A Joint Model with Marginal Interpretation for Longitudinal Continuous and Time-to-event Outcomes

Achmad Efendi[1], Geert Molenberghs[2,1], Edmund Njagi[1], Paul Dendale[3]

[1] I-BioStat, Katholieke Universiteit Leuven, Leuven, Belgium
[2] I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium
[3] Virga Jesse Hospital, Heart Center Hasselt, Hasselt, Belgium

E-mail for correspondence: achmad.efendi@student.kuleuven.be

Abstract: This paper proposes a marginalized joint model for longitudinal continuous and repeated time-to-event outcomes, extending work of Njagi et al. (2012), as well as a marginalized joint model for bivariate repeated time-to-event outcomes.

Keywords: Joint Modeling; Combined Model; Marginalization.

1 Introduction

Ever more commonly does one jointly collect longitudinal and time-to-event outcomes, the latter possibly censored. While an extensive amount of literature is available for Gaussian and other longitudinal outcomes, and literature on the joint modeling of a longitudinal outcome and a single time to event is rapidly growing, methods for the more general setting where at least two longitudinal sequences of perhaps different data types are jointly recorded has received less attention so far, especially when one or more of the sequences consist of times-to-event. Nevertheless, such designs are not uncommon in practice, as our two case studies, introduced in the next section, underscore. Recently, Njagi et al. (2012) formulated joint models for pairs of jointly measured outcomes where for each type of outcome, two sets of random effects are considered, the conjugate and the normal random effects, extending the so-called combined model introduced by Molenberghs et al. (2010). However, the joint model is formulated conditionally upon the random effects, with then the random-effects distribution specified, the parameters have a subject-specific interpretation. This poses difficulties when scientific research is geared towards marginal, population-averaged effects. To allow for such interpretation nevertheless, we supplement the work of Njagi et al. (2012) by a model with marginal interpretation. Focus

is on the case where a repeated continuous and a repeated time-to-event outcome are measured simultaneously (the base model referred to as JCS; the marginalized version JCS-M), as well as on the situation of a bivariate repeated time-to-event outcome (BSS and BSS-M). The marginalization is done following ideas of the so-called marginalized multilevel model (MMM) proposed by Heagerty (1999).

2 Motivating case study

The first set data are from a study with the objective to check whether the follow-up of chronic heart failure (CHF) patients, by means of a telemonitoring program, reduced mortality and re-hospitalization rates. Heart rate was longitudinally collected from 80 patients, recorded each day for a period of between 182 to 186 days. In addition, the following variables were also recorded: patient's gender, age, and heart rhythm at baseline. Our analysis of these data will be focusing on testing for a joint effect of heart rhythm on repeated time-to-hospitalization (as patients might experience multiple hospitalization) as well as on the longitudinal heart rate. The second set is a so-called comet assay. The data were collected in four groups of six male rats that received a daily oral dose of a compound in three dose levels (low, medium, high) or vehicle control. A cell suspension was prepared for each animal, from each of which three replicate samples were prepared for scoring. There were 50 randomly selected non-overlapping cells per sample, scored for DNA damage using a semi-automated scoring system. A total of 150 liver cells per animal was scored. DNA damage was assessed through the software system by measuring percentage of tail intensity and tail moment, these two responses has heavy tailed distribution and more or less similar to Weibull's. The data take the form of a multilevel structure where a cell suspension or slide, containing three replicate samples, is nested within an animal. In this paper, we target one clustering level, i.e., the slide. We also target two dose levels, low and medium.

