A semi-parametric joint model for two sequential times to events and one longitudinal covariate

Carles Serrat[1], Jaime-Abel Huertas[2], Guadalupe Gómez[3]

[1] Dept. Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Barcelona, Spain
[2] Dept. Estadística, Universidad Nacional de Colombia, Bogotá, Colombia
[3] Dept. Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, Spain

E-mail for correspondence: carles.serrat@upc.edu

Abstract: The present work proposes a joint model for two sequential times to events together with longitudinal information, as an extension of the joint model by Wolfsohn and Tsiatis (1997) for one time to event and one longitudinal variable. The model is applied to the TIBET, a clinical trial in which antiretroviral therapy interruptions guided by CD4 counts and plasma HIV-1 RNA levels in chronically HIV-1-infected patients are under evaluation. Details on the modelling strategy and the resulting estimates are given.

Keywords: Joint Modelling; Longitudinal Data; Sequential Times; Survival Analysis.

1 Introduction and the Motivating Clinical Trial

This research is motivated by the TIBET clinical trial (Ruiz *et al.*, 2007) and we want to model two sequential times to event with longitudinal measurements. T_1 is the time to re-initiate $HAART$ therapy and T_2 is the time to suspend therapy from the first re-initiation, and the longitudinal measurements are the levels of CD4 cell counts each four weeks. The goal is to use the longitudinal measurements as a marker for the times to event. The joint model will allows us to give prognosis for a time to event given covariates, the longitudinal process and the previous event time.

2 Notation

The variables of interest for each subject $i = 1, ..., n$ followed over an interval $[0,\ \tau)$ are $\{T_{1i}, T_{2i}, R_i(u), 0 \leq u \leq \tau, X_i\}$, where T_{1i} and T_{2i} are event times, $\{R_i(u), 0 \leq u \leq \tau\}$ is the longitudinal response trajectory

for all times $u \geq 0$ and $X_i = [X_{1i}^T \ \ X_{2i}^T]^T$ is a vector of baseline (time 0) covariates, X_{1i} with influence over T_1, and X_{2i} over T_2.
We will consider only a situation where T_1 and T_2 may be right censored by the censoring times C_1 and C_2 respectively, so instead of T_{ji} we observe (Y_{ji}, δ_{ji}), $j = 1, 2$, where $Y_{ji} = min\{T_{ji}, C_{ji}\}$ and $\delta_{ji} = I(T_{ji} \leq C_{ji})$ which indicates whether Y_{ji} is an uncensored right value of T_{ji}. On the other hand, for some set of times $t_{ij}, j = 1, ..., n_i$, instead of the true values $R_i(t_{ij})$ we observe $Z_i(t_{ij})$, then the observed data for subject i is $O_i = \{X_i, Y_i, \delta_i, Z_i, \tilde{t}_i\}$, where $\tilde{t}_i = (t_{i1}, ..., t_{in_i})^T$, $Z_i = (Z_i(t_{i1}), ..., Z_i(t_{in_i}))^T$, $Y_i = (Y_{1i}, Y_{2i})$, and $\delta_i = (\delta_{1i}, \delta_{2i})$.

3 Semi-parametric Joint Model of Two Sequential Times to Event and One Longitudinal Variable

We approach the problem of two sequential times to event with a sequence of conditional distributions (Lawless, 2003, Section 11.3). A natural choice for the survival model is to consider a distribution for a time to event given previous observed event times. Moreover, the sequence of conditional distributions for the times to event is jointly modeled with a mixed model, adapting the model of Wulfsohn and Tsiatis (1997).

3.1 Linear Trend for the Longitudinal Variable

First, we assume that the longitudinal variable is monotone not increasing (or viceversa) with linear trend as we show in the Figure 1 part (a). This is the simplest case for the longitudinal variable that we consider. Due to the existing association between the longitudinal and survival processes, there is a high probability that the longitudinal trend changes with the occurrence of the first event. Nevertheless we treat it because might happen cases where the monotonous trend over time persists. In the TIBET clinical trial we found that the GPT enzyme has linear trend over T_1 and T_2, so we need the methodology with longitudinal linear trend to analyze whether the effect of the enzyme in the times to event is significant.
We analyze our proposal linking the longitudinal and survival sub-models with the current value, and assuming a linear trend for the longitudinal data without fix part. A particular joint model analyzed, is

