Shared random-effect models for the joint analysis of longitudinal and time-to-event data: application to the prediction of prostate cancer recurrence

Mbéry Sène, Carine Bellera, Cecile Proust-Lima

Shared random-effect models for the joint analysis of longitudinal and time-to-event data: application to the prediction of prostate cancer recurrence

Mbéry Sène[1,2,3*], Carine A. Bellera[3,4] and Cécile Proust-Lima[1,2,3]

October 4, 2013

[1] INSERM, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, F-33000 Bordeaux, France

[2] Univ. Bordeaux, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, F-33000 Bordeaux, France

[3] INSERM, ISPED, CIC-EC7, F-33000 Bordeaux, France

[4] Department of Clinical Epidemiology and Clinical Research, Institut Bergonié, Regional Comprehensive Cancer Center, Bordeaux, France.

Correspondence to: Mbéry Sène, Mbery.Sene@isped.u-bordeaux2.fr

Abstract: In the last decade, joint modeling research has expanded very rapidly in biostatistics and medical research. This type of models enables the simultaneous study of a longitudinal marker and a correlated time-to-event. Among them, the shared random-effect models that define a mixed model for the longitudinal marker and a survival model for the time-to-event including characteristics of the mixed model as covariates received the main interest. Indeed, they extend naturally the survival model with time-dependent covariates and offer a flexible framework to explore the link between a longitudinal biomarker and a risk of event.

The objective of this paper is to briefly review the shared random-effect model methodology and detail its implementation and evaluation through a real example from the study of prostate cancer progression after a radiation therapy. In particular, different specifications of the dependency between the longitudinal biomarker, the prostate-specific antigen (PSA), and the risk of clinical recurrence are investigated to better understand the link between the PSA dynamics and the risk of clinical recurrence. These different joint models are compared in terms of goodness-of-fit and adequation to the joint model assumptions but also in terms of predictive accuracy using the expected prognostic cross-entropy. Indeed, in addition to better understand the link between the PSA dynamics and the risk of clinical recurrence, the perspective in prostate cancer studies is to provide dynamic prognostic tools of clinical recurrence based on the biomarker history.

Keywords: Joint models, Shared random-effect models, Dynamic predictions, Prognostic cross-entropy, Predictive accuracy, Prostate cancer.

Résumé: Dans la dernière décennie, la recherche en modélisation conjointe s'est développée très rapidement dans le domaine des biostatistiques et de la recherche médicale. Ce type de modèles permet d'étudier simultanément un marqueur longitudinal et un temps d'événement corrélés. Parmi eux, les modèles à effets aléatoires partagés, qui définissent un modèle mixte pour le marqueur longitudinal et un modèle de survie pour le temps d'événement incluant les caractéristiques du modèle mixte comme variables explicatives, ont reu le plus d'attention. En effet, ces modèles étendent naturellement le modèle de survie avec variables explicatives dépendantes du temps et offrent un cadre flexible pour explorer le lien entre le biomarqueur longitudinal et le risque d'événement.

L'objectif de cet article est de passer brièvement en revue la méthodologie du modèle à effets aléatoires partagés et de détailler son implémentation et son évaluation à travers un exemple réel d'étude de progression de cancer de la prostate après une radiothérapie. En particulier, différentes spécifications de la dépendance entre le biomarqueur longitudinal, l'antigène spécifique de la prostate (PSA), et le risque de rechute clinique sont investiguées pour bien comprendre le lien entre la dynamique du PSA et le risque de rechute clinique. Ces différents modèles conjoints sont comparés en termes de qualité d'ajustement et d'adéquation aux hypothèses du modèle conjoint mais aussi en termes de pouvoir prédictif en utilisant la cross-entropie pronostique. En effet, en plus de mieux comprendre le lien entre la dynamique de PSA et le risque de rechute clinique, la perspective dans les études sur le cancer de la prostate est de fournir des outils pronostiques dynamiques de rechute clinique basés sur toute l'histoire du biomarqueur.

Mots clés: Modèles conjoints, Modèles à effets aléatoires partagés, Prédictions dynamiques, Cross-entropie pronostique, Pouvoir prédictif, Cancer de la prostate.

1 Introduction

In cohort studies, time-to-event data and repeated measures of biomarkers data are often collected for each subject. There exist standard methods for separately analyzing these data; mixed models[16] are used to describe the biomarker repeated measures and assess the association with covariates while survival models, mainly proportional hazard models[3], are used to evaluate the risk of the event and the effect of its predictors. These approaches are however not optimal. First, these processes are usually dependent so that separated analyses may lead to biased inference, and second, interest is frequently in the understanding of their association so that the two processes need to be modeled simultaneously.

To assess the relationship between a longitudinal biomarker and a time-to-event, an intuitive model would be the proportional hazard Cox model[3] with the biomarker as a standard time-dependent covariate. But this technique is not adapted with biomarker repeated data. Indeed, first the Cox model assumes that the time-dependent covariate is external (or exogenous) which means that the value of the covariate at time t is not affected by the occurrence of an event at time u, with $t > u$[15]. This is not the case with biomarkers that constitute a typical example of internal (or endogenous) time-dependent covariate. Second, the Cox model assumes that the time-dependent covariate is observed at all the event times, which is most always not the case in practice.To solve this problem, missing values at the event times can be imputed using the LOCF (Last Observation Carried Forward) method that replaces an unobserved value at an event time by the most recent observed value. However, this method can strongly bias the estimates[31], especially with irregularly and infrequent measures of the biomarker. Finally, the standard Cox model assumes that the time-dependent covariate is measured without any error where as the observed biomarker is noisy measure.

To remedy this, joint models have been developed. They aim at modeling simulta-

neously the longitudinal biomarker and time-to-event processes, and at characterizing their relationship. As a consequence, they provide an interesting framework to (i) evaluate the longitudinal biomarker trajectory and its association with covariates without the bias introduced by the informative time-to-event, (ii) evaluate the risk of event and its association with covariates, including the longitudinal biomarker data and (iii) directly explore the association between the longitudinal and survival processes. The principle of joint models is to model the biomarker repeated measures using a mixed model, to model the risk of event using a survival model and to link the two models using a common latent structure[5–9–35]. This common latent structure captures the association between the processes so that the two processes are conditionally independent given the latent structure.

There exist two kinds of joint models for longitudinal and time-to-event data: the shared random-effect models (SREM) and the joint latent class models (JLCM). The SREM extends directly the idea of the survival model with time-dependent covariates by considering as covariates some characteristics of the mixed model defined for the longitudinal biomarker[5–9–13–14–30–31–34–35]. These characteristics are functions of the individual random-effects of the mixed model that capture the individual deviations to the mean trajectory of the biomarker. The most classical example of shared characteristic is the individual current error-free level of the biomarker predicted from the mixed model that is included as a time-dependent covariate in the survival model[5–35]. Another example is to include in the survival model directly the vector of random-effects. The alternative, the JLCM, relies on a different idea. It assumes that the population is heterogeneous and can be divided in a finite number of homogeneous subgroups (or classes), each class being characterized by a specific trajectory of the biomarker and a specific risk of event[19–21]. The latent class structure can be seen as a latent stratification that entirely captures the dependency between the two processes. See Proust-Lima et al.[20] for a review of JLCM.

In the last decade, joint modeling research has expanded very rapidly in biostatistics and medical research with a preference for the SREM approach. Indeed, it extends naturally the survival model with time-dependent covariates and offers a flexible framework to explore the link between a longitudinal biomarker and a risk of event. Besides, it allows to improve the efficiency in assessment of treatment effects and other prognostic factors[13]. Different extensions of these models have been developed; with multiple events[11], or multiple longitudinal data[27], or multivariate longitudinal and multivariate survival data[1], or with cured fraction[17–36–37], or in presence of data clustering with common frailty[23], or with a proportional cumulative hazard time-to-event submodel instead of standard proportional hazard[4].

