Alternative subgroup joint analysis proposal of nonlinear longitudinal and time-to-event data for modeling pregnancy miscarriage

Rolando De la Cruz
Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile.
E–mail: rolando.delacruz@uai.cl
Marc Lavielle
Inria & CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, France.
E–mail: Marc.Lavielle@inria.fr
Cristian Meza
CIMFAV - Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile.
E–mail: cristian.meza@uv.cl
Vicente Núñez-Antón
Department of Quantitative Methods, University of the Basque Country UPV/EHU, Bilbao, Spain.
E–mail: vicente.nunezanton@ehu.eus

Abstract

Pregnancies achieved through *in-vitro* fertilization (IVF) are associated with adverse first trimester outcomes in comparison to spontaneously achieved pregnancies. Human chorionic gonadotrophin β subunit (β-HCG) is a well-known and accurate biomarker for the diagnosis and monitoring of pregnancy after IVF. Low levels of β-HCG during the first trimester of pregnancy are related to miscarriage, ectopic pregnancy and failure of IVF procedure. Longitudinal profiles of β-HCG can be utilized to distinguish between normal and abnormal pregnancies, and to assist and guide the clinician in better management and monitoring of post-IVF pregnancies. Therefore, being able to assess the association between longitudinally measured β-HCG and time to early miscarriage is of crucial interest to clinicians. A common joint modeling approach proposal to achieve this objective is to use subject-specific random effects in a mixed effects model for longitudinal β-HCG data as predictors in a model for the time-to-event (TTE) data. This work was motivated by an observational study with normal and abnormal pregnancies where serum concentrations of β-HCG were measured in 173 young women during a gestational age of 9-86 days at a private clinic in Santiago, Chile. Some women experienced a miscarriage event but the rest did not. For those women who experienced miscarriage, their exact event times were unknown, in such case we have interval censored data, assuming that the event occurred between the last time of the observed β-HCG measurement and ten days after. For our dataset we consider a nonlinear mixed effects (NLME) model for both normal and abnormal pregnancies, but the joint model is considered only for the subgroup of miscarriage women. All the estimation procedures are based on the Stochastic Approximation of the EM (SAEM) algorithm implemented in the Monolix software.

Key Words: Time-to-event data; Joint modeling; Longitudinal data; Nonlinear mixed models.

1 Introduction

There is an increasing interest in the analysis of time-to-event (TTE) data and, most importantly, in the joint modeling of TTE and longitudinal data. Joint models are a class of statistical models that allow the researcher to jointly model TTE data and longitudinal data. A mixed effects model is a statistical model containing both fixed and random effects. These models are useful in many areas as diverse as agriculture, biology, economics, geophysics, manufacturing, and medicine. They are particularly useful in settings where repeated measurements are made on the same statistical units (i.e., longitudinal data settings), or where measurements are made on clusters of related statistical units. In this specific case, observations in the same units/clusters usually cannot be considered to be independent and, thus, mixed effects models constitute a convenient tool for modeling units/clusters dependence. Mixed effects models are commonly used in longitudinal data analysis since they can cope with missing observations and unbalanced data and, in addition, take into account individual variations from a common pattern. A commonly encountered complication in the analysis of longitudinal data is the variable length of follow-up due to interval censoring, or any other type of censoring mechanism. This can be further exacerbated by the possible dependency between the TTE data and the longitudinal measurements. Our interest, therefore, lies in proposing a combination of a parametric model for the time-to-event data and a (nonlinear) mixed effects model for the longitudinal measurements, including an alternative joint modeling proposal for a subgroup of the available data. The dependency is handled via random effects which are naturally incorporated. Estimation procedures based on the Stochastic Aproximation of the EM (SAEM) algorithm are considered.

If we center on medical settings, we usually have a set of individuals where specific TTE data under study are available (e.g., injuries, cancer recurrences, or death). One may be interested in modeling the phenomenon inducing the event under study using a suitable chosen hazard function to be able to describe the instantaneous probability of the specific occurrence of the event. At the same time, for these same individuals, we may also have a longitudinal measurement, such as, for example, a biomarker, and try modeling its evolution along time. The relevance and importance of joint modeling lies in the possibility of studying the effect the longitudinal biomarker or measurement has on the TTE data or phenomenon under study. Initial proposals on joint modeling and its application to biological/medical settings were introduced by Self and Pawitan (1992) and DeGruttola and Tu (1994). More recent and more commonly used proposals, where the standard joint model was introduced, include Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997). In addition and as expected, developments in the area of joint modeling have continued and constitute a very much active research area in the field. For a complete introduction to joint modeling, we refer the reader to Rizopoulos (2012). As a brief description of this approach, joint modeling proposals try

to characterize the relationship between a longitudinal measurement's evolution and the risk of a given event under study, putting special emphasis on the proposal of a reasonable model for the longitudinal measurement evolution itself.

The evolution of the longitudinal measurement has been usually modeled with the use of linear mixed effects models (Laird and Ware, 1982; Harville, 1977; Verbeke and Molenberghs, 2000), splines (Ruppert, Wand and Carroll, 2003), B-splines with random effects (Rizopoulos, Verbeke and Molenberghs, 2009; Brown, Ibrahim and DeGruttola, 2005), or penalized splines with random effects (Durbán et al., 2005). These approaches take into account the correlation within measurements available on the same individual, a common feature in longitudinal data, and they also allow for between-individuals random variability with the use of random subject-specific slopes and/or intercepts in the longitudinal model. Thus, means for the specific parameters can be estimated, and the evolution of the longitudinal measurement for each individual in the study can be followed and/or modeled. Parameter estimation is usually performed by using maximum likelihood approaches, under normality and model's linearity assumptions that, in general, either do not hold or are not realistic in the specific application. Therefore, nonlinear mixed effects models are an alternative approach to overcome this situation (Lindstrom and Bates, 1990; Davidian and Giltinan, 1995; Vonesh and Chinchili, 1997), but these methods involve computationally intensive calculations when estimating by maximum likelihood methods. Recent approaches, such as the stochastic approximation to the expectation-maximization (SAEM) algorithm (Kuhn and Lavielle, 2004), implemented in the Monolix software, have led to much faster methods, are very flexible and, most importantly, well suited for modeling longitudinal data within the linear and nonlinear mixed effects models settings. With regard to the time-to-event data, approaches usually include specifying the survival and hazard functions. Hazard functions may be constant or time-varying. Joint modeling approaches focus on the possibility of studying the effect the longitudinal measurement has on the TTE data or phenomenon under study. That is, joint modeling is achieved by allowing the hazard function at time t to potentially depend on the value of the longitudinal measurement predicted at time t. Under a general mixed effects models setting, one or several random effects can enter the longitudinal and TTE models in many different ways. This paper proposes an alternative combination of a parametric model for the time-to-event data and a nonlinear mixed effects model for the longitudinal measurements for a subgroup of the dataset under study. The dependency is handled via random effects which are naturally incorporated. Estimation procedures based on the SAEM algorithm implemented in the Monolix software are considered.

The paper is organized as follows. In Section 2 we introduce the dataset that motivated this work, and in Section 3 we specify the alternative joint model formulation for the TTE and the longitudinal data for a subgroup of the motivating dataset. In Section 4 we present the results. Finally, we end with some conclusions in Section 5.