$$Z_{ij} = b_{0i} + b_{1i}\ t_{ij} + e_i(t_{ij}) \tag{1}$$

$$\lambda_{T_1}(t_1 \mid b_i, X_{1i}; \eta_1, \beta_1) = \lambda_{1,0}(t_1) \exp\{\eta_1 X_{1i} + \beta_1(b_{0i} + b_{1i}t_1)\} \tag{2}$$

$$\lambda_{T_2}(t_2 \mid t_{1i}, b_i, X_{2i}; \eta_2, \beta_2, \gamma) = \lambda_{2,0}(t_2) \exp\{\eta_2 X_{2i} + \beta_2(b_{0i} + b_{1i}(t_{1i} + t_2)) + \gamma t_{1i}\}, \tag{3}$$

where η_1 and η_2 are vectors of parameters associated to baseline covariates, β_1 and β_2 are parameters of association between the longitudinal and

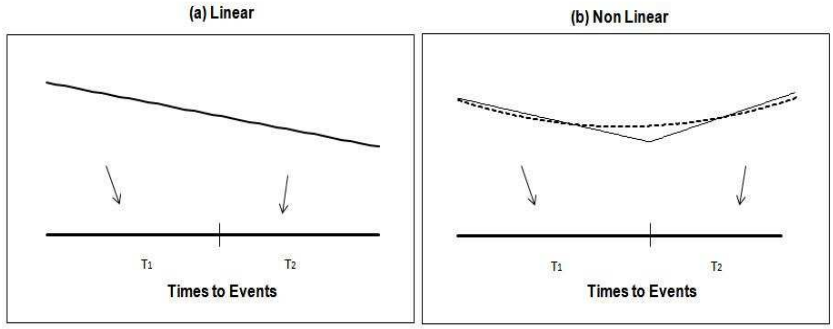

FIGURE 1. Two types of trends for the longitudinal data over T_1 and T_2

survival processes, γ describes the relation among the times to event, and both baseline risks $\lambda_{1,0}(\cdot)$ and $\lambda_{2,0}(\cdot)$ are left unspecified, and different.
For the likelihood function we apply the same assumptions made by Wulfsohn and Tsiatis (1997). The assumption of non-informative censoring extend to this case of censoring process. The errors e_i are assumed to be mutually independent, normally distributed with mean 0 and variance σ_e^2, and independent of b_i and conditionally independent of all other variables given (b_i, X_i). If we assume that, given random effects and covariates, Z, T_1, and $T_2 \mid T_1$, are all independent, then the observed likelihood is

$$L(\Omega) = \prod_{i=1}^{n} \int_{b_i} \left\{ \prod_{j=1}^{n_i} f(z_{ij} \mid b_i; \sigma_e^2) \right\} f(Y_i, \delta_i \mid b_i, X_i; \psi_{T|b}) f(b_i; B, \Gamma) \mathrm{d}b_i, \quad (4)$$

where $\Omega = (\psi_{T|b}, B, \Gamma, \sigma_e^2)$ and $\psi_{T|b} = (\eta_1, \eta_2, \beta_1, \beta_2, \gamma, \lambda_{1,0}, \lambda_{2,0})$. The vector of random effects $b_i = [b_{0i}\ b_{1i}]^T$ is taken to be normally distributed with mean B and covariance matrix Γ.
The algorithm by Wulfsohn and Tsiatis (1997) is extended in many cases of joint models with longitudinal and survival data. In our case the model is extended and applied in two parts. The first step models Z and T_1+T_2 using the Wulfsohn and Tsiatis's algorithm ignoring the sequence of the times to event, and the second step fits T_1 and $T_2 \mid T_1$ likely the modelling was direct and jointly with Z. We named this method as EM modified algorithm for joint model with longitudinal information and sequential times to event.

