84. A SHARED-PARAMETER JOINT MODEL FOR PROSTATE CANCER RISK AND PSA LONGITUDINAL PROFILES

C. Serrat
Universitat Politècnica de Catalunya-BarcelonaTECH, Dept. of Applied Mathematics I, Spain
carles.serrat@upc.edu

M. Rué, H. Perpiñán and A. Forte
Universitat de Lleida, Dept. of Basic Medical Sciences, Spain
montse.rue@cmb.udl.cat

C. Armero
Universitat de València, Dept. of Statistics and OR, Spain,
{carmen.armero, hector.perpinan, anabel.forte}@uv.es

X. Piulachs
Universitat de Barcelona, Dept. of Econometrics, Spain,
xavierpiulachs@hotmail.com

Á. Páez
Hospital Universitario de Fuenlabrada, Urology Dept., Spain,
alvaro.paez@salud.madrid.org

G. Gómez
Universitat Politècnica de Catalunya-BarcelonaTECH, Dept. of Statistics and OR, Spain,
lupe.gomez@upc.edu

ABSTRACT

The paper describes the use of frequentist and Bayesian shared-parameter joint models of longitudinal measurements of prostate specific antigen (PSA) and the risk of prostate cancer (PCa). The motivating dataset corresponds to the screening arm of the Spanish branch of the European Randomized Screening for Prostate Cancer (ERSPC) study. The results show that PSA is highly associated with the risk of being diagnosed with PCa and that there is an age-varying effect of PSA on PCa risk. Both the frequentist and Bayesian paradigms produced very close parameter estimates and subsequent 95% confidence and credibility intervals. Dynamic estimations of disease-free probabilities obtained using Bayesian inference highlight the potential of joint models to guide personalized risk-based screening strategies.

1. INTRODUCTION

Joint models for longitudinal and time-to-event data are increasingly used to assess relationships between serial measurements of one or several markers and time to an event of interest. Joint models were introduced during the 90s (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997) and since then have been applied to a great variety of studies in epidemiological and biomedical areas. In turn, these studies have fed a wide methodological research on the subject, with models focused on event times, longitudinal patterns, or both (Neuhaus *et al.* (2009) and Tsiatis and Davidian (2004) are excellent reviews up to date).

In a setting of dependent longitudinal and time-to-event data, shared-parameter models consider that the longitudinal and survival processes depend jointly on a common set of random effects. Given the random effects, the two processes are assumed independent. Shared-parameter models allow one to quantify the effect of the underlying longitudinal outcome on the risk of an event, and obtain individualized time-dynamic predictions. Recently, Rizopoulos has made a great contribution facilitating the use of the joint modeling methodology, first by means of an overview of the theory and applications of joint modeling (Rizopoulos, 2012) and secondly by developing the `JM` (Rizopoulos, 2010) and `JMbayes` (Rizopoulos, 2013) `R` packages for the frequentist and Bayesian shared-effects' approaches, respectively.

	Number of visits per subject				
	1	2	3	≥ 4	Overall
Sample size (%)	573 (23.7)	1499 (62.1)	293 (12.1)	50 (2.1)	2415 (100.0)
PSA descriptive					
Mean	2.40	1.40	1.92	5.46	1.78
StDev	5.67	1.67	2.25	2.79	3.20
Min	0.00	0.00	0.00	0.50	0.00
1st Q	0.60	0.60	0.60	3.79	0.63
Median	1.10	1.00	1.10	4.69	1.06
3rd Q	2.20	1.67	2.44	6.28	1.95
Max	68.90	42.30	20.36	15.53	68.90
PCa diagnosed (%)	51 (44.0)	50 (43.1)	13 (11.2)	2 (1.7)	116 (100.0)

Table 1. Distribution of the subjects in the screening arm of the Spanish ERSPC study, descriptive statistics of the PSA measurements and distribution of the PCa diagnosed cases, stratified by the number of visits and overall.

Proust-Lima and Taylor (2009) assessed the prognostic value of longitudinal PSA measurements as a marker of PCa progression, in patients treated with radiotherapy. Joint modeling is an interesting approach that allows to illustrate how classical and Bayesian statistics can be combined to reach a complex goal, taking advantage of the strengths of both paradigms. The goal of the paper is to show, from both approaches, how shared-parameter joint models can be used to incorporate past and current PSA levels into a predictive model of PCa, in a screening setting.