2 Assisted Reproductive Therapy in Chilean Women

It is well known in obstetrics that, among other clinical variables, the beta subunit of human chorionic gonadotrophin (β-HCG) is one of the clinical variables that shows dramatic changes in women during pregnancy. It has also been established that values of the β-HCG vary from women who have normal pregnancies with terminal deliveries to women who have spontaneous abortions or other types of adverse pregnancy outcomes (France et al., 1996). This association has made it possible to predict (with some uncertainty) the pregnancy outcomes. The dataset motivating our proposal corresponds to a follow-up study done in a private assisted reproduction center in Santiago, Chile, in which the β-HCG values and related hormones were measured during the first 90 days of gestational age for 173 pregnant women under the age of 30. The study was conducted for about two years and a total of 375 observations were recorded. The number of measurements per woman ranges between one and six, with a median of two. About 30% of the subjects had one β-HCG measurement, 31% had two, 33% had three, and 6% had four or more measurements. By the time of delivery, women were classified as *normal*, if they had a normal delivery, or as *abnormal* if they had any complication resulting in a non-terminal delivery and loss of the fetus.

Figure 1 shows the profiles for both groups against the time of pregnancy. Observe that the mean values of the $\log_{10}$ β-HCG show a nonlinear trend over the time of pregnancy. We notice the existence of a common growth pattern and that an exponential function (i.e., a nonlinear function) should be appropriate to represent the longitudinal behavior of the values of the β-HCG in a $\log_{10}$ scale. We can also observe the differences between the two groups. Such differences are more evident at lower concentrations of β-HCG, where the abnormal group shows higher values than the normal group (lower concentrations being almost four times more larger in the abnormal group). In addition, the abnormal group exhibits a larger variability in comparison with the normal group. These differences need to be accounted for by the model of choice, suggesting the use of a specific nonlinear model with woman-specific asymptotes in each group. In this specific dataset, we advocate for the use of TTE survival analyses, mainly because the event of interest under study, more specifically abnormal delivery of the child, occurs at any time in the period of study between 0 and 96 days, which should be considered in the joint model for the survival hazard rate for the i-th woman We would like to mention that our proposed approach is innovative in the sense that we advocate for the use of a joint longitudinal model for both groups and a TTE survival Weibull model only for the abnormal group. This is the main motivation to label our alternative joint methodological proposal as a subgroup joint modeling approach. That is, a joint modeling approach is used for the group of women having an *abnormal* delivery (i.e., those who had a complication resulting in a non-terminal delivery and loss of the fetus). In addition, it is important to mention that the subgroup of women having a *normal* delivery was included in the analysis. Moreover, as will be described in the proposal in Section 3 and the analysis in Section 4, the advantage of the proposed methodology is that it allows for the inclusion of several specific random effects in the nonlinear model and convergence occurs in

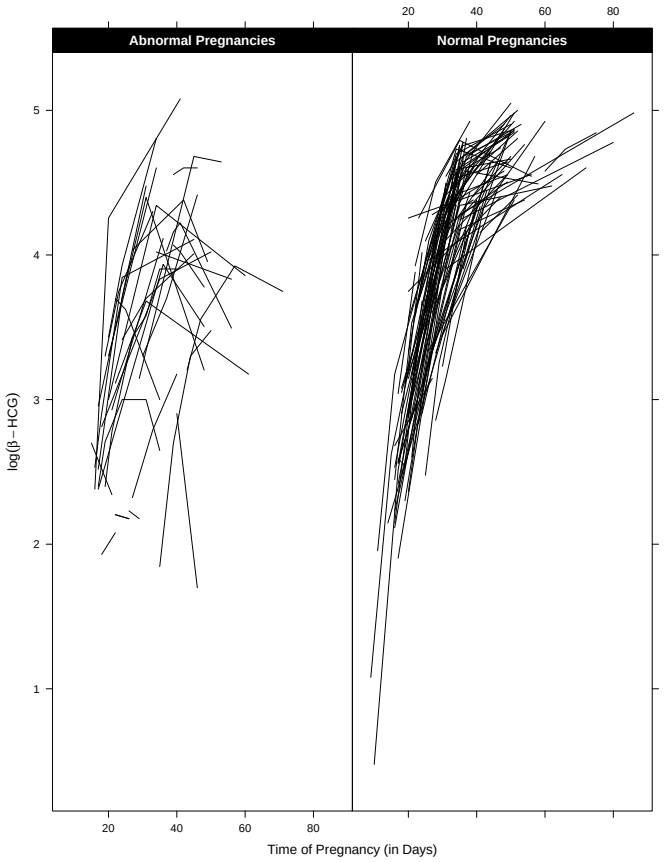


Figure 1: Time profiles for abnormal and normal pregnancies.

a natural way in the SAEM algorithm, whereas the use of this nonlinear model in a two-stage joint modeling approach (i.e. at the first stage, we summarize the longitudinal information with nonlinear mixed effects model and, at the second stage, we include the Empirical Bayes estimates of the subject-woman parameters as predictors in the TTE survival Weibull model, see Murawska, Rizopoulos and Lesaffre, 2012) is not possible with well known software packages, such as, for example, NLME in R, because covergence cannot be attained. The main reason for this proposal is that, in our case, clinicians were mainly interested in this subgroup because pregnancy did not come to an end within women belonging this subgroup, and they were interested in both modeling this behavior with a TTE survival model but, at the same time, be able to also follow the longitudinal evolution for the behavior of the β-HCG biomarker for both groups. That is, we only applied the TTE survival model to a subgroup of the original data, because clinicians were really interested in studying the time from pregnancy to fetus loss, which was a phenomenon only occurring in the abnormal group.

The analysis of the dataset is challenging, among other things, because the observations

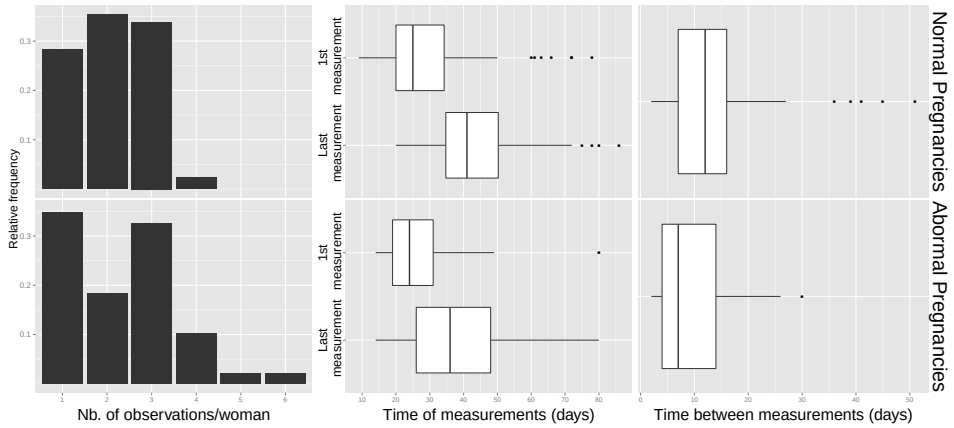


Figure 2: Characteristics of the measurements included in the pregnancy data set. The top row corresponds to the group of normal pregnancies and the bottom one to the group of abnormal pregnancies.

are unbalanced. Specifically, the number of observations per woman is very small and the measurement time grid is very irregular. In the group of normal pregnancies, 28% of the women only have one measurement and almost 98% of the women have three or fewer measurements, whereas these percentages are of 35% and 86%, respectively, in the group of abnormal pregnancies. In addition, the times at which the measurement were taken present a large variability among women. Moreover, the time between two consecutive measurements for the same woman exhibits a variability that goes from 2 up to 51 days. The number of women in the normal group, the number of women in the abnormal group and the total number of women in the study are, respectively, $N_0 = 124$, $N_1 = 49$ and $N = N_0 + N_1 = 173$. Furthermore, we let I_0 and I_1 be the set of indices of women in the normal and abnormal groups, respectively. See Figure 2 for a summary of these features.