3 Method and Estimation

3.1 Ingredients

There are three components to be involved to propose the joint models: linear mixed model (for longitudinal continuous outcomes), the combined model (for repeated, overdispersed time-to-event data), and the marginalization approach. We refer to Verbeke and Molenberghs (2000) to review the linear mixed model (LMM),

$$Y_{ij} = x'_{ij}\boldsymbol{\xi} + z'_{ij}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{ij}. \tag{1}$$

where Y_{ij} denotes the response of interest, for the ith subject, measured at time τ_{ij}, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_i$. The $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are p- and q-vectors of known covariates, with $\boldsymbol{\xi}$ a p-dimensional vector containing the fixed effects. The $\boldsymbol{b}_i$ and ε_i are assumed to be independent and distributed $\boldsymbol{b}_i \sim N(0, D)$ and $\varepsilon_i \sim N(0, \Sigma_i)$, respectively. This assumption is also applied to the rest of the paper. Meanwhile, the combined model (Molenberghs et al, 2012) for time-to-event outcomes can be the Weibull-gamma-normal model, specified as

$$\begin{aligned} Y_{ij}|\boldsymbol{b_i}, \theta_{ij} &\sim \text{Weibull}(\rho, k_{ij}), && (2)\\ k_{ij} &= \lambda\theta_{ij} e^{\widetilde{\boldsymbol{x}}'_{ij}\boldsymbol{\xi} + \widetilde{\boldsymbol{z}}'_{ij}\boldsymbol{b}_i}, && (3)\\ \theta_{ij} &\sim \text{Gamma}(\alpha, \beta), && (4) \end{aligned}$$

with Y_{ij} the time-to-event outcome of individual i at occasion j. The design vectors $\widetilde{\boldsymbol{x}}_{ij}$ and $\widetilde{\boldsymbol{z}}_{ij}$ play a role similar to their counterparts in the linear mixed model. Further, κ_{ij} is the mean function, ρ is the shape parameter, and the parametrization of the linear predictor is chosen in analogy with (1). Furthermore, regarding the marginalization, we adopt the idea of Heagerty (1999). A fully general MMM formulation is:

$$\begin{aligned} g(\mu^m_{ij}) &= \widetilde{\boldsymbol{x}}'_{ij}\boldsymbol{\xi}^m, && (5)\\ g(\mu^c_{ij}) &= \Delta_{ij} + \widetilde{\boldsymbol{z}}'_{ij}\boldsymbol{a}_i, && (6)\\ \boldsymbol{a}_i &\sim F_a(0, D), && (7)\\ Y^c_{ij} = Y_{ij}|\boldsymbol{a}_i &\sim F_{Y^c}(\mu^c_{ij}, v). && (8) \end{aligned}$$

Retaining notational conventions used so far, (5) and (6) can be seen as specifying the marginal and conditional means, respectively, thereby linking them through so-called connector function Δ_{ij}. Each outcome Y_{ij} follows an exponential family model with distribution F_{Y^c}, as specified in (8). The $g(\cdot)$ is a link function applied to both means. The function Δ_{ij} depends on the covariates, marginal parameters, and random-effects specification. It connects the marginal and conditional means and can be obtained from solving the integral equation: $g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\xi}^m) = \mu^m_{ij} = \int_a g^{-1}(\Delta_{ij} + \widetilde{\boldsymbol{z}}'_{ij}\boldsymbol{a}_i)\, dF_a$. The MMM idea applies without difficulty to the combined model, with the integral equation now becomes:

$$g^{-1}(\widetilde{\boldsymbol{x}}'_{ij}\boldsymbol{\xi}^m) = \mu^m_{ij} = \int_a \int_\theta g^{-1}(\Delta_{ij} + \widetilde{\boldsymbol{z}}'_{ij}\boldsymbol{a}_i)\, d\Theta_\theta\, dF_a.$$

For the Weibull-gamma-normal MMM, with gamma distributed overdispersion effects as in (4), the connector becomes:

$$\Delta_{ij} = -\log(\alpha\beta) + \widetilde{\boldsymbol{x}}'_{ij}\boldsymbol{\xi}^m - \widetilde{\boldsymbol{z}}'_{ij} D \widetilde{\boldsymbol{z}}_{ij}/2. \qquad (9)$$

3.2 The Proposed Joint Models

We introduce the following notation for a general Weibull model with both conjugate and normal random effects:

$$\omega(t,\lambda,\rho,\theta,\mu,b)=\lambda\rho t^{\rho-1}\theta e^{\mu+b}e^{-\lambda^{\rho}\theta e^{\mu+b}}.$$