In addition to evaluating the association between a longitudinal and event process, joint models (JLCM and SREM) have also allowed very recently the development of a new type of prediction tools that incorporate the repeated biomarker measures to predict the risk of event[20–21–25]. Because of the longitudinal nature of the biomarker, these predictive tools are dynamic: whatever the time s, they provide the risk of event from s given all the information collected until s so that, in practice, the dynamic prediction can be systematically updated when new information is collected. As a consequence, these dynamic predictive tools potentially provide more accurate predictions than standard prognostic tools based only on information at baseline. Dynamic prognostic tools are of particular interest in chronic diseases such as cancers for predicting a clinical event (recurrence, death,...) where in addition to standard prognostic factors, repeated measures of a biomarker are collected for monitoring the patients[21–37].

In this context, the aim of this paper is to review briefly the SREM methodology and detail its implementation and evaluation through a real example from the study of prostate cancer progression after a radiation therapy. In particular, different specifications of the dependency between the longitudinal biomarker, the prostate-specific antigen

(PSA), and the risk of clinical recurrence are investigated to better understand the link between the PSA dynamics and the risk of clinical recurrence. These different joint models are compared in terms of goodness-of-fit and adequation to the joint model assumptions but also in terms of predictive accuracy. Indeed, in addition to better understand the link between the PSA dynamics and the risk of event, the perspective in prostate cancer studies is to provide powerful dynamic prognostic tools of clinical recurrence based on the biomarker history.

The paper is organized as follows. Section 2 presents the SREM model and its estimation. Section 3 focuses on dynamic prognostic tools derived from these models and their validation using a measure of predictive accuracy. The method is applied to prostate cancer progression in section 4. Finally, section 5 concludes.

2 Shared random-effect models

For subject i, $i = 1, ..., N$, we note T_i^* the time to the event of interest and C_i the censoring time. We observe the time $T_i = \min(T_i^*, C_i)$ and $E_i = \mathbb{1}_{\{T_i^* \leq C_i\}}$, the indicator of event. We also observe $Y_i = (Y_i(t_{i1}), ..., Y_i(t_{in_i}))$, the n_i-vector of repeated measures of the biomarker collected intermittently at times $(t_{i1}, ..., t_{in_i})$ until the observed time T_i. In general, covariates are denoted by X. For clarity, we distinguish in the following X_{Li} the matrix of (possibly time-dependent) covariates involved in the longitudinal submodel and X_{Si} the vector of time-independent covariates involved in the survival submodel.

2.1 Longitudinal submodel

Repeated measures of the biomarker are analyzed by a linear mixed model[16]. Specifically, we assume that the repeated measures $Y_i(t_{ij})$ are noisy measures of $Y_i^*(t_{ij})$ the true unobserved biomarker value for $j = 1, ..., n_i$. We model the mean change over time of

$Y_i^*(t_{ij})$ by taking into account the correlation within the biomarker repeated measures of a same subject:

$$\begin{aligned} Y_i(t_{ij}) &= Y_i^*(t_{ij}) + \epsilon_i(t_{ij}) \\ &= X_{Li}(t_{ij})^T \beta + Z_i(t_{ij})^T b_i + \epsilon_i(t_{ij}) \end{aligned} \quad (1)$$

where $X_{Li}(t_{ij})$ and $Z_i(t_{ij})$ are p-vector and q-vector of time-dependent covariates associated respectively with the p-vector of fixed effects β and the q-vector of Gaussian random effects b_i with mean $\mathbf{0}$ and variance-covariance matrix B. In the following, we will also use X_{Li} and Z_i the design matrices with respectively $X_{Li}(t_{ij})^T$ and $Z_i(t_{ij})^T$ as row vectors j ($j = 1, ..., n_i$). In (1), the fixed part $X_{Li}\,\beta$ represents the mean trajectory of the biomarker over time while the random part $Z_i\,b_i$ defines the individual deviation relative to the mean trajectory. The vector of errors of measurements is $\epsilon_i = (\epsilon_i(t_{i1}), \cdots, \epsilon_i(t_{in_i}))^T$. We assume ϵ_i are independent and follow a multivariate Gaussian distribution of mean $\mathbf{0}$ and diagonal variance-covariance matrix $\Sigma_i = \sigma^2 I_{n_i}$; ϵ_i and b_i are independent. In addition to the independent errors, autocorrelated errors could be included to further take into account the correlation between the repeated measures for a same subject. For example, Henderson et al.[9] considered a stationary Gaussian process with mean zero while Wang and Taylor considered[33] an integrated Ornstein-Uhlenbeck stochastic process. Originally, Tsiatis et al.[32], Faucett and Thomas[5] and Wulfsohn and Tsiatis[35] used a linear mixed model with only a random intercept and a random slope but any functions of time can be considered in $X_{Li}(t_{ij})$ and $Z_i(t_{ij})$ to capture at best the trajectory of the biomarker.

2.2 Survival submodel

The risk of event could be modelled using any survival model but in practice, proportional hazard models are mostly considered and defined as follows:

$$\lambda_i\left(t|X_{Si},b_i\right)=\lambda_0\left(t\right)e^{X_{Si}^T\gamma+h(b_i,t)^T\eta} \tag{2}$$

where $\lambda_0\left(t\right)$ is the baseline hazard function, γ is the r-vector of coefficients defining the association between the r-vector of covariates X_{Si} (that can be time dependent) and the survival time, and $h(b_i,t)$ is a multivariate function of the random-effects b_i defined in (1) and associated with the vector of parameters η. The coefficients η measure the association between the longitudinal and survival processes while $h(b_i,t)$ defines the nature of the dependence between the two processes. Different functions $h(b_i,t)$ can be defined. Usual examples include the current true level of the biomarker $(h(b_i,t)=Y_i^*(t))$[4–5–13–35], the current slope of the biomarker $(h(b_i,t)=\partial Y_i^*(t)/\partial t)$, both $(h(b_i,t)=(Y_i^*(t),\partial Y_i^*(t)/\partial t)^T)$[37] or directly the individual deviations $(h(b_i,t)=b_i)$[14]. In the application, we will go back on these functions and detail more specific ones in the context of prostate cancer progression.

Other types of dependency can be assumed. For example, instead of considering strictly random-effects shared between the two submodels, Henderson et al.[9] considered two correlated zero-mean Gaussian processes, one for each submodel with the assumption that the correlation between the longitudinal and the survival processes were entirely captured by the correlation between the two zero-mean Gaussian processes.

In the present paper, we chose to not detail further this option and concentrate on models where the dependency between the submodels is captured by standard random-effects and functions of them.

2.3 Maximum likelihood estimation

Shared random-effect models can be estimated within the maximum likelihood framework by considering the joint likelihood from the two submodels.

Let θ be the whole vector of parameters defined in (1) and (2). Using the assumption of independence between the longitudinal and the survival processes conditionally to the random effects, the joint log-likelihood of the observed data is:

$$\begin{aligned} l(\theta) &= \log\left[\prod_{i=1}^{N}\left(\int_{b_i} f_Y\left(Y_i|X_{Li},b_i;\theta\right)\ f_T\left(T_i|X_{Si},b_i;\theta\right)\ f_b\left(b_i;\theta\right)\ db_i\right)\right] \\ &= \sum_{i=1}^{N}\log\left(\int_{b_i} f_Y\left(Y_i|X_{Li},b_i;\theta\right)\ \lambda_i\left(T_i|X_{Si},b_i;\theta\right)^{E_i}\ S_i\left(T_i|X_{Si},b_i;\theta\right)\ f_b\left(b_i;\theta\right)\ db_i\right) \end{aligned} \tag{3}$$

where f_b and f_Y are multivariate Gaussian density functions of b and Y with respectively mean 0 and $X_{Li}\,\beta + Z_i\,b_i$, and variance-covariance matrix B and Σ_i; $\lambda_i(T_i|X_{Si},b_i;\theta)$ is the hazard function defined in (2) and taken at the observed time T_i and $S_i(T_i|X_{Si},b_i;\theta) = e^{-\int_0^{T_i}\lambda_i(t|X_{Si},b_i;\theta)dt}$ is the derived survival function.

From this, the maximum likelihood estimates can be obtained by iterative algorithms such as EM or Newton-Raphson algorithm[9–24–28–35]. This estimator has good asymptotic properties as shown by Zeng and Cai[38]. We note that other authors used a bayesian approach to estimate these joint models[5–8–12–33].