Our joint modeling approach includes a nonlinear mixed effects model for the longitudinal measurements of the β-HCG values, which will provide the corresponding predicted values of the random effects included therein. These predicted values will be the input for the specific hazard rate for the i-th woman. In our specific settings, joint modeling characterizes the relationship between a longitudinal biomarker's evolution over time and the risk of a given event (i.e., the TTE under study), while also providing a reasonable model of the biomarker's evolution itself (i.e., the longitudinal model). That is, when both longitudinal measurement data (i.e., the β-HCG values) and time-to event data (i.e., the event of interest under study, normal or abnormal delivery of the child) are observed, a joint modeling proposal will be the best approach in terms of efficiently use of both types of data, specially in the case where predicted values from the former model enter in the specification of the latter, as is the case here and will be later described in detail in Section 3. In this work, we follow the idea proposed in Mbogning, Bleakley and Lavielle (2015), but we handle the dependency in both parts (i.e., the longitudinal and the TTE data) via random effects, which

are naturally incorporated. That is, only some random individual effects are included in the survival model. We then consider a nonlinear mixed effects model to model the longitudinal data, and a parametric model to explain the TTE data, where both parts share a common parameter. In the case of the TTE data, the recorded observations are the times at which events occur, where we will consider that the event can be interval censored. Our proposal fits into the so-called "shared parameter" approach in the sense that predicted values from the random effects in the nonlinear mixed effects model are included in the hazard ratio specification for the time-to-event model specification. We generalize the approach in Mbogning, Bleakley and Lavielle (2015), and propose an estimation procedure using the SAEM algorithm. That is, we propose a shared parameter model to both the time-to-event data and continuous longitudinal data. A nonlinear mixed effects model for the longitudinal data and a parametric survival model with the shared predicted values from the random effects in the longitudinal data model are joined to predict the probability of the event under study. Our joint model proposal centers on the time-to-event outcome of interest, taking advantage of the longitudinal data measurements, which basically means that the probability of the event under study will be more precisely predicted by incorporating the longitudinal β-HCG value measurements.

3 The Joint Model Formulation for the Pregnant Women Dataset

As pointed by Mbogning, Bleakley and Lavielle (2015), we can use joint models as a class of statistical methods for modeling together longitudinal data and time-to-event (TTE) data into a unified approach. In biometrics setting, we often have, for a set of patients, time-to-event data of interest, for instance the loss of the fetus during a pregnancy. One may be interested in modeling the process inducing the event, using for example a suitable chosen (time-dependent or not) hazard function to describe the instantaneous chance of an event occurrence. Simultaneously, for each patient we may be able to measure a longitudinal outcome and model its progression. It is common that a given longitudinal biomarker has a real influence on the TTE process, which is the context within which this paper centers.

Liu and Ying (2007) introduced a combination of a semiparametric transformation model for the TTE data and a linear mixed effects model for the longitudinal measurements. Mbogning, Bleakley and Lavielle (2015) proposed a nonlinear mixed-effects framework to jointly model longitudinal and repeated time-to-event data using a parametric mixed-effects hazard model for repeated event times. The link between both types of data (i.e., the longitudinal and the TTE data) is the conditional expectation of the longitudinal observation given the random effects or, more simply, a function of the predicted longitudinal biomarker.

In this work, we follow the idea proposed in Mbogning, Bleakley and Lavielle (2015), but we handle the dependency in both parts (i.e., the longitudinal and the TTE data) via random effects, which are naturally incorporated. That is, only some random individual effects are included in the survival model. We then consider a nonlinear mixed effects model to model the longitudinal data, and a parametric model to explain the TTE data, where both parts share

a common parameter. In the case of the TTE data, the recorded observations are the times at which events occur. Here, we will consider that the event is interval censored, i.e. we know the event has happened during a time interval, but we do not know the exact time at which it occurred. For instance we have $(L_i < T_i < U_i)$, where T_i is the time of occurrence of the event of interest for the i-th woman of the abnormal group (i.e. $i \in I_1$).

Suppose that the responses of interest are repeatedly measured for each of the N observations over a period of time. For the i-th observation, $i = 1, \ldots, N$, observation times are confined to a woman-specific time interval $[0, T_i]$.

3.1 Model for the longitudinal data

Let y_{ij} be the measured concentration of the β–HCG hormone for the i-th woman at time t_{ij}. We consider the longitudinal data arising from a nonlinear mixed effects model, which assumes a nonlinear regression model with woman-specific random effects. More specifically, the model can be written as:

$$y_{ij} = m(t_{ij}, \phi_i) + \varepsilon_{ij}, \quad i = 1, \cdots, N, \; j = 1, \ldots, n_i, \tag{1}$$

where m is a real-valued function of time that depends on a subject-specific parameter ϕ_i and where the residual error terms are independent and normally distributed such that $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. For a given woman i, $m(t, \phi_i)$ is then the predicted concentration of β–HCG at time t. To model the hormone concentration as a function of time, we use the following three-parameter logistic model:

$$m(t, \phi_i) = \frac{a_i}{1 + \exp\left\{-\frac{(t-b_i)}{c_i}\right\}}, \quad i = 1, \cdots, N, \; j = 1, \ldots, n_i, \tag{2}$$

where $\phi_i = (a_i, b_i, c_i)'$ is a vector of individual parameters.

3.2 Model for the time-to-event data

Here, the event of interest is the loss of the fetus in the abnormal group, which occurred within 10 days after the last measurement of the hormone. Two functions have a key role in TTE analysis: the survival function $S(t)$ and the hazard function $h(t)$, where $S(t)$ is the probability that the event happens after time t and $h(t)$ is the instantaneous rate of an event, given that it has not already occurred. Both functions are related by the following equation:

$$S(t) = P(T \geq t) = e^{-\int_0^t h(u)d(u)}$$

A common way to estimate $S(t)$ non-parametrically is to calculate its Kaplan-Meier estimate. Here, we are working under a population approach, so these functions, $S(t)$ and $h(t)$, are, thus, individual-specific functions, i.e., each subject has its own. Depending on the goal of the time-to-event analysis, different modeling approaches can be used: non-parametric,

semi-parametric (Cox models) and parametric. In this work, we will consider individual parametric functions S_i and h_i for the TTE analysis for women in the abnormal group (i.e. $i \in I_1$). In other words, these functions depend on the individual parameters ϕ_i, so that we have: $S_i(t) = S(t, \phi_i) = P(T_i > t; \phi_i)$ and $h_i(t) = h(t, \phi_i)$ for $i \in I_1$.

To describe the various shapes that the survival function can take, several hazard functions have been proposed. In this work, we will use the Weibull model for which the hazard is defined by

$$h_0(t) = \gamma \alpha t^{\alpha-1} \tag{3}$$

Two extensions of this baseline model will be considered in order to take into account the link between the β–HCG concentration and the time of the loss:

$$h_i^1(t\,, \phi_i) = h_0(t) e^{\beta_1 a_i} \tag{4}$$
$$h_i^2(t\,, \phi_i) = h_0(t) e^{\beta_2 m(t, \phi_i)} \tag{5}$$

In the first model, only the individual limiting concentration a_i has an impact on the individual hazard, while the second model assumes that the hazard at time t depends on the instantaneous concentration $m(t, \phi_i)$. Coefficients β_1 and β_2 measure the strength of the association between the different characteristics of the underlying woman-specific nonlinear evolution of the longitudinal profiles and the risk of losing the fetus.