Then, the joint distribution for the continuous and time-to-event outcomes, conditional upon the random effects is:

$$\begin{aligned} f(\boldsymbol{t}_i,\boldsymbol{y}_i|\boldsymbol{b}_i,\boldsymbol{\theta}_i) &= \prod_k \omega(t_{ik},\lambda,\rho,\theta_{ik},\mu_{ik}=\widetilde{\boldsymbol{x}}'_{ik}\boldsymbol{\xi},\widetilde{\boldsymbol{z}}'_{ik}\boldsymbol{b}_i) \\ &\times \frac{1}{(2\pi)^{\frac{n_i}{2}}|\Sigma_i|^{\frac{1}{2}}}e^{-\frac{1}{2}(\boldsymbol{y}_i-X_i\boldsymbol{\xi}-Z_i\boldsymbol{b}_i)'\Sigma_i^{-1}(\boldsymbol{y}_i-X_i\boldsymbol{\xi}-Z_i\boldsymbol{b}_i)} \end{aligned} \quad (10)$$

Here, $\Sigma_i=\sigma^2 I_{n_i}$, with I_n denoting the identity matrix of dimension n. Then, implementing the MMM requires marginalization over the Weibull model only, given that the linear mixed model contribution trivially marginalizes. This implies that the connector function (9) applies without any problem. Moreover, in the same spirit, one can consider a joint model for two repeated time-to-event sequences. The association is induced by shared normal random effects:

$$\begin{aligned} f(\boldsymbol{t}_{1i},\boldsymbol{t}_{2i}|\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{b}_i) &= \prod_j \omega(t_{1ij},\lambda_1,\rho_1,\theta_{1ij},\mu_{1ij},\boldsymbol{b}_i) \\ &\cdot\prod_k \omega(t_{2ik},\lambda_2,\rho_2,\theta_{2ik},\mu_{2ik},\gamma\boldsymbol{b}_i). \end{aligned} \quad (11)$$

The $\boldsymbol{\theta}_{1i}$ and $\boldsymbol{\theta}_{2i}$ are assumed to be independent. This process is closely related to the marginalization of a single sequence of repeated time-to-event outcomes, presented above. Also here, connector function (9) is used. Finally, regarding estimation, the fitting method of Molenberghs et al. (2010) is employed. It consists of analytically integrating the marginal form of (10) and (11) over the gamma and numerically over the normal random effects. This result that a standard software, such as the SAS procedure NLMIXED can be used to fit the model.

4 Application

With $\psi(t,\lambda,\rho,\mu,b,\alpha,\beta)=\frac{\lambda\rho t^{\rho-1}e^{\mu+b}\alpha\beta}{(\lambda t^{\rho}\beta e^{\mu+b}+1)^{\alpha+1}}$
and $\xi(C,\lambda,\rho,mu,b,\alpha)=\frac{1}{\left(\frac{\lambda C_k^{\rho}e^{\mu_k+b}}{\alpha}+1\right)^{\alpha}}$, respectively the marginal conditional density of the Weibull combined model and its form with allowing right censoring, we fit the model of

$$\begin{aligned} f(y_{ij},t_{ik}|b_i) &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}[y_{ij}-(\beta_0+\beta_1x_i+\beta_2\tau_{ij}+\beta_3x_i\tau_{ij}+b_i)]^2} \\ &\cdot\xi(C_{ik},\lambda,\rho,\mu_{ik},\gamma b_i,\alpha). \end{aligned}$$

to analyze the chronic heart failure data, and the model of

$$f(t_{1ij}, t_{2ij}|b_i) = \psi(t_{1ij}, \lambda_1, \rho_1, \mu_{1ij}, b_i, \alpha_1) \cdot \psi(t_{2ij}, \lambda_2, \rho_2, \mu_{2ij}, \gamma b_i, \alpha_2).$$