Equation (3) involves two integrals that do not have analytic solutions. They are usually approximated by numerical integration with Gauss-Hermite and Gauss-Kronrod quadratures for the integrals over respectively the random effects and the survival function[9–24–35]. The numerical approximations of the integrals, mostly the Gauss-Hermite one for the random effects, slow the calculations. Indeed, the integral over the random-effects is usually multidimensional (size q). When q remains small (less than 3), the

Gauss-Hermite quadrature remains the standard method but in higher dimension settings, alternative methods may be preferred to reduce the computational time like the Laplace method as proposed by Rizopoulos et al.[28] or a Monte Carlo method. We note that the structure of B does not intervene in the computational complexity, only the number q of random-effects is limiting.

To improve the accuracy and reduce the number of nodes in Gaussian quadratures, adaptive versions have been proposed. They consist in centering and rescaling the integral[18] around its modal value, so that the nodes are systematically placed at the optimum position. This technique remains however time consuming as it requires a subject-and-iteration-specific optimization to define the optimum position. To retain the same precision but simplify the numerical aspect, Rizopoulos in[26] proposed a pseudo-adaptive version where the integral is centered and rescaled according to the posterior distribution of the random-effects defined in the linear mixed model in (1) but estimated separately and once for all in a first step.

There exist two R packages to estimate SREM: joineR (http://cran.r-project.org/web/packages/joineR) and JM[24]. In this work, we used directly the JM program online, or modified the JM source code when necessary to include specific functions $h(b_i, t)$. We used a Newton-Raphson algorithm for the maximization of the likelihood and a pseudo-adaptive Gaussian quadrature rule with nine points to approximate the integral over the random effects in (3)[26]. The variance-covariance matrix of the parameters $\widehat{\mathrm{V}(\hat{\theta})}$ was estimated by the inverse of the Hessian matrix.

3 Dynamic predictions

3.1 Individual prediction

Individual dynamic predictions of event can be derived from any joint model. They consist in the individual predicted probability of event between times s and $s+t$ given the biomarker data $Y_i^{(s)} = \{Y_i(t_{ij}),\, j = 1, ..., n_i, \text{such as } t_{ij} \leq s\}$ collected until time s and the matrix of explanatory covariates $X_{Li}^{(s)} = \{X_{Li}(t_{ij}),\, j = 1, ..., n_i, \text{such as } t_{ij} \leq s\}$ and X_{Si}[20–21–25]. In a SREM, this reduces to:

$$\begin{aligned} p_i(s,t;\theta) &= \mathbb{P}\left(T_i \leq s+t | T_i \geq s, Y_i^{(s)}, X_{Li}^{(s)}, X_{Si}; \theta\right) \\ &= \frac{\displaystyle\int_{b_i} f_{Y^{(s)}}\left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta\right) \left[S_i\left(s | X_{Si}, b_i; \theta\right) - S_i\left(s+t | X_{Si}, b_i; \theta\right)\right] f_b\left(b_i; \theta\right) db_i}{\displaystyle\int_{b_i} f_{Y^{(s)}}\left(Y_i^{(s)} | X_{Li}^{(s)}, b_i; \theta\right) S_i\left(s | X_{Si}, b_i; \theta\right) f_b\left(b_i; \theta\right) db_i} \end{aligned} \tag{4}$$

where f_b and S_i are defined in section 2.3, and $f_{Y^{(s)}}$ is the multivariate Gaussian density function with mean $X_{Li}^{(s)}\beta + Z_i^{(s)} b_i$ and variance-covariance matrix $\Sigma_i^{(s)}$, superscript (s) meaning that only information before time s is considered.

3.2 Dynamic prognostic tool

Once a joint model is estimated, dynamic prognostic tools can be constructed from its estimates $\hat{\theta}$ and the variance of its estimates $\widehat{\mathrm{V}(\hat{\theta})}$.

At a time s of prediction, for a new subject i with biomarker history $Y_i^{(s)}$ and covariates $X_{Li}^{(s)}$ and X_{Si}, a dynamic prognostic tool can be computed as the predicted probability of event defined in (4) and computed at the point estimates $\hat{\theta}$. However, as this dynamic prognostic tool is an estimate, its uncertainty should be taken into account. This can

be done by approximating the posterior distribution of $p_i(s,t;\theta)$ defined in (4) using a Monte-Carlo method. It consists in generating a large set of $\theta^{(d)}$ $(d = 1, ..., D)$ from the asymptotic distribution of the estimates $\mathcal{N}\left(\hat{\theta}, \widehat{\mathrm{V}(\hat{\theta})}\right)$, and compute $p_i(s,t;\theta^{(d)})$. The median of $p_i(s,t;\theta^{(d)})$ gives a point estimate over the D draws while the 2.5% and 97.5% percentiles give the 95% confidence bands[20–21]. Using also a Monte-Carlo method to generate a large set of $\theta^{(d)}$ $(d = 1, ..., D)$, Rizopoulos[25] proposed to generate random-effects $b_i^{(d)}$ from their posterior distribution and compute the predicted probability of event given $b_i^{(d)}$ instead of computing directly $p_i(s,t;\theta^{(d)})$.

3.3 Evaluation of the predictive accuracy

To evaluate the predictive accuracy of dynamic prognostic tools, different measures can be used. The quadratic error of prediction called the Brier score and developed in survival models[6–7] was extended to the dynamic setting of joint models[10–20–21–29]. It consists in the expectation of the squared difference between the observed status of event and the predicted survival. The methodology of the area under the ROC curve (AUC) was also extended to the longitudinal setting[25–39] to assess how a longitudinal biomarker discriminated between patients with low and high risks of event. Recently, another predictive accuracy measure was developed from the information theory. This is the expected prognostic observed cross-entropy (EPOCE)[2–20] which is defined as $E\left[-\log\left(f_{T|Y^{(s)},T^*\geq s}(T)\right)|T^* \geq s\right]$, that is the expectation of the log of the conditional density $f_{T|Y^{(s)},T^*\geq s}$ of the time of event given the history of the marker until the time of prediction s.

One major problem when evaluating the predictive performances of a model is the expected over-optimism of the performances when estimated on the data on which the model was estimated. One way to counter that is to validate the model on external data but this is not always possible in practice. An alternative is to compute the measure on

the estimating data but using a cross-validation technique[7]. However this is numerically too intensive with joint models. In contrast to the AUC or Brier score, a formula approximating the leave-one-out cross-validation was proposed by Commenges et al.[2] for the EPOCE estimator so that this estimator provides a correct measure of predictive accuracy that can be directly computed on the estimating data in a very short computation time.

In this context, we chose to focus in this work only on the EPOCE to evaluate the predictive accuracy of the joint models on the estimating data by correcting the over-optimism. The approximated cross-validated estimator of the EPOCE called CVPOL_a is defined by:

$$\text{CVPOL}_\text{a}(s) \quad = \quad -\frac{1}{N_s}\sum_{i=1}^{N_s} F_i\left(\hat{\theta}, s\right) + N\,\text{Trace}\left(H^{-1}K_s\right) \tag{5}$$

where N_s is the number of subjects still at risk at time s, $F_i\left(\hat{\theta}, s\right)$ is the individual contribution to the conditional log-likelihood and is defined below in (6), H is the Hessian matrix of the joint log-likelihood defined in (3), and $K_s = \frac{1}{N_s(N-1)}\sum_{i=1}^{N} \mathbb{1}_{\{T_i \geq s\}} \hat{v}_i(s)\hat{d}_i^T$ with $\hat{v}_i(s)$ and $\hat{d}_i^T$ the gradients of the individual contributions respectively to the conditional log-likelihood computed in s using only $Y_i^{(s)}$ and defined in (6), and the joint log-likelihood defined in (3) and computed using all measurements of Y_i of subject i. These gradients are computed numerically using finite differences. In (5), the individual contribution to the conditional log-likelihood of the time-to-event given that the subject is still at risk at time s and given his biomarker history until s is:

$$F_i\left(\hat{\theta}, s\right) \quad = \quad \mathbb{1}_{\{T_i \geq s\}} \log\left(\frac{\displaystyle\int_{b_i} f_Y\left(Y_i^{(s)}|X_{Li}^{(s)}, b_i; \theta\right) \lambda\left(T_i|X_{Si}, b_i; \theta\right)^{E_i} S_i\left(T_i|X_{Si}, b_i; \theta\right) f_b\left(b_i; \theta\right)\, db_i}{\displaystyle\int_{b_i} f_Y\left(Y_i^{(s)}|X_{Li}^{(s)}, b_i; \theta\right) S_i\left(s|X_{Si}, b_i; \theta\right) f_b\left(b_i; \theta\right)\, db_i}\right) \tag{6}$$