Note that the hazard rates (4) and (5) are specified only for the abnormal pregnancy group, that is, the joint model is formulated for a subgroup of the available dataset ($i \in I_1$).

3.3 Model for the individual parameters

Let $(z_i, 1 \leq i \leq N)$ be a sequence of label variables such that $z_i = 0$ if $i \in I_0$, i.e. if the i-th woman belongs to the normal group, and $z_i = 1$ if $i \in I_1$. We will then assume that the ϕ_i's are independent and normally distributed such that

$$\phi_i \sim \mathcal{N}\left(\mu_i\,,\ \Omega\right), \tag{6}$$

where

$$\mu_i = \begin{pmatrix} a_{\text{pop}} + \delta\, z_i \\ b_{\text{pop}} \\ c_{\text{pop}} \end{pmatrix} \quad ; \quad \Omega = \begin{pmatrix} \omega_a^2 & 0 & 0 \\ 0 & \omega_b^2 & 0 \\ 0 & 0 & \omega_c^2 \end{pmatrix}$$

According to the structural model (2), this model for the individual parameters assumes that the asymptotic β-HCG level, a_i, has mean a_{pop} for the normal pregnancy group and $a_{\text{pop}} + \delta$ for the abnormal pregnancy group. Then, testing if the profiles of β–HCG concentration are the same in the two groups reduces to testing if $\delta = 0$.

3.4 Parameter estimation and model selection

In this study, we compare several models based on two different covariate models for the individual parameter a_i ($\delta = 0$ or $\delta \neq 0$) and two different hazard functions defined in equations (4) and (5). The four models to compare are summarized in Table 1.

Table 1: Four joint models to be compared.

Model	covariate model	hazard	θ
$\mathcal{M}_1$	$\delta = 0$	h^1	$(\gamma, \alpha, \beta_1, a_{\text{pop}}, b_{\text{pop}}, c_{\text{pop}}, \omega_a^2, \omega_b^2, \omega_c^2, \sigma^2)$
$\mathcal{M}_2$	$\delta \neq 0$	h^1	$(\gamma, \alpha, \beta_1, a_{\text{pop}}, b_{\text{pop}}, c_{\text{pop}}, \delta, \omega_a^2, \omega_b^2, \omega_c^2, \sigma^2)$
$\mathcal{M}_3$	$\delta = 0$	h^2	$(\gamma, \alpha, \beta_2, a_{\text{pop}}, b_{\text{pop}}, c_{\text{pop}}, \omega_a^2, \omega_b^2, \omega_c^2, \sigma^2)$
$\mathcal{M}_4$	$\delta \neq 0$	h^2	$(\gamma, \alpha, \beta_2, a_{\text{pop}}, b_{\text{pop}}, c_{\text{pop}}, \delta, \omega_a^2, \omega_b^2, \omega_c^2, \sigma^2)$

We propose to fit this joint model to all longitudinal data (normal and abnormal groups), using the hazard rate previously defined in (4) or (5) only for the abnormal group.

For each of these four models, the vector of parameters θ is estimated by maximizing the joint observed likelihood

$$\ell(\theta; y, T) = \prod_{i \in I_0} \ell_i(\theta; y_i) \prod_{i \in I_1} \ell_i(\theta; y_i, T_i),$$

where the contribution of the i-th woman in the normal pregnancy group to the likelihood is given by:

$$\begin{aligned} \ell_i(\theta; y_i) &= p(y_i; \theta) \quad , \quad i \in I_0 \\ &= \int_{\Re^3} p(y_i|\phi_i; \theta) p(\phi_i; \theta) d\phi_i, \end{aligned} \tag{7}$$

and the contribution of the i-th woman in the abnormal pregnancy group to the log-likelihood is given by:

$$\begin{aligned} \ell_i(\theta; y_i, T_i) &= p(y_i, T_i; \theta) \quad , \quad i \in I_1 \\ &= \int_{\Re^3} p(y_i|\phi_i; \theta) P(L_i < T_i < U_i|\phi_i; \theta) p(\phi_i; \theta) d\phi_i \end{aligned} \tag{8}$$

Here, $p(y_i|\phi_i; \theta)$ is the probability density function of the longitudinal observations conditionally on the individual parameter ϕ_i. That is,

$$p(y_i|\phi_i; \theta) = (2\pi)^{-\frac{n_i}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n_i} \left[y_{ij} - m(t_{ij}, \phi_i) \right]^2 \right\},$$

and $P(L_i < T_i < U_i|\phi_i; \theta)$ is the conditional distribution of the time-to-event T_i,

$$P(L_i < T_i < U_i|\phi_i; \theta) = P(T_i < U_i|\phi_i; \theta) - P(T_i < L_i|\phi_i; \theta) \tag{9}$$

$$= e^{-\int_0^{L_i} h(t,\phi_i) dt} - e^{-\int_0^{U_i} h(t,\phi_i) dt}, \tag{10}$$

where $h = h^1$ is defined in (4) under models $\mathcal{M}_1$ and $\mathcal{M}_2$ and $h = h^2$ is defined in (5) under models $\mathcal{M}_3$ and $\mathcal{M}_4$. In our dataset, some women experienced a miscarriage event but the rest did not. For those women who experienced miscarriage, their exact event times were unknown, in such case we have interval censored data, assuming that the event occurred between the last time of the observed β-HCG measurement and ten days after.

Finally, $p(\phi_i;\theta)$ is the probability density function of the individual parameters ϕ_i

$$p(\phi_i;\theta) = (2\pi)^{-\frac{3}{2}}|\Omega|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\phi_i-\mu_i)'\Omega^{-1}(\phi_i-\mu_i)\right\},$$

where $\mu_i = (a_{\text{pop}}, b_{\text{pop}}, c_{\text{pop}})'$ under models $\mathcal{M}_1$ and $\mathcal{M}_3$ and $\mu_i = (a_{\text{pop}}+\delta, b_{\text{pop}}, c_{\text{pop}})'$ under models $\mathcal{M}_2$ and $\mathcal{M}_4$.

For each of the four models in Table 1, the likelihood $\ell(\theta; y, T)$ was maximized with respect to θ using the SAEM algorithm (Delyon et al., 1999; Kuhn and Lavielle, 2004) implemented in the Monolix software version 2019R2 (http://lixoft.com). The likelihood was estimated by Monte Carlo integration using an Importance Sampling algorithm. All calculations were performed in an Intel 8th Gen Core i7 processor 64bits, 2.77 GHz 4 Core with 16GB RAM.