2. ERSPC STUDY AND DATA DESCRIPTION

Our motivating dataset consists of the screening arm of the Spanish subset of the European Randomized Screening for Prostate Cancer (ERSPC) study (Luján *et al.*, 2012). The ERSPC study is a multicenter trial initiated in 1991 in the Netherlands and Belgium, with six more European countries joining during the 90s. The objective of the ERSPC was to evaluate whether PCa screening with PSA reduces PCa specific mortality in asymptomatic men. The Spanish subset recruited 4278 individuals between February 1996 and June 1999, aged 45 to 71 years, of whom 2415 were assigned to the screening arm and the remaining to the control arm. Those screened were given a blood test to measure PSA levels. Table 1 summarizes the description of the data.

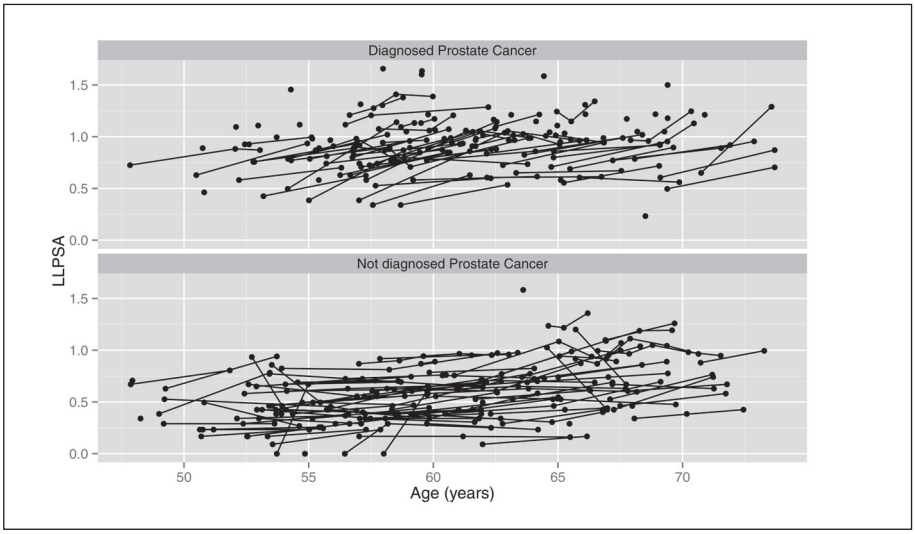

Figure 1. *LLPSA* profiles versus time (age in years) for men who developed PCa during follow-up (top) and for 116 randomly selected men without a PCa diagnosis (bottom), participants in the screening arm of the Spanish ERSPC study.

We defined the time-to-event, T, as the time elapsed from the protocol screening start (age 45) to diagnosis of prostate cancer. We estimated that more than 90% of patients were prostate cancer free at age 80 years. In particular 116 men (4.8%) were diagnosed with PCa (median time, over age 45, equal to 17.32 years) and the remaining 2,299 men (95.2%) contributed to the study as right-censored data.

3. MODEL PROPOSAL

For the i-th subject, $i = 1, \ldots, n$, denote by $m_i(t)$ the true $LLPSA$ value at time t (i.e. at age $45 + t$) and by $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s \leq t\}$ the whole longitudinal history of the true marker levels up to time t. The observable data for the i-th subject at the specific occasions t_{ij} at which measurements were taken, consists of the observed longitudinal $LLPSA$ profile $\{y_{ij} = LLPSA_{ij} = \log(1 + \log(1 + PSA_{ij})), j = 1, \ldots, n_i\}$. Let $y_i(t)$ be the hypothetical observed value of $LLPSA$ at time t which is assumed to deviate from the true value $m_i(t)$ by a certain error measurement $\varepsilon_i(t)$ in such a way that $y_i(t) = m_i(t) + \varepsilon_i(t)$ holds.

The joint model that we propose binds a linear mixed-effects model for the longitudinal part and a relative risk survival model for the time to PCa diagnosis, which incorporates the true historical subject profile of the longitudinal process (Rizopoulos, 2012). The survival submodel is based on the Weibull distribution due to its flexibility for representing different types of risks. To the best of our knowledge, this is the first time that a straightforward shared-parameter joint model of PSA and age at PCa diagnosis is fitted, in a screening setting, with the objective of exploring if longitudinal PSA measurements may orientate a personalized risk-based screening. The joint model is specified as

$$\begin{cases} (y_{ij} \mid \boldsymbol{\theta}_i) = m_i(t_{ij}) + \varepsilon_{ij} = \beta_0 + b_{i0} + (\beta_1 + b_{i1})t_{ij} + \varepsilon_{ij} \\ h_i(t|\mathcal{M}_i(t), w_{1i}, \boldsymbol{\theta}_i) = h_0(t \mid \lambda)\exp\{\gamma_0 + \gamma_1 w_{1i} + \alpha m_i(t)\}, \end{cases} \quad (1)$$

where $h_i(t) = \lim_{dt\to 0} \Pr(t \le T_i^* < t + dt \mid T_i^* \ge t)/dt$ denotes the hazard function, T^* denotes the true event time, and