The CVPOL_a has two components. The first part $\left(-\frac{1}{N_s}\sum_{i=1}^{N_s} F_i\left(\hat{\theta}, s\right)\right)$ estimates the apparent cross-entropy in the sample while the second term $(N \,\text{Trace}\,(H^{-1}K_s))$ provides the correction for over-optimism computed by the approximated leave-one-out cross-validation. As a measure of risk, the lower the CVPOL_a, the better the predictive accuracy of the model. Difference in EPOCE can also be computed with a 95% tracking interval[2] making possible the comparison of predictive accuracy between two models.

We note that for this work, the predictive accuracy was assessed in the entire remaining window of time from the time of prediction s but the EPOCE estimate could also be computed to evaluate the predictive accuracy of a model in a shorter window of horizons in order to focus on the ability to predict the event in the next τ years.

4 Application to prostate cancer data

The objective of this application was to better understand the dependency between the dynamics of the prostate specific antigen (PSA) and the risk of clinical recurrence after a first treatment by radiation therapy for a localized prostate cancer, and to explore the predictiveness of joint models based on the PSA dynamics.

Through this application, we also aimed at showing how shared random-effect models could be used in practice, how their goodness-of-fit could be evaluated, and how dynamic predictive tools could be derived and validated.

4.1 Description of the dataset

The data come from an hospital cohort of the University of Michigan that included patients treated with external beam radiation therapy (EBRT) for a localized prostate cancer and followed-up from 1988 to 2004. The present sample is constituted of 459 men

among whom 74 (16.1%) had a clinical recurrence defined as any distant metastases, nodal recurrence or any palpable or biopsy-detected local recurrence three years or later after radiation, any local recurrence within three years of EBRT if the last PSA value was over 2ng/ml and death from prostate cancer. PSA measurements were collected from the end of EBRT to clinical recurrence or lost to follow-up. Patients included in the sample were required to have at least 1-year follow-up without clinical recurrence, did not initiate any second treatment during the follow-up and had at least two PSA measurements before the end of follow-up.

Three main prognostic factors known at diagnosis were considered in the analysis and included in the vector of time-independent covariates X_i:

1. The level of PSA at diagnosis was defined as a continuous covariate in the log scale; The median was 2.1 (in log of ng/ml) with an interquartile range of [1.7, 2.6] (in log of ng/ml)

2. T-stage indicates how far the cancer has developed; it ranges from 1 when the cancer is small and contained within the prostate, to 4 when the cancer has spread to other organs, or bones. In the application, T-stage was split in 2 categories, 3-4 *versus* 1-2, with respectively 8.9% and 91.1% of the patients in each category.

3. The Gleason score quantifies the aggressiveness of prostate cancer; it ranges from 2 to 10, 10 indicating the higher aggressiveness. In the application, Gleason score was considered in 3 categories, 7 and 8-10 *versus* 2-6, with respectively, 37.7%, 7.4% and 54.9% of the patients.

Patients had a median follow-up time of 5.2 years with an interquartile range of [2.7, 7.7] years and a median time to clinical recurrence of 2.8 years with an interquartile range of [1.9, 4.3] years after EBRT.

PSA repeated measures were analyzed in the logarithmic scale $\log(PSA + 0.1)$ to satisfy the normality assumption of the linear mixed model. In the following, the term biomarker and log PSA will refer to $\log(PSA + 0.1)$. Figure 1 presents the individual observed trajectories of log PSA according to the occurrence of a clinical recurrence during the follow-up. In the two subgroups, the log PSA trajectory is characterized by a very rapid PSA drop in the first year after EBRT and a long-term increase. However, the intensity of the long-term increase differs greatly according to the subgroup: patients who recurred tend to have a faster PSA rise than patients who did not.

[Figure 1 about here.]

4.2 Specification of the joint models

To describe the dynamics of the PSA after EBRT and its link with time to clinical recurrence, we explored different SREM models that only differed by the function $h(b_i, t)$ assumed to summarize the dependency between the PSA dynamics and the risk of clinical recurrence.

The log PSA trajectory was systematically described by the following linear mixed model:

$$\begin{aligned} Y_i(t) &= \log(PSA + 0.1) \\ &= Y_i^*(t) + \epsilon_i(t) \\ &= \left(\mu_0 + X_i^T\beta_0 + b_{0i}\right) + \left(\mu_1 + X_i^T\beta_1 + b_{1i}\right) f_1(t) + \left(\mu_2 + X_i^T\beta_2 + b_{2i}\right) t + \epsilon_i(t), \forall t \in \mathbb{R}^+ \end{aligned} \tag{7}$$

The biphasic trajectory of log PSA was captured by $f_1(t) = ((1+t)^{-1.5} - 1)$ that

modeled the initial decline of log PSA after the end of EBRT and t that modeled the linear long-term rise in log PSA. Proust-Lima et al.[22] showed through a pooled analysis of five large cohorts, by profile maximum likelihood technique, that these functions correctly captured the nonlinear biphasic PSA trajectory after EBRT. The vector $\mu = (\mu_0, \mu_1, \mu_2)$ represents the intercepts in the different phases; the vector of random effects $bi = (b_{i0}, b_{i1}, b_{i2})$ is multivariate Gaussian with vector of mean $\mathbf{0}$ and unstructured variance-covariance matrix. We remind that the vector X_i includes the three baseline prognostic factors.

The time-to-recurrence was systematically modeled using the following proportional hazard model:

$$\lambda_i\left(t|X_i, b_i\right) = \lambda_0\left(t\right) e^{X_i^T \gamma + h(b_i,t)^T \eta} \tag{8}$$

in which the log-baseline hazard, $\log\left(\lambda_0(t)\right)$ was approximated using B-splines with three internal knots placed at the quartiles of the observed event times.

We explored the nature of the dependency between the log PSA dynamics described in (7) and the risk of clinical recurrence described by considering the following functions for $h(b_i, t)$ in (8):

(a) Current level of log PSA: $h_{\mathrm{a}}(b_i, t) = Y_i^*(t)$.

This is the most classical function assumed in a SREM. It assumes that the current true level of the biomarker is predictive of the risk of event.

(b) Current slope of log PSA: $h_{\mathrm{b}}(b_i, t) = \partial Y_i^*(t)/\partial t$.

This model assumes that the current change of the biomarker is predictive of the risk of event.

(c) Current level and current slope of log PSA: $h_{\mathrm{c}}(b_i, t) = \left(Y_i^*(t), \partial Y_i^*(t)/\partial t\right)^T$.

With $h_{\mathrm{c}}(b_i, t)$, the risk of event is linked both to the current true level of the biomarker and the current change in the biomarker.

(d) Lagged level and lagged slope of log PSA: $h_{\mathrm{d}}(b_i, t) = (Y_i^*(t-1), \partial Y_i^*(t-1)/\partial t)^T$.
Instead of considering that the current biomarker information explains the risk of event, a lag can be introduced as in $h_{\mathrm{d}}(b_i, t)$ where we assumed a latency of 1 year between the log PSA characteristics and their impact on the risk of event.