4 Results

In model (2), the vector $\phi_i = (a_i, b_i, c_i)'$ characterizes the profile for the i-th woman. The parameter a_i refers to the asymptotic β-HCG hormone level, b_i refers to the time at which the woman reaches half of the asymptotic β-HCG hormone level, and c_i is the time elapsed for the woman to reach between half and three fourths of its asymptotic β-HCG hormone level. We fit the four models considered in Table 1 for the pregnant women dataset described in Section 2. Table 2 shows the estimated parameters, standard errors, as well as the BIC and AIC values. The results for the estimation of the coefficients a_{pop}, b_{pop}, c_{pop}, δ and the variance components σ, ω_a, ω_b and ω_c for the longitudinal part present some important differences, the same happens with the estimation of the coefficients β, γ, and α for the TTE model. Here, $\beta = \beta_1$ for models $\mathcal{M}_1$ and $\mathcal{M}_2$, whereas $\beta = \beta_2$ for models $\mathcal{M}_3$ and $\mathcal{M}_4$. The inclusion of the covariable z_i in the longitudinal part significantly improves the values of the AIC and BIC criteria in both models $\mathcal{M}_2$ (vs $\mathcal{M}_1$) and $\mathcal{M}_4$ (vs $\mathcal{M}_3$). Additionally, the Wald tests for the hypothesis $H_0 : \delta = 0$ in models $\mathcal{M}_2$ and $\mathcal{M}_4$ are 8.43 and 9.81 respectively ($p < 0.001$). These results confirm that each group should have a different mean asymptote (i. e. the random coefficient a_i has mean $a_{\text{pop}} + \delta$ for the abnormal group and a_{pop} for the normal group).

We observe that the best joint models, based on AIC and BIC criteria, are the proposed models $\mathcal{M}_2$ and $\mathcal{M}_4$. The AIC and BIC criteria for joint models $\mathcal{M}_2$ and $\mathcal{M}_4$ are similar, but the nature of the dependency assumed between the longitudinal β-HCG biomarker and the time to early miscarriage is different for these specific models. Model $\mathcal{M}_2$ included the individual deviations of the longitudinal β-HCG biomarker, that are the random effects, as

predictors in the TTE model. Model $\mathcal{M}_4$ assumes that the prediction of the longitudinal β-HCG biomarker is predictive of the time-to-event of early miscarriage. From our results for model $\mathcal{M}_2$, we estimate that a_i have a mean equal to $\hat{a}_{\text{pop}} = 4.76$ for the normal pregnant women and $\hat{a}_{\text{pop}} + \hat{\delta} = 3.975$ for the abnormal pregnant women. We also have that the time at which the woman reaches half of the asymptotic β-HCG hormone level is around 16 days, and that the time elapsed for the woman to reach between half and three fourths of its asymptotic β-HCG hormone level is around 7 days.

Whatever the assumed dependency between the longitudinal β-HCG biomarker and the time-to-event of early miscarriage is, the four models considered here concluded that there exists a significant association between the two processes. The corresponding Wald tests for the hypothesis $H_0 : \beta = 0$ are 5.5, 2.6, 39.1 and 13.4 for models $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$, respectively ($p < 0.05$). However, based on the two-stage joint model of the formulation corresponding to model $\mathcal{M}_2$, the Wald test is 1.85 (results not shown here for brevity of exposition), concluding that the β-HCG asymptote did not have a significant impact on the time-to miscarriage. The results show the benefits of use joint models in comparison with the two-stage joint model in order to understand the association between the longitudinal β-HCG biomarker and the time-to-event of early miscarriage.

Table 2: Estimated parameters and their standard errors (in parentheses). Here, $\beta = \beta_1$ for models $\mathcal{M}_1$ and $\mathcal{M}_2$, whereas $\beta = \beta_2$ for models $\mathcal{M}_3$ and $\mathcal{M}_4$.

Parameters	Model $\mathcal{M}_1$	Model $\mathcal{M}_2$	Model $\mathcal{M}_3$	Model $\mathcal{M}_4$
a_{pop}	$4.61_{(0.0486)}$	$4.76_{(0.048)}$	$4.57_{(0.0487)}$	$4.75_{(0.045)}$
δ	-	$-0.784_{(0.093)}$	-	$-0.775_{(0.079)}$
b_{pop}	$15.91_{(0.555)}$	$15.80_{(0.479)}$	$16.08_{(0.532)}$	$15.70_{(0.477)}$
c_{pop}	$7.24_{(0.384)}$	$7.01_{(0.405)}$	$6.90_{(0.396)}$	$6.98_{(0.340)}$
ω_a	$0.435_{(0.0357)}$	$0.323_{(0.0396)}$	$0.468_{(0.0396)}$	$0.326_{(0.0349)}$
ω_b	$4.72_{(0.427)}$	$3.51_{(0.708)}$	$4.43_{(0.489)}$	$3.95_{(0.534)}$
ω_c	$1.05_{(0.310)}$	$1.99_{(0.40)}$	$1.51_{(0.325)}$	$1.72_{(0.366)}$
σ	$0.252_{(0.0171)}$	$0.253_{(0.0216)}$	$0.236_{(0.017)}$	$0.249_{(0.0192)}$
β	$1.58_{(0.287)}$	$0.717_{(0.280)}$	$0.633_{(0.0162)}$	$0.468_{(0.0349)}$
γ	$12.42e^{-11}_{(2.09e^{-11})}$	$4.97e^{-8}_{(2.33e^{-8})}$	$8.52e^{-7}_{(2.04e^{-7})}$	$1.36e^{-6}_{(6.62e^{-8})}$
α	$4.61_{(0.418)}$	$3.63_{(0.325)}$	$3.01_{(0.032)}$	$3.06_{(0.027)}$
BIC	730.63	668.01	739.27	668.25
AIC	699.09	633.32	707.74	633.57

Next, for the selection of the joint model, model evaluation relied on the analysis of the model individual predictions, population predictions and residuals, as well as on the analysis of the survival predictions. The goodness-of-fit plots for model $\mathcal{M}_2$ are shown in Figures 3, 4 and 5. These plots are useful to detect any possible mispecifications in the structural and residual error models. The population and individual models allow us to calculate predic-

tions $m(t_{ij}, \hat{\mu}_i)$ and $m(t_{ij}, \hat{\phi}_i)$ for each woman at the observation times t_{ij}. Observed β-HCG hormone concentrations versus population predicted (left) and individual predicted (right) β-HCG hormone concentrations are included in Figure 3, where it is clear that there is no evidence for any model misspecification. That is, the data are well described by model $\mathcal{M}_2$ as assessed by visual inspection of the corresponding diagnostic plots. The residuals for the longitudinal part were assessed by using the Individual Weighted Residuals (IWRES), defined as $\text{IWRES}_{ij} = \left[y_{ij} - m(t_{ij}, \hat{\phi}_i)\right] / \hat{\sigma}$ where $\hat{\sigma}$ is the estimated standard deviation of the error term and $\hat{\phi}_i$ is the vector of the predicted individual parameters, i.e., the Empirical Bayes Estimates. The analysis of the IWRES did not suggest any model misspecification in the model fitting across time (left panel in Figure 4). The residuals at the population level were assessed using the normalized prediction distribution errors (NPDE) which is a nonparametric version of the population weighted residuals (PWRES), based on a rank statistic (Comets, Brendel and Mentré, 2008). The PWRES are defined as $\text{PWRES}_{ij} = \left[y_{ij} - \hat{\text{E}}(y_{ij})\right] / \hat{\text{std}}(y_{ij})$, where $\hat{\text{E}}(y_{ij})$ and $\hat{\text{std}}(y_{ij})$ are the mean and variance of y_{ij} estimated by Monte Carlo. From Figure 4 (right panel) no major systematic bias is observed for the NPDE. The points were equally distributed around zero and most of the data in included in the -2 to +2 range, indicating acceptable agreement between observed and predicted concentrations. Figure 5 includes the comparison between the empirical and theoretical probability density function (PDF) and cumulative distribution function (CDF) for the IWRES, in the left and right panels, respectively. These Figures suggest that the normality assumption for the error terms is a reasonable one. Figure 6 presents the visual predictive check (VPC) for the longitudinal part for each group. The VPC is a diagnostic tool which allows the researcher to be able to summarize, in the same graphical display, the structural and statistical models by computing quantiles for the empirical distribution of the data after having regrouped them into bins over successive intervals. Using Monte Carlo, we compute prediction intervals for these quantiles under the selected model (i.e., model $\mathcal{M}_2$ in our case). More specifically, in Figure 6, we report the median of the empirical distribution of the data (solid line), the predicted median and the respective 95% prediction interval. For more details on how a VPC is constructed in Monolix, see Lavielle (2015). The VPC for the abnormal pregnancy group seems to indicate that the structural model can be improved by considering a more complex model specification. Moreover, the analysis of the IWRES and NPDE do not suggest any model misspecification, showing that both the structural model and the residual error model properly fit the data. In addition to the residuals analysis and in order to evaluate the overall prediction for the survival, the mean survival curve was also calculated and compared to the Kaplan-Meier curve. As can be seen in Figure 7, the mean survival curve was close to the Kaplan-Meier curve, concluding that an approximate 23 day lag is evident before an event occurs, which is a direct consequence of the use of the time since conception as a timeline. Finally, from the analysis of the dataset under study and the results reported above, it can be seen that one of the clear advantages of the proposed methodology is that it allows for the inclusion of several specific random effects in the joint models and two-stage approach and convergence

