodel, the results had shown an evident relationship between the two processes in both approaches. This fact justifies the need for a joint modelling approach, and the advantages of the use of this methodology is highlighted when comparing results with separate analysis. Different conclusions were obtained considering separate analysis or a joint analysis, as shown in the previous section. Considering independent approaches, the focus is on the effect on parameters estimates and their standards errors ignoring the link between the longitudinal and survival processes and the longitudinal response measured with error within the survival process [14]. For separate analysis the effects of covariates that are significant, became not significant when a joint analysis approach is done. This might be due to variability that is being overestimated in a separate analysis, which is due to association between the two processes, longitudinal and survival. When this is taking into account this effect disappears.

In conclusion, joint modelling for longitudinal and time-to-event outcomes

in the presence of competing risks is useful in different areas of applications when the interest is the evaluation of the relationship between these two types of outcomes. In clinical studies diverse information about the patient is collected along a disease stages or treatment duration, and these models become an appropriate approach. Then it is necessary to alert clinicians for the implications and the advantages of a proper data collection and a correct data analysis.

ACKNOWLEDGMENTS

REFERENCES

[1]   ANDRINOPOULOU, E.; RIZOPOULOS, D.; JIN, R.; BOGERS, A.; LESAFFRE, E. and TAKKENBERG, J. (2012). An introduction to mixed models and joint modeling: Analysis of valve function over time, *Annals of Thoracic Surgery*, **93**, 6, 1765–1772.

[2]   BARRETT, J.; AND DIGGLE, P.; HENDERSON, R. and TAYLOR-ROBINSON, D. (2015). Joint modelling of repeated measurements and time-to-event outcomes: Flexible model specification and exact likelihood inference, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 1, 131–148.

[3]   COX, D.R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)*, **77**, 187–220.

[4]   DIGGLE, P.; HEAGERTY, P.; LIANG, K.-Y. and ZEGER, S.L. (2002). *Analysis of Longitudinal Data*, 2nd edition, Oxford University Press, Oxford.

[5]   DIGGLE, P.J.; SOUSA, I. and CHETWYND, A.G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: The fourth Armitage lecture, *Statistics in Medicine*, **27**, 16, 2981–2998.

[6]   ELASHOFF, R.M.; LI, G. and LI, N. (2007). An approach to joint analysis of longitudinal measurements and competing risks failure time data, *Statistics in Medicine*, **26**, 14, 2813–2835.

[7]   ELASHOFF, R.M.; LI, G. and LI, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types, *Biometrics*, **64**, 3, 762–771.

[8]   GRAY, R.J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk, *The Annals of Statistics*, **16**, 3, 1141–1154.

[9]   GUO, X. and CARLIN, B.P. (1988). Separate and joint modeling of longitudinal and event time data using standard computer packages, *The American Statistician*, **58**, 1, 16–24.

[10] HENDERSON, R.; DIGGLE, P. and DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data, *Biostatistics*, **1**, 4, 465–480.

[11] HICKEY, G.L.; PHILIPSON, P.; JORGENSEN, A. and KOLAMUNNAGE-DONA, R. (2018). A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **181**, 4, 1105–1123.

[12] HOGAN, J.W. and LAIRD, N.M. (2018). Model-based approaches to analysing incomplete longitudinal and failure time data, *Statistics in Medicine*, **16**, 3, 259–272.

[13] LI, N.; ELASHOFF, R.M. and LI, G. (2009). Robust joint modeling of longitudinal measurements and competing risks failure time data, *Biometrical Journal*, **51**, 1, 19–30.

[14] McCRINK, L.M.; MARSHALL, A.H. and CAIRNS, K.J. (2013). Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients, *International Statistical Review*, **81**, 2, 249–269.

[15] MEHROTRA, R.; DUONG, U.; JIWAKANON, S.; KOVESDY, C.P.; MORAN, J.; KOPPLE, J.D. and KALANTAR-ZADEH, K. (2011). Serum albumin as a predictor of mortality in peritoneal dialysis: Comparisons with hemodialysis, *American Journal of Kidney Diseases*, **58**, 3, 418–428.

[16] NEUHAUS, A.; AUGUSTIN, T.; HEUMANN, C. and DAUMER, D. (2009). A review on joint models in biometrical research, *Journal of Statistical Theory and Practice*, **3**, 4, 855–868.

[17] PHILIPSON, P.; DIGGLE, P.; SOUSA, I.; KOLAMUNNAGE-DONA, R.; WILLIAMSON, P. and HENDERSON, R. (2012). *joineR: Joint modelling of repeated measurements and time-to-event data*, R package version 1.0.1.

[18] PROUST-LIMA, C.; SÉNE, M.; TAYLOR, J.M.G. and JACQMIN-GADDA, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review, *Statistical Methods in Medical Research*, **23**, 1, 74–90.

[19] RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R*, CRC Press, Boca Raton.

[20] SAVILLE, B.; HERRING, A. and KOCH, G. (2010). A robust method for comparing two treatments in a confirmatory clinical trial via multivariate time-to-event methods that jointly incorporate information from longitudinal and time-to-event data, *Statistics in Medicine*, **29**, 1, 75–85.

[21] SOUSA, I. (2011). A review on joint modelling of longitudinal measurements and time-to-event, *REVSTAT - Statistical Journal*, **9**, 1, 57–81.

[22] TEIXEIRA, L.; RODRIGUES, A.; CARVALHO, M.J.; CABRITA, A. and MENDONÇA, D. (2013). Modelling competing risks in nephrology research: an example in peritoneal dialysis, *BMC Nephrology*, **14**, 1, 110.

[23] TSIATIS, A.A. and DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview, *Statistica Sinica*, **14**, 3, 809–834.

[24] WALDMANN, E.; TAYLOR-ROBINSON, D.; KLEIN, N.; KNEIB, T.; PRESSLER, T.; SCHMID, M. and MAYR, A. (2017). Boosting joint models for longitudinal and time-to-event data, *Biometrical Journal*, **59**, 6, 1104–1121.

[25] WILLIAMSON, P.R.; KOLAMUNNAGE-DONA, R.; PHILIPSON, P. and MARSON, A.G. (2008). Joint modelling of longitudinal and competing risks data, *Statistics in Medicine*, **27**, 30, 6426–6438.