(e) Transformed current level of log PSA: $h_{\mathrm{e}}(b_i, t) = \Gamma\left(Y_i^*(t)\right)$.
A SREM with $h_{\mathrm{a}}(b_i, t) = Y_i^*(t)$ makes a log-linear assumption between the current true level and the risk of event. To relax this assumption, a transformation of the current level can be considered with function Γ. In the PSA progression after EBRT, Proust-Lima et al.[22] found that the transformation $\Gamma\left(Y_i^*(t)\right) = \mathrm{logit}^{-1}\left(\left(Y_i^*(t) - 0.71\right)/0.44\right)$ better modeled the effect of the current level of PSA and the risk of recurrence.

(f) Transformed current level and current slope of PSA: $h_{\mathrm{f}}(b_i, t) = \left(\Gamma\left(Y_i^*(t)\right), \partial Y_i^*(t)/\partial t\right)^T$.
To relax both models (c) and (e), the most adapted transformed current true level of log PSA and the current slope of log PSA can be simultaneously considered to describe the impact of PSA dynamics on the risk of recurrence.

(g) Random effects: $h_{\mathrm{g}}(b_i, t) = (b_{0i}, b_{1i}, b_{2i})^T$.
Instead of considering some functions of the level of the biomarker, the individual deviations can be directly included in the survival model to explain the impact of the PSA dynamics. In that case, the function $h_{\mathrm{g}}(b_i)$ linking both processes is time-independent.

We only considered the seven functions above but other functions $h(b_i, t)$ could be investigated according to the question of research.

4.3 Joint models estimations

[Table 1 about here.]

Estimation of these seven models ((a) - (g)) is summarized in Table 1 with goodness-of-fit statistics and estimates of parameters η that summarize the link between the PSA dynamics and the risk of recurrence adjusted for other prognostic factors. Whatever the assumed dependency between the PSA dynamics and the risk of recurrence, the seven models concluded to a significant association between the two processes. However, according to the Akaike criterion, the goodness-of-fit of the models varied substantially.

Models (a) and (b) underlined that the current true level of log PSA and the current slope of log PSA were associated with the risk of recurrence, higher current level or slope predicting higher risk of recurrence. Model (c) showed that these two features of PSA dynamics were independent predictors of the risk of recurrence. Their effects were partly reduced when considered simultaneously as predictors of the risk of recurrence but they both remained highly significant ($p < 0.0001$ and $p < 0.0001$).

Model (d) illustrates how some lag in the effect of a time-dependent covariate can be introduced in joint models, with the exploration here of a latency of 1 year between the characteristics of the PSA trajectory and their impact on the risk of recurrence. In the present example, the AIC was better when assuming an association between the risk of recurrence and the level and slope of log PSA at the current time (model (c)) rather than one year before (model (d)). However, the parameter estimates were roughly the same in the two models.

Models (a), (c) and (d) assume a log-linear effect of the level of log PSA on the risk of recurrence, which means that the effect of 1 additional unit of log PSA is constant whatever the level of log PSA. As suggested by previous works[22], a non-linear effect of the current level of log PSA was explored via a logistic transformation in model (e) and

combined with a joint effect of the current slope in model (f). From the seven models investigated, the latter model assuming a dependency via the transformed current log PSA level and the current slope provided the best fit with an improvement of 46 points of AIC compared to the model assuming a dependency via the crude current level of log PSA in addition to the slope. The logistic transformation makes the effect of the level of PSA increase in the range 0 to 4 ng/mL, and become maximal around 4ng/mL. This result is in accordance with clinical assumptions regarding the range of meaningful PSA levels. Let note that when considering the logistic transformation of the level of log PSA (model (f)) in the survival model instead of the crude true log-PSA level (model (c)), the association between the current slope and the risk of recurrence was still significant but its intensity was divided by two.

Finally, instead of assuming a dependency of the risk of recurrence on the PSA dynamics via the transformation and/or derivative of the current level of log PSA, the last model (g) included the individual deviations of the PSA dynamics, that are the random-effects, as predictors of the risk of recurrence in the survival model. Although the model used 1 additional parameter to summarize the dependency, it gave a worse fit than the models including the current slope of log PSA and the transformed (or crude) current level of log PSA with an AIC higher by 52.3 points compared to model (f) (and respectively 6.3 points compared to model (c)). Interestingly, as a linear long-term trend was assumed for the log PSA trajectory in (7), $\partial Y_i^*(t)/\partial t$ reduces approximately to $\left(\mu_2 + X_i^T\beta_2 + b_{2i}\right)$ after a few years, and effects of $\partial Y_i^*(t)/\partial t$ and b_{2i} adjusted for X_i in the survival model can be quantitatively compared. So with an effect estimated around 4, model (g) gives a larger impact of the long-term slope on the risk of recurrence than models (c) and (f) with estimates of the effect of the long-term slopes around 2 and 1 respectively.

4.4 Goodness-of-fit assessment

[Figure 2 about here.]

In joint models, like in all statistical models, model assumptions can be checked with a comparison of the predicted *versus* the observed values and an inspection of the residuals.

Figure 2 compares respectively the marginal $(X_{Li}(t_{ij})^T \hat{\beta})$ and the subject-specific $(X_{Li}(t_{ij})^T \hat{\beta} + Z_i(t_{ij})^T \hat{b}_i$ with $\hat{b}_i$ the empirical Bayes estimate) predictions from the linear mixed model with the observed log PSA values. Observations and predictions are averaged by intervals of times having the same amount of PSA measures. First, whatever the assumed dependency between processes, the longitudinal predictions are very similar between models. This was expected since the structure of the longitudinal model was the same in all the models and all the models took into account (even if differently) the association with the risk of recurrence. Second, the marginal predictions deviate from the observed mean trajectory from 2-3 years after the end of EBRT but the subject-specific predictions show a very good fit to the observed data. This deviation of the marginal predictions is due to the informative dropout caused by the clinical recurrence and was again expected. In the seven models, the residuals of the longitudinal submodel were very similar and did not exhibit any heteroscedasticity problem or model misspecification (results not shown).

For the goodness-of-fit of the survival submodel, we only focused on the four joint models (b), (c), (f) and (g). Figure 3 displays the predicted survival curves and the Kaplan-Meier estimate of the time-to-recurrence. The predicted survival curves were computed as $\frac{1}{N}\sum_{i=1}^{N} S_i\left(t|\hat{b}_i\right)$ where $\hat{b}_i$ is the empirical Bayes estimate of b_i. All the predicted survival curves remained in the 95% confidence interval of the Kaplan-Meier estimate showing an overall correct goodness-of-fit even if the curves slightly differed one

from the other.

[Figure 3 about here.]

We used also the Cox-Snell residuals computed as $\left(\Lambda_i\left(T_i|\hat{b}_i\right) = \int_0^{T_i}\lambda_i\left(t|\hat{b}_i\right)dt\right)$ to evaluate the overall goodness-of-fit of the survival models. Figure 4 displays the Kaplan-Meier estimate of the Cox-Snell residuals from the 4 models. They are relatively close to the theoretical survival curve of a unit exponential random variable ($\exp(-t)$) showing the overall satisfying goodness-of-fit of the four survival models.

Using the Martingale residuals computed as $\left(E_i - \Lambda_i\left(T_i|\hat{b}_i\right)\right)$, we specifically assessed in Figure 5 the log-linearity assumption of the covariates linking the PSA dynamics and the risk of recurrence. In model (c) with the current level and slope, the residuals are satisfying according to the slope of log PSA but exhibit a deviation in the larger values of the current level of log PSA. When assuming instead a transformation for the current log PSA level in model (f), this deviation disappears, showing a very good adequation to the log-linearity assumption with the logistic transformation of the current level of log PSA and the current slope. In model (b) assuming a dependency via the slope only, the residuals also highlight a deviation in the large values of the predicted slope. The same phenomenon is observed in model (g) where the dependency is directly on the individual random-effects. While the residuals are correct for the random intercept and the random short-term drop, a deviation from the zero-mean is observed for larger random long-term slopes.

[Figure 4 about here.]

[Figure 5 about here.]