$$\boldsymbol{\theta}_i = (\beta_0, \beta_1, b_{i0}, b_{i1}, \sigma_{b_0}, \sigma_{b_1}, \rho_{b_0 b_1}, \sigma, \lambda, \gamma_0, \gamma_1, \alpha)^{\mathsf{T}}$$

is the vector containing all parameters, random effects and hyperparameters associated with the individual i. For the longitudinal submodel, $(\beta_0, \beta_1)^{\mathrm{T}}$ and $(b_{i0}, b_{i1})^{\mathrm{T}}$ are the fixed and the random effects for the intercept and the slope term, respectively, and ε_{ij} the error term for the i-th subject at the j-th measurement with $(\varepsilon ij \mid \sigma) \sim \mathcal{N}(0, \sigma^2)$; furthermore we assume that both random effects given σ_{b0}, σ_{b1} and their correlation coefficient, $\rho_{b0\,b1}$, follow a bivariate normal distribution with a mean of zero and an unstructured co variance matrix. For the survival submodel, the baseline hazard is given by $h_0(t \mid \lambda) = \lambda\, t^{\lambda-1}$, where λ is the shape parameter; $w_1 = a_1 \times y_1$ is a covariate which describes the interaction between the age, a_1, and the LLP SA measurement, y_1, at first visit, respectively; $(\gamma_0, \gamma_1)^{\mathrm{T}}$ are the regression coefficients for the baseline hazard and the w_1 interaction. Finally, α stands for the association parameter assessing the relationship between the longitudinal and the survival submodels.

Since preliminary analysis for joint model (1) showed a large correlation between b_{i0} and b_{i1} we discarded the slope random effect and restrict posterior analyses, both frequentist and Bayesian, interpretations and conclusions to the following model

$$h_i(t|\mathcal{M}_i(t), w_{1i}, \boldsymbol{\theta}_i) = h_0(t \mid \lambda)\exp\{\gamma_0 + \gamma_1 w_{1i} + \alpha(\beta_0 + b_{i0} + \beta_1 t)\}, \quad (2)$$

where $(b_{i0} \mid \sigma_{b_0}) \sim \mathcal{N}(0, \sigma_{b_0}^2)$.

4. ESTIMATION PROCEDURE, RESULTS AND INTERPRETATION

In the frequentist approach, the joint model expression in (2) was fitted through the `jointModel` function, implemented in the `R` package `JM` (Table 2). The estimates reflect a strong positive association between PSA values and risk of PCa, and allows estimation of the impact of different increments of PSA at specific PSA values.

Parameters	Value	95% Conf. Int.
Longitudinal Submodel		
β_0	0.369	(0.351, 0.388)
β_1	0.014	(0.013, 0.015)
σ_2	0.011	(0.011, 0.012)
σ^2_{b0}	0.050	(0.044, 0.057)
Survival Submode		
γ_0	-13.706	(-15.299, -12.113)
γ_1	-0.068	(-0.107, -0.028
λ	1.887	(1.459, 2.315)
Association		
α	7.207	(6.041, 8.372)

Table 2. Joint model estimates for frequentist analysis of PCa diagnosis and longitudinal PSA values.

Similar results are obtained when estimating the full vector of uncertainties in the model, $\boldsymbol{\theta} = \cup_{i=1}^{n} \boldsymbol{\theta}_i$, from a Bayesian perspective, by using non-informative distributions as a priors and the common prior distribution given by

$$\pi(\boldsymbol{\theta}) = \pi(\lambda)\,\pi(\gamma_0)\,\pi(\gamma_1)\,\pi(\alpha)\;\pi(\beta_0)\,\pi(\beta_1)\,\pi(\mathbf{b}_0 \mid \sigma_{b_0})\,\pi(\sigma_{b_0})\,\pi(\sigma).$$

Computations were carried out using the statistical software `WinBUGS` (Lunn *et al.*, 2000) and `R`-package `JMbayes`.

The proposed methodology allows to dynamically estimate disease-free probabilities. Figure 2 is an illustration of these updated survival probabilities based on the observed longitudinal subject-specific marker profile available up to date.