occurs in a natural way when using the Monolix software, whereas the use of the two-stage approach is not possible with well known software packages, such as, for example, NLME in R, because convergence cannot be attained. A simulated dataset and the code needed to fit model $\mathcal{M}_2$, compute the residuals and generate the diagnostic figures presented above are available as supplementary materials with a README file.

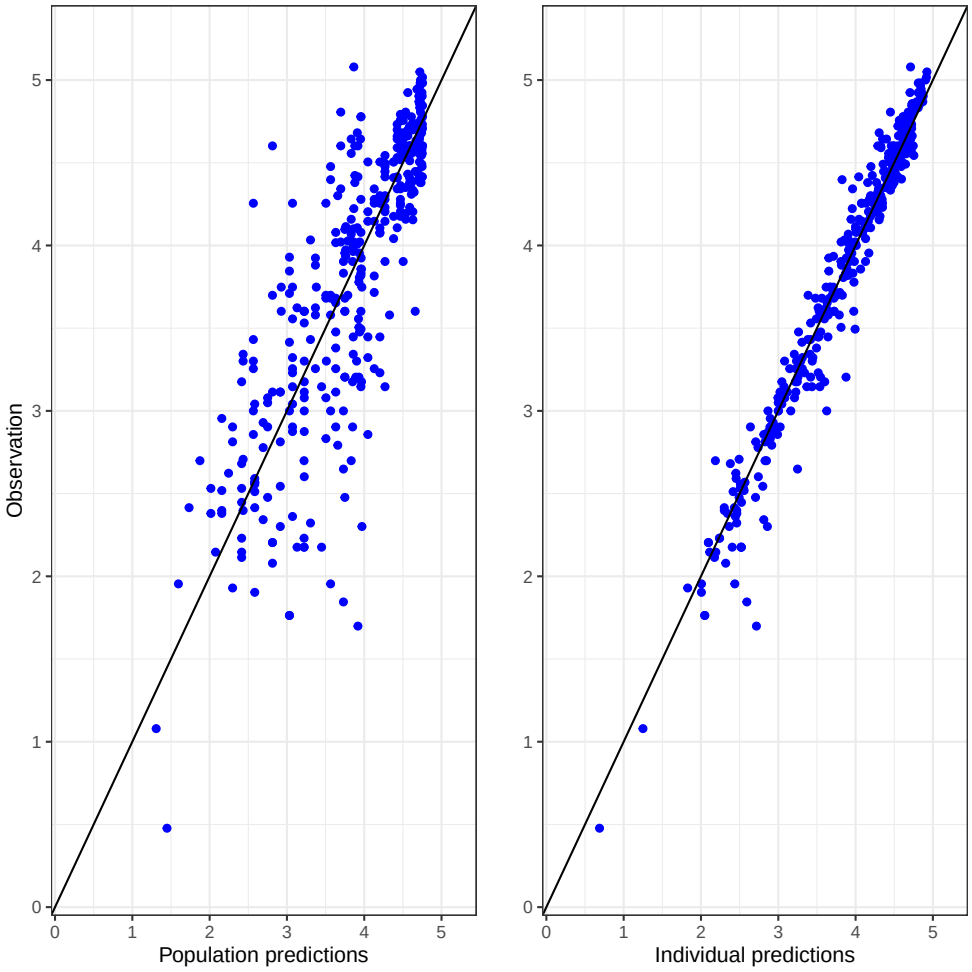


Figure 3: Observations versus predictions for the β-HCG hormone concentrations for model $\mathcal{M}_2$, with population predictions on the left and individual predictions on the right.

5 Discussion

Alternative proposals in joint modeling of longitudinal biomarkers and time-to-event data are important and of interest to better understand the possible connection between biological changes in time and the occurrence of an event clinicians are interested in studying. In

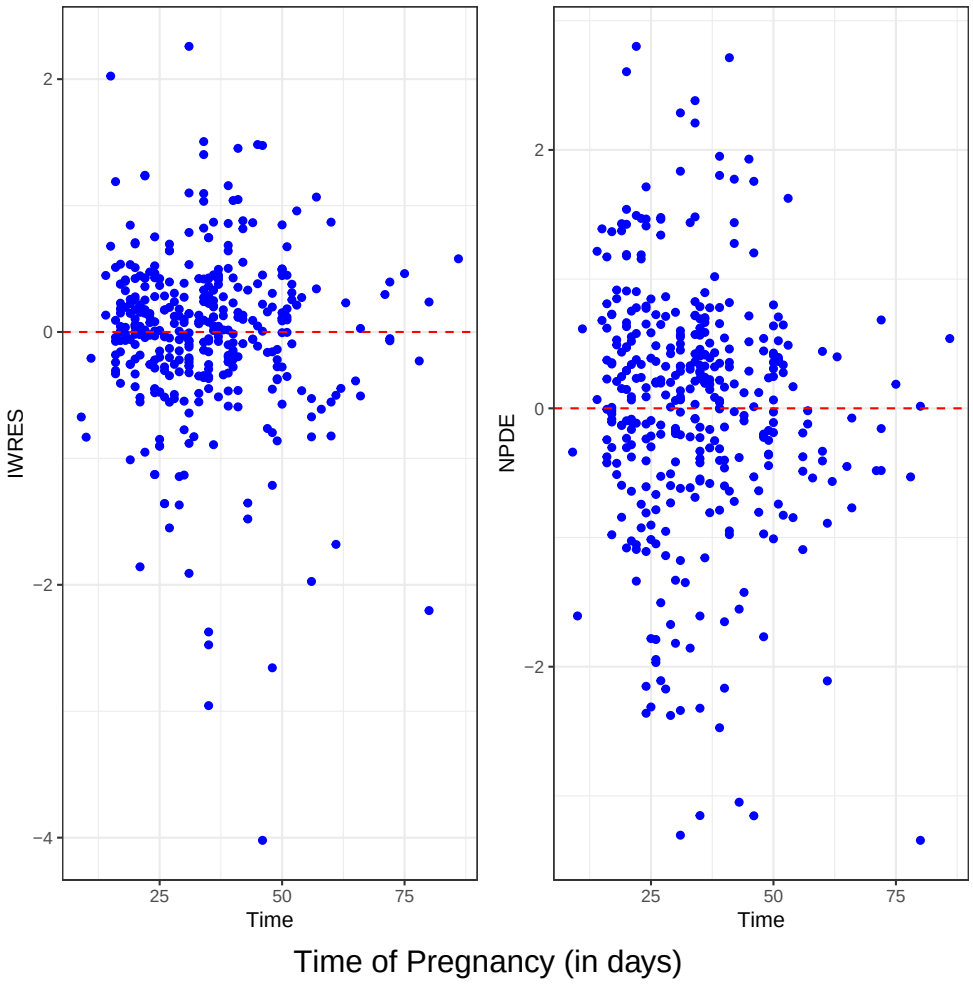


Figure 4: Individual weighted residuals versus time (left) and normalized prediction distribution errors versus time (right) for the β-HCG hormone concentrations for model $\mathcal{M}_2$.