4.5 Predictive accuracy assessment

To evaluate the predictiveness of these joint models, we computed the approximated cross-validated prognostic likelihoods that estimate the EPOCE in Figure 6 (A), and the differences between models with 95% tracking intervals in Figure 6 (B). The abscissa of the graphs represents the times s of prediction so that each point gives the model ability to predict the risk of event from time s (in Figure 6 (A)) and the difference in model abilities (in Figure 6 (B)) for subjects still at risk at time s according to the PSA repeated measures collected before s. We remind that the EPOCE is a risk so that the lower the EPOCE estimate, the better the predictive accuracy of the model.

[Figure 6 about here.]

Interestingly, the model with the best goodness-of-fit, that is model (f), does not have the best predictive accuracy for all the times of prediction between 1 year and 6 years after the end of EBRT. Its predictive accuracy is the best (lowest EPOCE estimate) when using information in the first 2.25 years after the end of EBRT on patients still at risk at these times. However, after 2.25 years (and especially after 3 years), its predictive accuracy becomes a lot worse compared to the ones of models (b), (c) and (g).

Compared to the goodness-of-fit measures, the predictive accuracy measures focus only on subjects still at risk at the time of prediction. Yet, in our dataset, 50% of the events occurred before 2.8 years and 75% occurred before 4 years. This means that most of the information available is concentrated in the first years after the end of EBRT, giving more importance to that period of time in the estimation process compared to longer times after EBRT. As a consequence, models that gave a very good fit to the data in the first years as model (f) had better overall goodness-of-fit. This is confirmed by the predicted survival curves in Figure 3 which show a slightly better fit of model (f) in the first years after the end of EBRT (and a slightly better fit of model (g) in longer times

after end of EBRT).

Finally, we note that in the longer times of prediction, the models having the smallest EPOCE were those assuming a dependency through the random-effects or through the current log PSA slope. This suggests that after a few years, the slope of log PSA could be the major predictor of the risk of recurrence on subjects still at risk of recurrence. Indeed, although most of the models included the slope as a predictor, the intensity of the effect of the current slope was a lot more important in model (g) with the random-effects (parameter estimate around 4) and in model (b) with only the current slope (parameter estimate around 3) compared to models (c) and (f) including the current log PSA level also (and where the parameter estimates of the effect of the current slope were around 2 and 1).

4.6 Individual predictions

To illustrate how joint models can be used to provide dynamic individual predictions using the methodology described in section 3.2, Figure 7 displays dynamic individual predictions of the risk of clinical recurrence computed at an horizon of 3 years and updated every 6 months from the joint model (f). The patient had a T-stage of 2, a Gleason score of 7, an initial PSA of 9.7ng/mL and had a clinical recurrence after 3.8 years following the end of EBRT. The figure illustrates how the dynamic predictive tool can be updated at each new measurement of the biomarker, with here a progressive increase in the predicted risk of recurrence over years.

[Figure 7 about here.]

5 Conclusion

This paper aimed at explaining through an application on a real dataset how shared random-effect models could be used in practice, how their assumptions could be evaluated, and how individual predictions could be derived from them.

We specifically focused here on SREM that differed by the nature of the dependency assumed between the biomarker dynamics and the risk of the event. Indeed, most applications of the SREM assume that only the current level of the biomarker impacts the risk of event. Yet, first the log-linear assumption induced by the inclusion of the current level of the biomarker in the proportional hazard model may not hold, and second, other characteristics of the biomarker trajectory can better summarize the dependency between the two processes than the current level. For example, in the application, we first found using Martingale residuals that the log-linear assumption of the current true level of the biomarker did not hold and that a logistic transformation of this current true level should be considered instead. Second, we found that the current slope of the biomarker was the most important predictor of the risk of recurrence.

As we were interested both in the fit of the models to the data, and the models ability to predict the risk of event, we evaluated both the goodness-of-fit and the predictive accuracy of the models. Interestingly, we found that the model with the best Akaike criterion and adequate Martingale residuals (i.e. the best goodness-of-fit) was not the best model for predicting the risk of recurrence whatever the time elapsed since the end of the treatment. This illustrates the difference between these two types of evaluation. The goodness-of-fit measures use all the information available while the predictive accuracy measures use only a part of the longitudinal biomarker (the history of the biomarker up to time s) and only a part of the sample (the subjects still at risk at time s). This raises the question of the methodology which should be favored when developing dynamic predictive tools. When interested only in the predictive ability of a model and the development of a

dynamic predictive tools, predictive accuracy measures may be preferable over goodness-of-fit measures.

Among predictive accuracy measures, we only used the prognostic cross-entropy[2–20] which offers a natural way to focus on the ability to predict the event from the history of the marker. Yet, other measures could be investigated as the Brier score[6–7–20] and AUC[25–39]. Nevertheless, it should be kept in mind that the over-optimism with predictive accuracy measures computed on the internal data always needs to be corrected. This was done using an approximated cross-validation formula for the EPOCE but no such approximation is available for the Brier score or AUC yet, and classical cross-validation technique is computationally too intensive with joint models to be considered in practice. We also note that the EPOCE has the advantage to provide a tracking interval of the difference between EPOCE of two models. This is very useful to appreciate the gain in predictive accuracy of a model over another, and has not been developed yet for the other measures.

Acknowledgements

The authors thank Jeremy M G Taylor from the Department of Biostatistics, University of Michigan Ann Arbor (Michigan, USA) for providing the dataset and the French National Institute of Cancer INCa for financing PREDYC project.

References

1. Yueh-Yun Chi and Joseph G. Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–45, June 2006.

2. Daniel Commenges, Benoit Liquet, and Cécile Proust-Lima. Choice of prognostic

estimators in joint models by estimating differences of expected conditional Kullback-Leibler risks. *Biometrics*, 68(2):380–7, June 2012.

3. David R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

4. Michael J. Crowther, Keith R. Abrams, and Paul C. Lambert. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine*, 31(30):4456–4471, 2012.

5. Cheryl L. Faucett and Duncan C. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine*, 15(15):1663–85, August 1996.

6. Thomas A. Gerds and Martin Schumacher. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. *Biometrical Journal*, 48(6):1029–1040, December 2006.

7. Thomas A. Gerds and Martin Schumacher. Efron-type measures of prediction error for survival analysis. *Biometrics*, 63(4):1283–7, December 2007.

8. Xu Guo and Bradley P. Carlin. Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*, 58(1):16–24, February 2004.

9. Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics (Oxford, England)*, 1(4):465–80, December 2000.

10. Robin Henderson, Peter Diggle, and Angela Dobson. Identification and efficacy of longitudinal markers for survival. *Biostatistics (Oxford, England)*, 3(1):33–50, March 2002.

11. Xin Huang, Gang Li, Robert Elashoff, and Jianxin Pan. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis*, 17:80–100, 2011.

12. Joseph G. Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. Bayesian Survival Analysis. *Springer*, 2001.

13. Joseph G Ibrahim, Haitao Chu, and Liddy M Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 28(16):2796–801, June 2010.

14. Héléne Jacqmin-Gadda, Rodolphe Thiébaut, and Jean F. Dartigues. Modélisation conjointe de données longitudinales quantitatives et délais censurés. *Revue d'épidémiologie et de santé publique*, 52(6):502–10, 2004.

15. JD Kalbfleisch and RL Prentice. The statistical analysis of failure time data, Second Edition. *John Wiley*, 2002.

16. Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

17. Ngayee J Law, Jeremy M G Taylor, and Howard Sandler. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics (Oxford, England)*, 3(4):547–63, December 2002.

18. Emmanuel Lesaffre and Bart Spiessens. On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):325–335, 2001.

19. Haiqun Lin, Bruce W. Turnbull, Charles E. McCulloch, and Elizabeth H. Slate. Latent class models for joint analysis of longitudinal biomarker and event process

data : Application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457), mar 2002.

20. Cécile Proust-Lima, Mbéry Sène, Jeremy M G. Taylor, and Héléne Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical methods in medical research*, April 2012.

21. Cécile Proust-Lima and Jeremy M G. Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics (Oxford, England)*, 10(3):535–49, July 2009.

22. Cécile Proust-Lima, Jeremy M G. Taylor, Willams Scott, D. Ankerst, N. Liu, L. Kestin, Bae K, and Sandler Howard. Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. *International Journal of Radiation Oncology Biology Physics*, aug 2008.