employed for the comet assay data. The analysis results can be seen in Table 1. In the comet analysis, we observe similar point estimates and precision for both the BSS and the BSS-M. There is statistically significant evidence that the dose level has an effect on the hazard of the tail intensity and of the tail moment in both models (conditional and marginal). Direct marginal interpretation is possible. The shared parameter's presence is statistically significant, indicating that the two survival processes are correlated. Also here, the likelihood ratios are similar. Turning attention to the heart failure analysis, a few observations are in place. First, estimates of the two models are similar. Second, there is no statistically significant evidence that heart rhythm has an effect on the evolution of heart rate, both in the joint model and its marginalized one. In contrast, however, there is no significant effect of heart rhythm on the hazard of time-to-hospitalization in the marginalized joint model whereas this is not true in the joint model. This is important and requires careful qualification. Third, the shared estimate is statistically significant, pointing to non-negligible correlation between the continuous and survival processes. Overall, such a result should not be treated as problematic, but rather as resulting from genuine differences in parameter interpretation between the marginal and conditional formulations.

5 Concluding Remarks

Our work builds upon and extends work of Molenberghs et al.(2010), Molenberghs et al. (2012), Njagi et al. (2012), and Heagerty (1999), bringing in additional features e.g. the model can be marginalized in the sense of carrying marginal parametric regression functions that have a population-averaged interpretation; and the time-to-event outcomes are allowed to be right censored. Furthermore, even though the model is relatively complex in the sense that it extends and amends a conventional generalized linear mixed model in various ways, the marginalization using a so-called connector function on the one hand and the numerical technique of partial marginalization, renders the model relatively easy to fit, through standard statistical GLMM software, with minimal additional programming. While focus has been placed on bivariate longitudinal sequences, the methodology could be extended without trouble to more than two outcomes. Additionally, left-censoring and even interval censoring could be considered as well. For conciseness, this has not been made explicit here.

TABLE 1. The Chronic Heart Failure Data (With censoring) and The Comet Data. 'JCS' refers to joint continuous survival model; 'BSS' refers to the bivariate survival model; 'M' and 'Cens' means marginalized and with censoring, respectively.

Par.	JCS-Cens	JCS-Cens-M	Par.	BSS	BSS-M
	Est.(s.e.)	Est.(s.e.)		Est.(s.e.)	Est.(s.e.)
	Longitudinal process			*The first survival process*	
β_0	3.4683(0.2393)	3.6728(0.0852)	ξ_1	-3.3509(0.1109)	-3.3521(0.1109)
β_1	-0.1853(0.3543)	-0.1487(0.1140)	λ_1	2.7610(0.2343)	2.8741(0.2468)
β_2	-0.0003(0.0001)	-0.0004(0.0001)	α_1	9.7570(2.3903)	9.7713(2.4071)
β_3	-0.0003(0.0002)	-0.0002(0.0002)	${\sigma_1}^2$	0.0773(0.0236)	0.0770(0.0234)
σ^2	0.1530(0.0021)	0.1531(0.0021)			
	Survival process			*The second survival process*	
ξ	-0.1812(0.0438)	-0.3724(0.3233)	ξ_2	-2.4161(0.0911)	-2.4167(0.0910)
λ	0.0018(0.0024)	0.0047(0.0012)	λ_2	0.2351(0.0158)	0.2411(0.0164)
α	10.241(3.5079)	5.1688(6.7443)	α_2	52.870(40.013)	52.871(39.968)
${\sigma_b}^2$	3.9922(9.1030)	0.1821(0.1413)	${\sigma_2}^2$	0.0391(0.0142)	0.0393(0.0146)
γ	0.3775(0.1740)	1.1670(0.4422)	γ	0.7958(0.1161)	0.7958(0.1161)
-2LL	11745	11688		19624	19624

References

Heagerty, P.J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**, 688 – 698.

Molenberghs, G., Verbeke, G., Demetrio C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325 – 347.

Molenberghs, G., Verbeke, G., Efendi, A., Braekers, R.,and Demétrio, C.G.B. (2012). A combined gamma frailty and normal random-effects model for repeated, overdispersed time-to-event data. *Submitted for publication.*

Njagi, E. N., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2012). A flexible joint-modeling framework for longitudinal and time-to-event data with overdispersion. *Submitted for publication.*

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