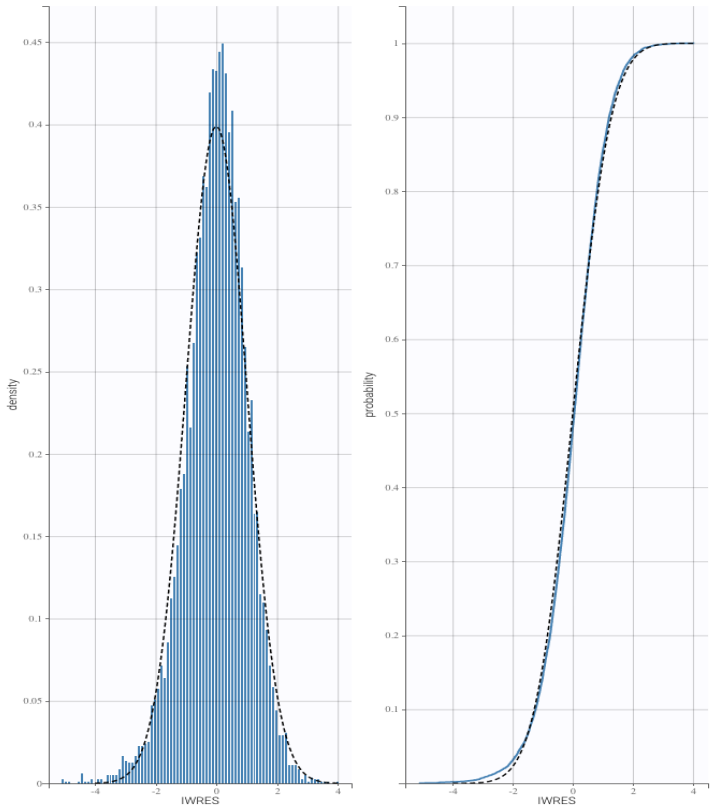


Figure 5: Empirical and theoretical probability density function (PDF) (left panel) and cumulative distribution function (CDF) (right panel) comparison for the individual weighted residuals (IWRES) for model $\mathcal{M}_2$.

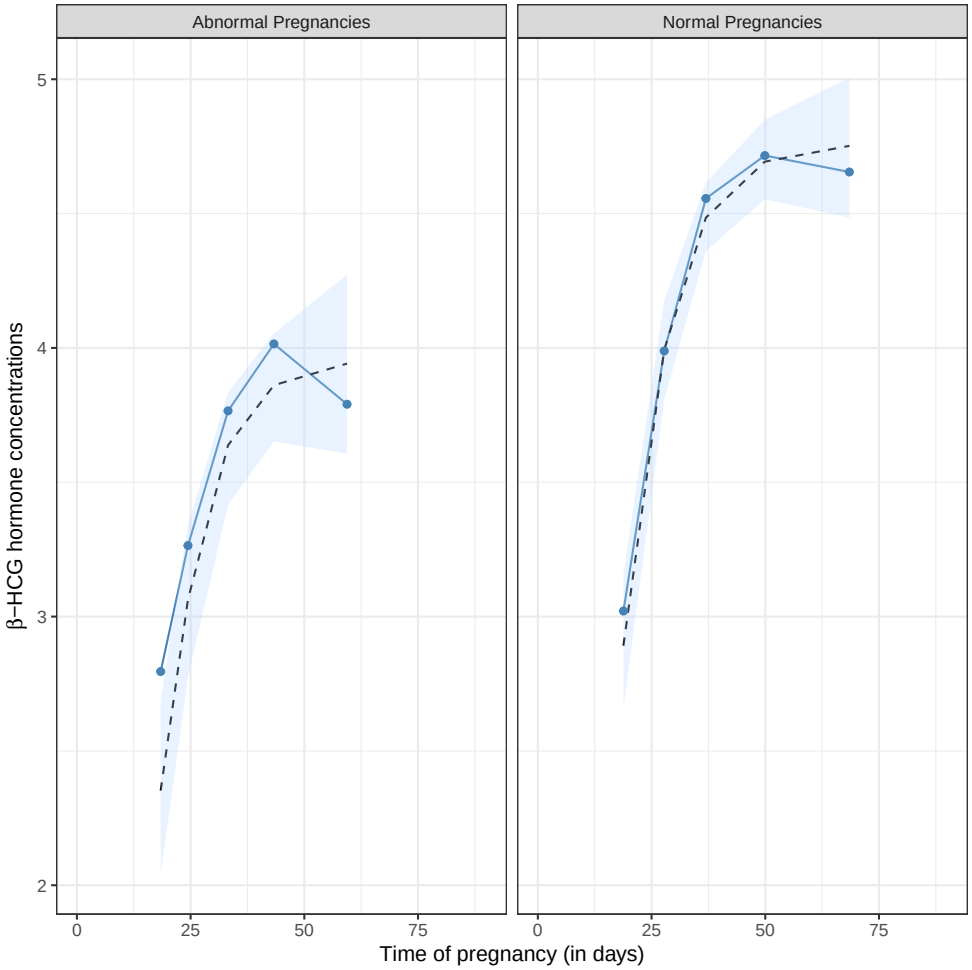


Figure 6: Visual predictive check (VPC) for the β-HCG hormone concentrations with the 95% prediction intervals for the 50th percentile, empirical median (in solid line) and predicted median (in dashed line), separated by group for model $\mathcal{M}_2$.