23. Sarah J. Ratcliffe, Wensheng Guo, and Thomas R. Ten Have. Joint Modeling of Longitudinal and Survival Data via a Common Frailty. *Biometrics*, 60:892–899, dec 2004.

24. Dimitris Rizopoulos. JM : An R Package for the Joint Modelling of longitudinal and time-to-event data. *Journal Of Statistical Software*, 35(9), 2010.

25. Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–29, September 2011.

26. Dimitris Rizopoulos. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics and Data Analysis*, 56(3):491 – 501, 2012.

27. Dimitris Rizopoulos and Pulak Ghosh. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12):1366–80, May 2011.

28. Dimitris Rizopoulos, Geert Verbeke, and Emmanuel Lesaffre. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):637–654, June 2009.

29. Rotraut Schoop, Martin Schumacher, and Erika Graf. Measures of prediction error for survival data with longitudinal covariates. *Biometrical journal. Biometrische Zeitschrift*, 53(2):275–93, March 2011.

30. Michael J. Sweeting and Simon G. Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763, 2011.

31. Anastasios A. Tsiatis and Marie Davidian. Joint modeling of longitudinal and time-to-event data : An overview. *Statistica Sinica*, 14:809–834, 2004.

32. Anastasios A. Tsiatis, DeGruttola Victor, and Michael S. Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Stastical Association*, 90(429), March 1995.

33. Yan Wang and Jeremy M G. Taylor. Jointly Modeling Longitudinal and Event Time Data With Application to Acquired Immunodeficiency Syndrome. *Journal of the American Statistical Association*, 96(455), September 2001.

34. Lang Wu, Wei Liu, G Y Yi, and Yangxin Huang. Analysis of Longitudinal and

Survival Data: Joint Modeling, Inference Methods, and Issues. *Journal of Probability and Statistics*, 2012.

35. Michael S. Wulfsohn and Anastasios A. Tsiatis. A joint model of survival and longitudinal data measured with error. *Biometrics*, 53:330–339, March 1997.

36. Menggang Yu, Ngayee J. Law, Jeremy M G. Taylor, and Howard M. Sandler. Joint Longitudinal-Survival-Cure Models and Their Application to Prostate Cancer. *Statistica Sinica*, 14:835–862, 2004.

37. Menggang Yu, Jeremy M G. Taylor, and Howard M. Sandler. Individual Prediction in Prostate Cancer Studies Using a Joint Longitudinal Survival-Cure Model. *Journal of the American Statistical Association*, 103(481):178–187, March 2008.

38. Donglin Zeng and Jianwen Cai. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, 33(5):2132–2163, 2005.

39. Yingye Zheng and Patrick J. Heagerty. Prospective accuracy for longitudinal markers. *Biometrics*, 63(2):332–341, 2007.

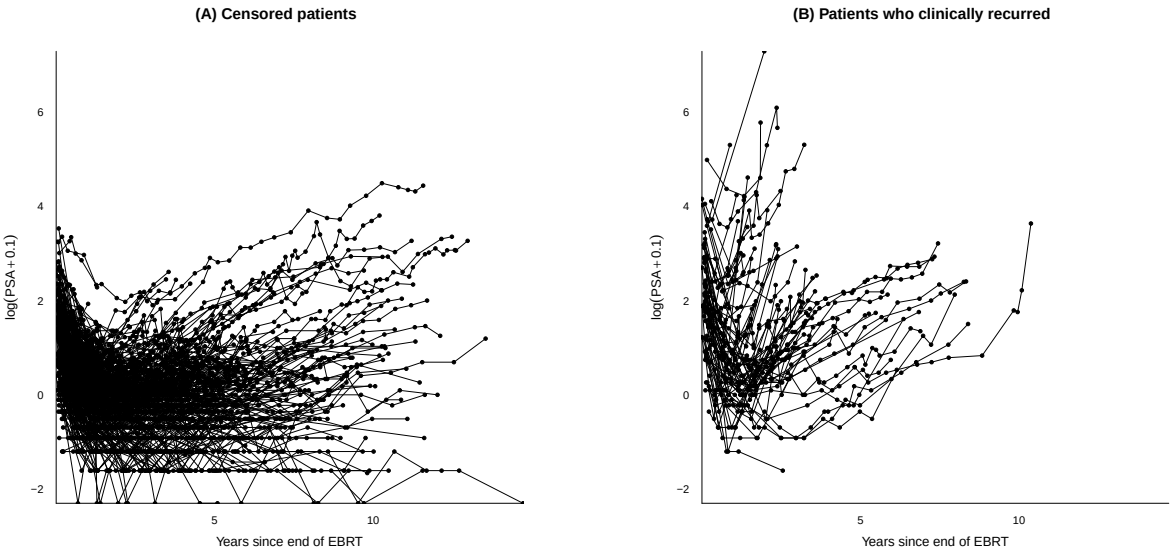

Figure 1: Individual observed trajectories of $\log(PSA + 0.1)$ over time for (A) censored patients and (B) patients who clinically recurred.

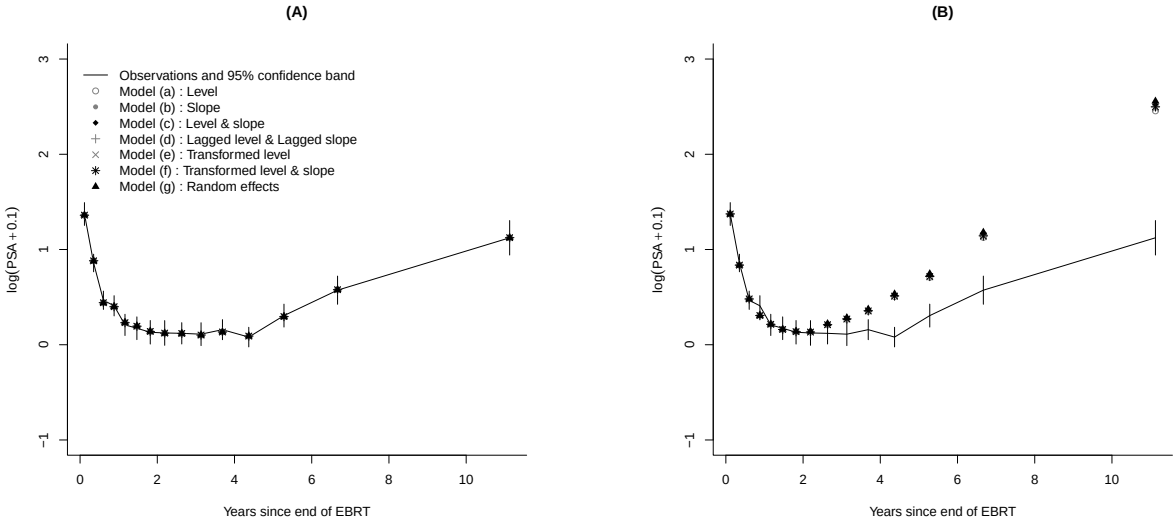

Figure 2: Comparison of the subject-specific predictions (A) and the marginal predictions (B) with the observed PSA values averaged by intervals of times having the same amount of PSA measures.

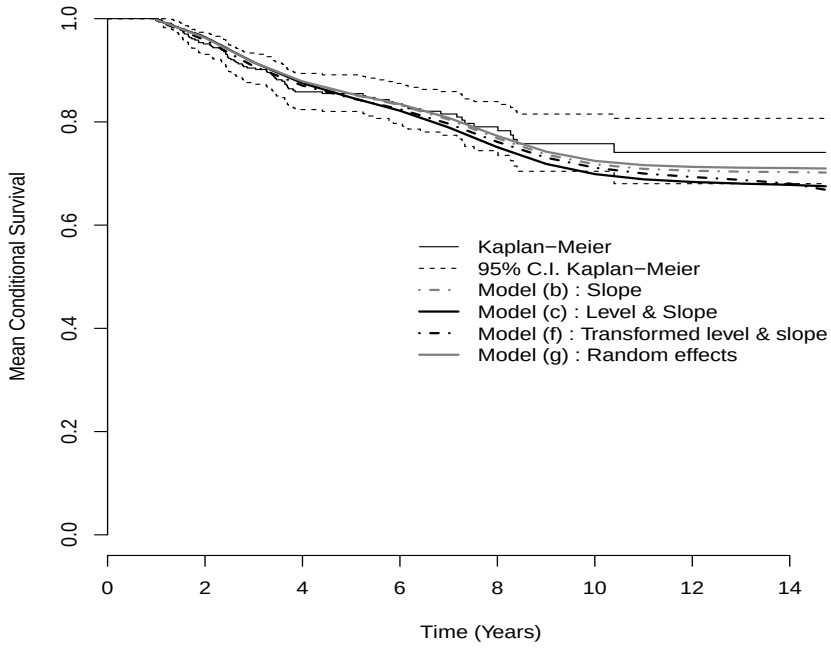

Figure 3: Mean predicted subject-specific survival curves from four joint models and observed Kaplan-Meier survival curve with 95% confidence bands.