3.2 Non Linear Trend for the Longitudinal Variable

When the first time to event occurs the conditions in the statistical units may change, affecting the evolution of the longitudinal variables. For our TIBET dataset, the CD4 cell counts changes its trend when the therapy is

restated. Figure 1 part (b) shows how the trend may be linear in two pieces or parabolic. We focus the analysis for this two cases of non linear trend.

a)

$$Z_{ij} = b_{0i} + (b_{1i}\ t_{ij} + b_{2i}(t_{ij} - t_{1i})I(t_{ij} \geq t_{1i})) + e_i(t_{ij}) \tag{5}$$

$$\lambda(t_1 \mid b_i; \beta_1) = \lambda_{1,0}(t_1) \exp\{\beta_1(b_{0i} + b_{1i}t_1)\} \tag{6}$$

$$\lambda(t_2 \mid t_{1i}, b_i; \beta_2, \gamma) = \lambda_{2,0}(t_2) \exp\{\beta_2(b_{0i}+b_{1i}\cdot t_{1i}+(b_{1i}+b_{2i})t_2)+\gamma\ t_{1i}\}. \tag{7}$$

b)

$$Z_{ij} = b_{0i} + b_{1i}\ t_{ij} + b_{2i}\ t_{ij}^2 + e_i(t_{ij}) \tag{8}$$

$$\lambda(t_1 \mid b_i; \beta_1) = \lambda_{1,0}(t_1) \exp\{\beta_1(b_{0i} + b_{1i}t_1 + b_{2i}t_1^2)\} \tag{9}$$

$$\lambda(t_2 \mid t_{1i}, b_i; \beta_2, \gamma) = \lambda_{2,0}(t_2) \exp\{\beta_2(b_{0i}+b_{1i}(t_{1i}+t_2)+b_{2i}(t_{1i}+t_2)^2)+\gamma t_{1i}\}, \tag{10}$$

4 Application to the TIBET Dataset

In the TIBET dataset T_1 is the time to the first restart of therapy, and T_2 is the time from the first restart of therapy to the suspension of it. We have 74% observed cases for T_1 and 68% observed cases for T_2.
The evolution of the CD4 is decreasing until the first time to event, and then is increasing, so we model it with a two piecewise based on (5)-(7), and with a parabolic trend based on (8)-(10). The analysis of these models only aims to compare them in order to determine the best alternative to model the longitudinal variable for prognosis. Table 1 shows the results for both models fitted with the EM modified algorithm, and Figure 2 shows some fitted cases with these models.

The selected joint model is as follow, and the results obtained with the EM modified algorithm are shown in Table 2.

$$Z_{ij} = b_{0i} + (b_{1i}\ t_{ij} + b_{2i}(t_{ij} - t_{1i})I) + e_i(t_{ij}) \tag{11}$$

$$\lambda(t_1 \mid b_i, VL_i; \eta_1, \beta_1) = \lambda_{1,0}(t_1) \exp\{\eta_1 VL_i + \beta_{11}(b_{0i} + b_{1i}t_1) + \beta_{12}b_{1i}\} \tag{12}$$

$$\lambda(t_2 \mid t_{1i}, b_i, VL_i; \eta_2, \beta_2, \gamma) = \lambda_{2,0}(t_2) \exp\{\eta_2 VL_i + \beta_{21}(b_{0i} + b_{1i}(t_{1i} + t_2) + b_{2i}t_2) + \beta_{22}(b_{1i} + b_{2i}) + \gamma\ t_{1i}\}. \tag{13}$$

The influence of the intercepts b_0 and $b_0 + b_1t_1$ in T_1 and T_2 respectively, are not significative. It is logic since patients start the trial without therapy with good and similar conditions, and the restart of therapy is due to the threshold reached in the levels of CD4 and viral load.

TABLE 1. Joint models for T_1 and T_2 with two different model for CD4. It is assumed semi parametric form in the hazard risks.

	Two Piecewise			Parabolic		
Parameter	*Estimate*	*s.e.*	$p-value$	*Estimate*	*s.e.*	$p-value$
Mixed						
B_0	25.8263	0.4569	< 0.0001	26.5564	0.4709	< 0.0001
B_1	-0.0502	0.0041	< 0.0001	-0.0732	0.0092	< 0.0001
B_2	0.1248	0.0080	< 0.0001	0.0007	0.0001	< 0.0001
σ_{11}	20.8764	2.9524	< 0.0001	22.1763	3.1362	< 0.0001
σ_{12}	-0.1186	0.0221	0.6483	-0.1964	0.0477	0.7388
σ_{13}	0.0166	0.0364	< 0.0001	0.0001	0.0003	< 0.0001
σ_{22}	0.0017	0.0002	< 0.0001	0.0085	0.0012	< 0.0001
σ_{23}	-0.0016	0.0004	< 0.0001	-4E-5	6E-6	< 0.0001
σ_{33}	0.0063	0.0009	< 0.0001	3E-7	4E-8	< 0.0001
σ_e^2	6.1463	0.1881	< 0.0001	5.5382	0.1695	< 0.0001
Survival T_1						
β_1 (Assoc.)	-0.1881	0.0361	<0.0001	-0.1060	0.0335	0.0016
Survival T_2						
β_2 (Assoc.)	0.0465	0.0414	0.2614	-0.0002	0.0386	0.9958
γ (T_1)	-0.0172	0.0063	0.0063	-0.0178	0.0072	0.0134