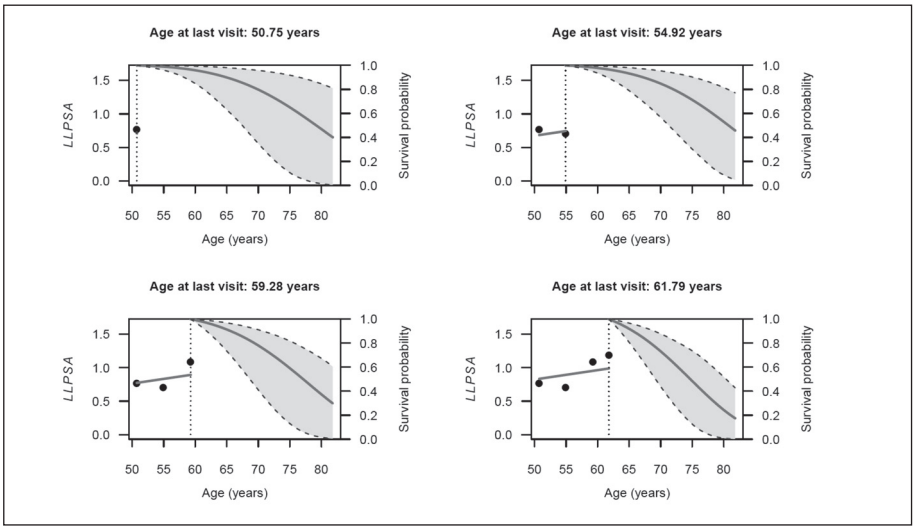

Figure 2. *LLPSA* longitudinal trajectories and dynamic PCa free survival probabilities (with 95% pointwise credible intervals at each time point) for a patient after their first to fourth successive PSA measurements.

5. DISCUSSION

In this study, using data from the Spanish arm of the ERSPC study, we have estimated a joint model of longitudinal measurements of PSA and time to diagnosis of PCa. Our model overcomes the limitations of simpler statistical tools because it accounts for a) the effect of PSA as an endogenous time-dependent covariate measured with error and b) the non-random dropout that results when an individual is diagnosed with PCa. In particular, we used both the frequentist and the Bayesian approaches, and built a shared-parameter joint model that combined some baseline covariates and longitudinal PSA values to predict PCa incidence. The resulting model is consistent with the literature and clinical knowledge. This work can be considered a first step that should be followed by a more comprehensive modeling process accounting for additional predictive factors measured over time, e.g. prostate volume or previous biopsy status (Roobol *et al.*, 2012).

ACKNOWLEDGMENTS

This paper has been partially supported by research grants MTM2012-38067-C02-01 and MTM2013-42323-P from the Spanish Ministry of Economy and Competitiveness and 2014 SGR 464 from the Departament d'Economia i Coneixement de la Generalitat de Catalunya.

REFERENCES

Faucett, C.L., and Thomas, D.C. (1996). "Simultaneously Modelling Censored Survival Data and Repeatedly Measured Covariates: A Gibbs Sampling Approach". *Statistics in Medicine*, 15, 1663-1685.

Lujan, M., Paez, A., Berenguer, A., and Rodriguez, J.A. (2012). "Mortality Due to Prostate Cancer in the Spanish Arm of the European Randomized Study of Screening for Prostate Cancer (ERSPC). Results after a 15-Year Follow-Up". *Actas Urológicas Españolas*, 36, 403-409.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). "WinBUGS - A Bayesian Modelling Framework: Concepts, Structure, and Extensibility". *Statistics and Computing*, 10, 325-337.

Neuhaus, A., Augustin, T., Heumann, C., and Daume, M. (2009). "A Review on Joint Models in Biometrical Research". *Journal of Statistical Theory and Practice*, 3, 855-868.

Proust-Lima, C., and Taylor, J.M. (2009). "Development and Validation of a Dynamic Prognostic Tool for Prostate Cancer Recurrence Using Repeated Measures of Posttreatment PSA: A Joint Modeling Approach". *Biostatistics*, 10, 535-549.

Rizopoulos, D. (2010). JM: "An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data". *Journal of Statistical Software*, 35, 1-33.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*. CRC Press, Boca Ratón, FL.

Rizopoulos, D. (2013). *JMbayes: Shared Parameter Models for the Joint Modeling of Longitudinal and Time-to-Event Data Using JAGS, WinBUGS, or OpenBUGS*. R package 0.4-1 http://cran.r-project.org/package=JMbayes/.

Roobol, M.J., Van Vugt, H.A., Loeb, S., Zhu, X., Bul, M., Bangma, C.H., Van Leenders, A.G., Steyerberg, E.W., and Schroder, F.H. (2012). "Prediction of Prostate Cancer Risk: The Role of Prostate Volume and Digital Rectal Examination in the ERSPC Risk Calculators". *European Urology*, 61, 577-583.

Tsiatis, A.A., and Davidian, M. (2004). "Joint Modeling of Longitudinal and Time-to-Event Data: An Overview". *Statistica Sinica*, 14, 809-834.

Wulfsohn, M.S., and Tsiatis, A.A. (1997). "A Joint Model for Survival and Longitudinal Data Measured with Error". *Biometrics*, 53, 330-339.