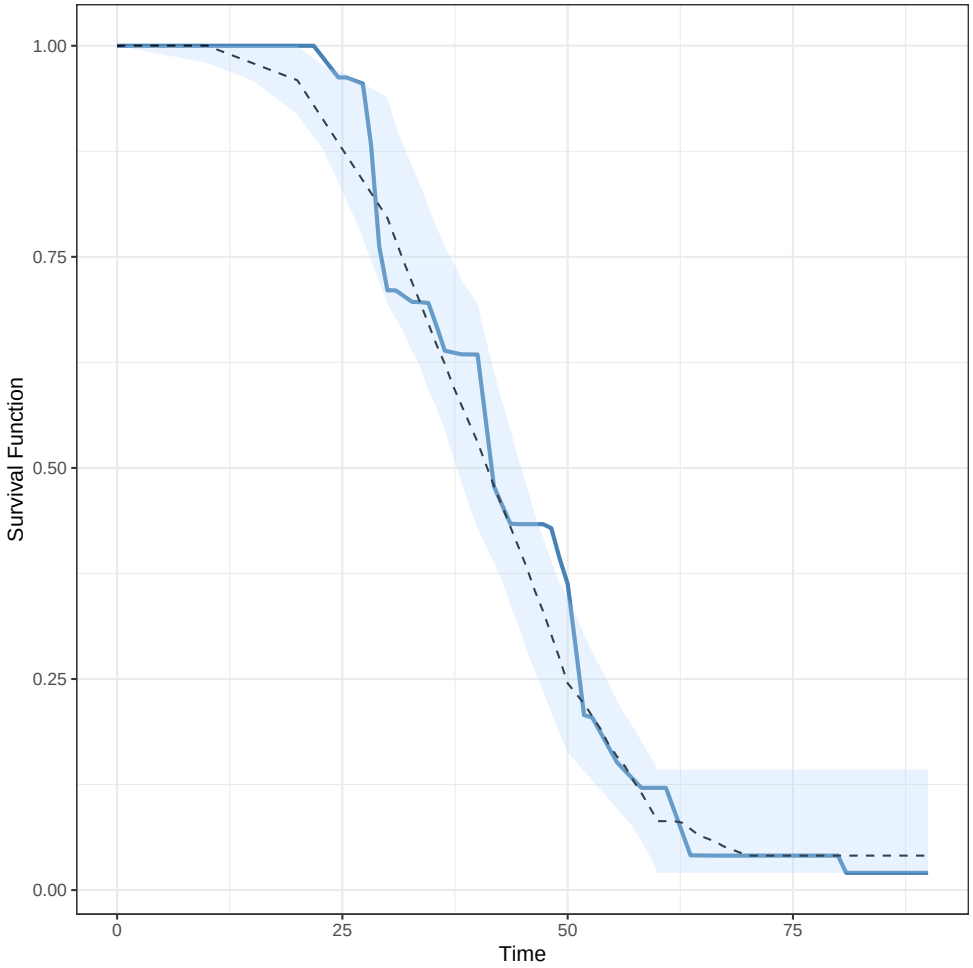


Figure 7: Survival function estimated for abnormal group using the SAEM algorithm via Monolix for model $\mathcal{M}_2$. In blue, the predicted interval for the Kaplan Meier plot, in dashed line the predicted median.

recent years, there has been much interest in putting forward proposals were linear mixed-effects models and time-to-single-event data are jointly modeled, and parameter estimation has been performed, often using maximum likelihood together with the EM algorithm. However, when there are several random effects in the model, likelihood calculations may have a convergence problem, discouraging the use of these methods. Motivated by the clinical question regarding the association between β-HCG values and the risk of an abnormal delivery or fetus loss for pregnant women under the age of 30 in Chile, we proposed an innovative joint modeling approach of a parametric Weibull model for the time-to-event data and a nonlinear mixed effects model for the longitudinal measurements, including an alternative joint modeling approach for a subgroup of the data under study, and also using the remaining part of the data as part of the analysis. The dependency was handled via random effects which were naturally incorporated into the models. Estimation procedures based on the SAEM algorithm were also proposed, where we have verified in our analysis and proposal that the subgroup of women having a *normal* delivery was included in the analysis, that the joint modeling approach was implemented to a subgroup of the dataset under study, and that one of the advantages of the proposed methodology is that it allows for the inclusion of several specific random effects in the model and convergence occurs in a natural way, whereas the use of this model is not possible with well known software packages, such as, for example, NLME in R, because covergence cannot be attained for the two-stage approach.

Here, we have shown that the SAEM algorithm can be considered as a valid methodological alternative for performing parameter estimation in joint models where the mixed-effects were nonlinear and a parametric Weibull TTE model was considered for a subgroup of the dataset under study; i.e., the one corresponding to the abnormal group, mainly because clinicians were interested in both modeling this behavior with a TTE survival model but, at the same time, be able to also follow the longitudinal evolution for the behavior of the β-HCG biomarker for both groups. That is, we only applied the TTE survival model to a subgroup of the original data, because physicians were really interested in studying the time from pregnancy to fetus loss, which was a phenomenon only occurring in the abnormal group. In order to be able to implement the SAEM algorithm for joint models for the aforementioned study and objectives, we derived specific expressions for the conditional likelihood of the observations given the individual parameters. Our analyses showed that convergence of the SAEM algorithm, within the joint modeling approach we have proposed, is fast, specially when compared to alternative likelihood approaches for which, in some cases, convergence was not achieved. As a consequence, we can also quickly estimate individual parameters and be able to carry out any inferential process related to the parameters included in the proposed models. In our modest view, this represents a clear advantage when comparing our proposed approach to previous ones in the area.

In summary, given that, in our view, this proposal represents a simple, fast and high-performance tool for joint modeling, we believe they should now be used more frequently in the relevant statistical practice.

6 Acknowlegments

This work was supported by Ministerio de Economía y Competitividad (Spain), Agencia Estatal de Investigación (AEI), and the European Regional Development Fund (ERDF), under research grant MTM2016-74931-P (AEI/ERDF, EU), by the Department of Education of the Basque Government (UPV/EHU Econometrics Research Group) under research grant IT-1359-19, by the National Agency for Research and Development (ANID) / FONDECYT / 1181662 & 1190801, and the Math AmSud 18-MATH-07 SaSMoTiDep project.

References

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

Comets, E., Brendel, K. and Mentré, F. (2008). Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Computer Methods and Programs in Biomedicine*, **90**, 154-166.

Davidian, M. and Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurements Data*. London: Chapman and Hall.

DeGruttola, V. and Tu, X. (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics*, **50**, 1003-1014.

Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Stat.*, **27**, 94-128.

Durbán, M., Harezlak, J., Wand, M.P. and Carroll, R.J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153-1167.

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

France, J.T., Keelan, J., Song, L., et al. (1996). Serum concentrations of human chorionic gonadotrophin and immunoreactive inhibin in early pregnancy and recurrent miscarriage: a longitudinal study. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, **36**, 325-330.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Biometrics*, **61**, 383-385.

Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM P.& S.*, **8**, 115-131.

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

Lavielle, M. (2015). *Mixed Effects Models for the Population Approach*. New York: Chapman and Hall/CRC.

Lindstrom, M.J. and Bates, D.M. (1990). Non-linear mixed-effects models for repeated measures. *Biometrics*, **46**, 673-687.

Liu, M. and Ying, Z. (2007). Joint analysis of longitudinal data with informative right censoring. *Biometrics*, **63**, 363-371.

Mbogning, C., Bleakley, K. and Lavielle, M. (2015). Joint modeling of longitudinal and repeated time-to-event data using nonlinear mixed-effects models and the stochastic approximation expectation-maximization algorithm. *Journal of Statistical Computation and Simulation*, **85(8)**, 1512-1528.

Murawska, M., Rizopoulos, D. and Lesaffre, E. (2012). A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *Journal of Probability and Statistics*, **2012**, article ID 194194, 18 pages.

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

Rizopoulos, D., Verbeke, G. and Molenberghs, G. (2009). Fully exponential Laplace approximation for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society-Series B*, **71**, 637-654.

Ruppert, D., Wand, M. and Carroll, R. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.

Self, S. and Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. In *AIDS Epidemiology: Methodological Issues* (N.P. Jewell, K. Dietz and V.T. Farewell, eds.). Boston, MA: Birkhaüser, 231-255.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Vonesh, E.G. and Chinchili, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

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