TABLE 2. Survival result for the joint model for T_1 and T_2 with two piecewise mixed model for the CD4 evolution, based in EM modified algorithms. It is assumed semi parametric form in the hazard risks.

Parameter	*Estimate*	*s.e.*	$p-value$
Survival T_1			
β_{11} $(b_{0i}+b_{1i}t_1)$	-0.2044	0.0405	< 0.0001
β_{12} (b_{1i})	-14.3298	3.5632	< 0.0001
η_1 (VL_i)	0.6521	0.1858	0.0004
Survival T_2			
β_{21} $(b_{0i}+b_{1i}(t_{1i}+t_2)+b_{2i}t_2)$	0.0257	0.0430	0.5500
β_{22} $(b_{1i}+b_{2i})$	6.7904	2.4090	0.0048
η_1 (VL_i)	-0.0169	0.2302	0.9414
γ (t_{1i})	-0.0210	0.0079	0.0079

The slope is the only significative random effect in T_2, and due to the fact that the effect of the viral load pre-therapy is diluted in T_2, then we have the slope of the longitudinal variable along T_2 and the observed values of T_1, the only significative covariate in the survival model for T_2.
The negative sign for $\hat{\gamma}$ indicates that for long times to restart therapy, we have long times to suspend therapy.

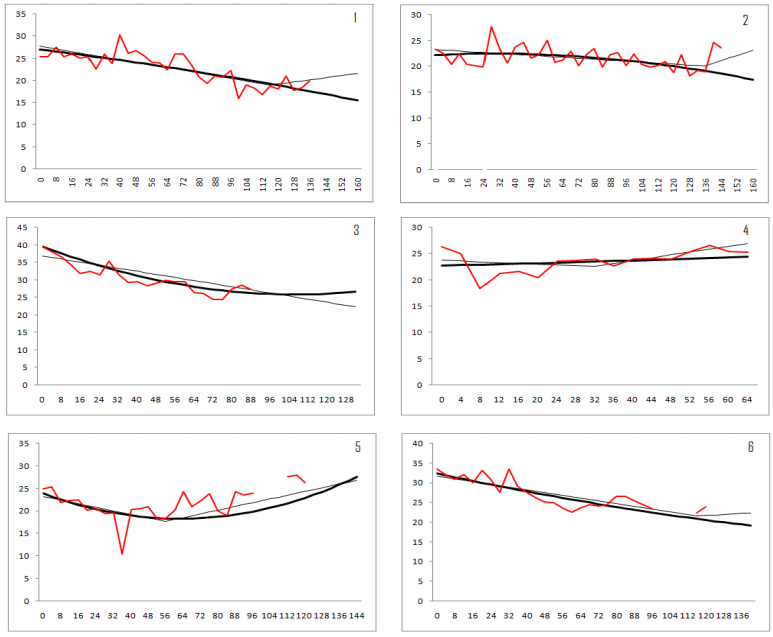

FIGURE 2. Evolution of the CD4 with models based in two piecewise and parabolic trend, for some cases of the TIBET dataset

Acknowledgments: This work has been partially supported by the grant MTM2008–06747–C02-01 from the Ministerio de Ciencia e Innovación of Spain. Authors are indebted to the GRASS group for fruitful discussions and to the Fundació Lluita contra la SIDA for providing the data.

References

Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data, 2nd edition. *Wiley* Hoboken.

Ruiz, L., Paredes, R., Gómez, G., Romeu, J., Domingo, P., Pérez-Alvarez, N., Tambussi, G., Llibre, J.M., Martnez-Picado, J., Vidal, F., Fumaz, C.R., Clotet, B., and Group, T.I.B.E.T.S. (2007). Antiretroviral therapy interruption guided by CD4 cell counts and plasma HIV-1 RNA levels in chronically HIV-1-infected patients. *AIDS*, **21**(2), 169–178.

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