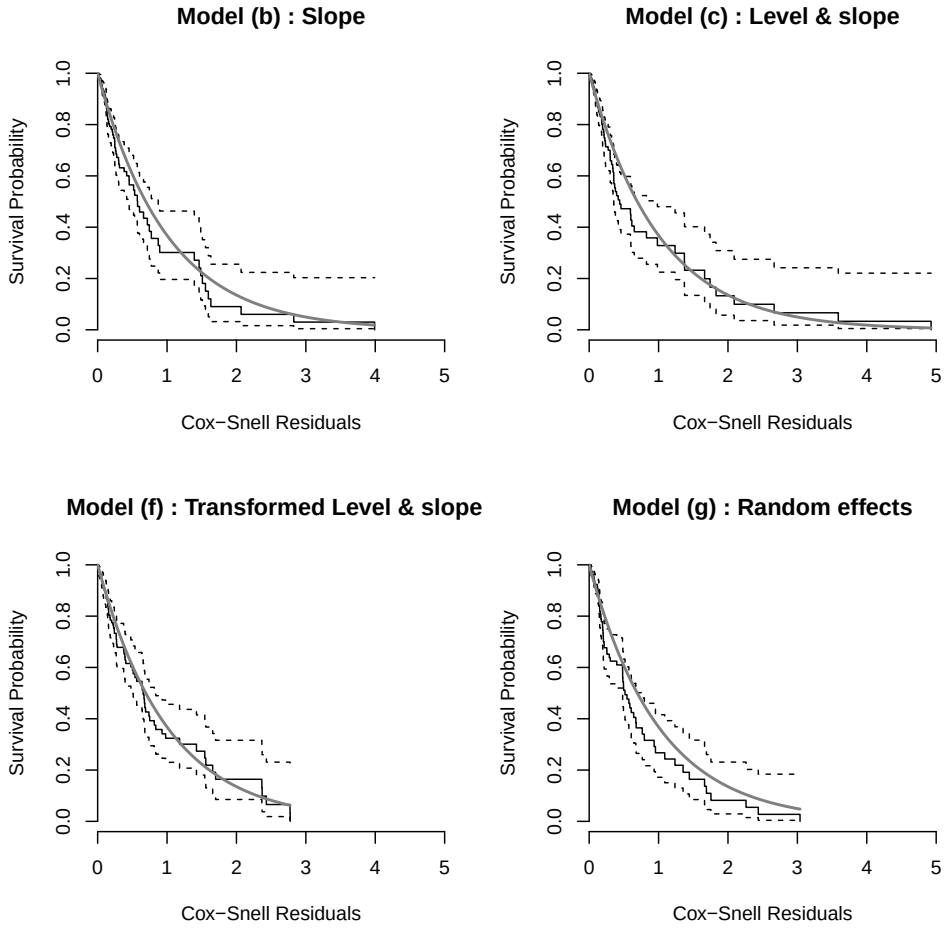

Figure 4: Kaplan-Meier estimates of the Cox-Snell residuals from four joint models ((b), (c), (f) and (g)) in plain black line, their 95% confidence bands in dashed black line, and the theoretical survival curve of a unit exponential random variable in plain grey line.

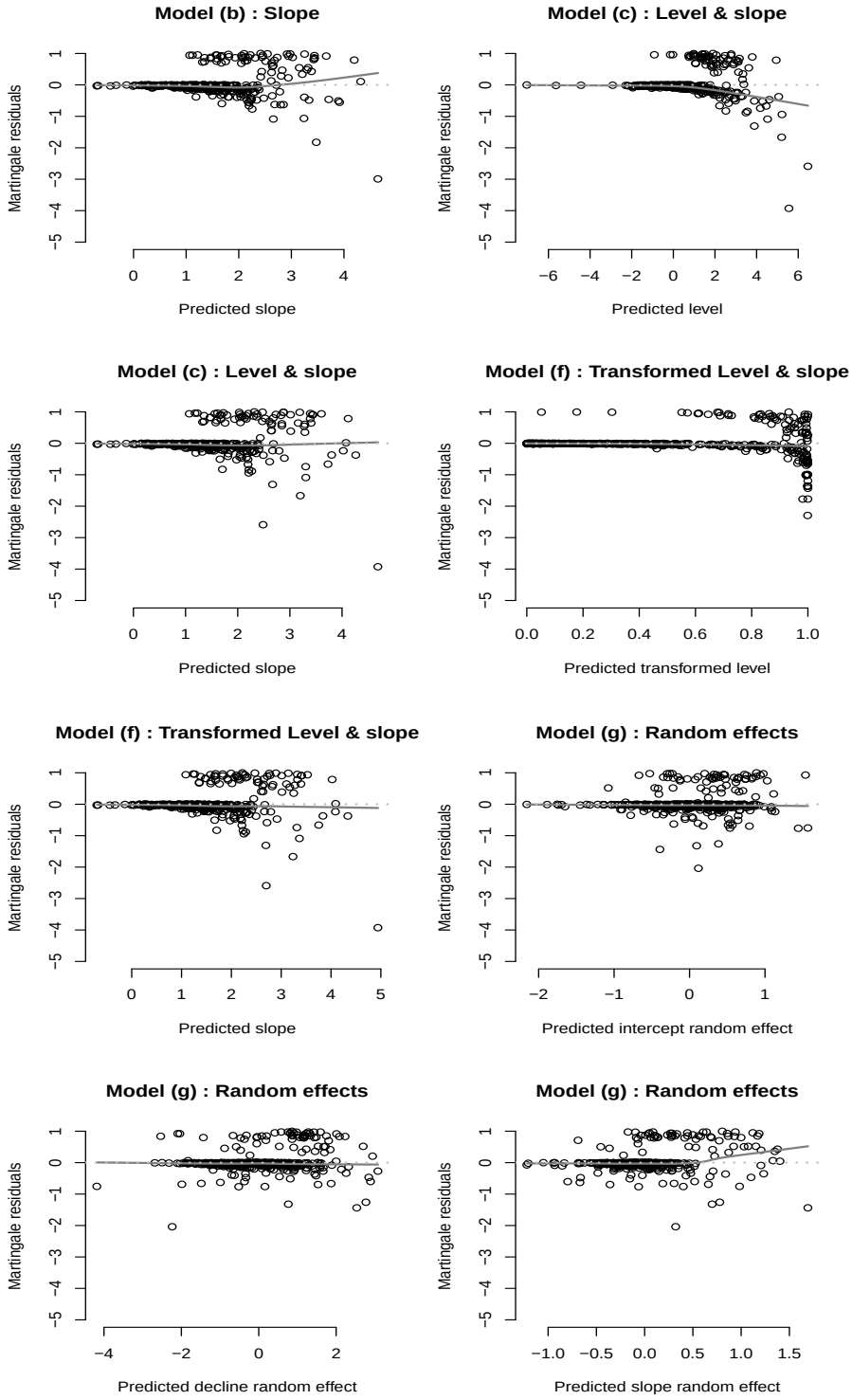

Figure 5: Martingale residuals *versus* predicted covariates summarizing the dependency between the PSA dynamics and the risk of recurrence in four joint models ((c), (f), (b) and (g)).

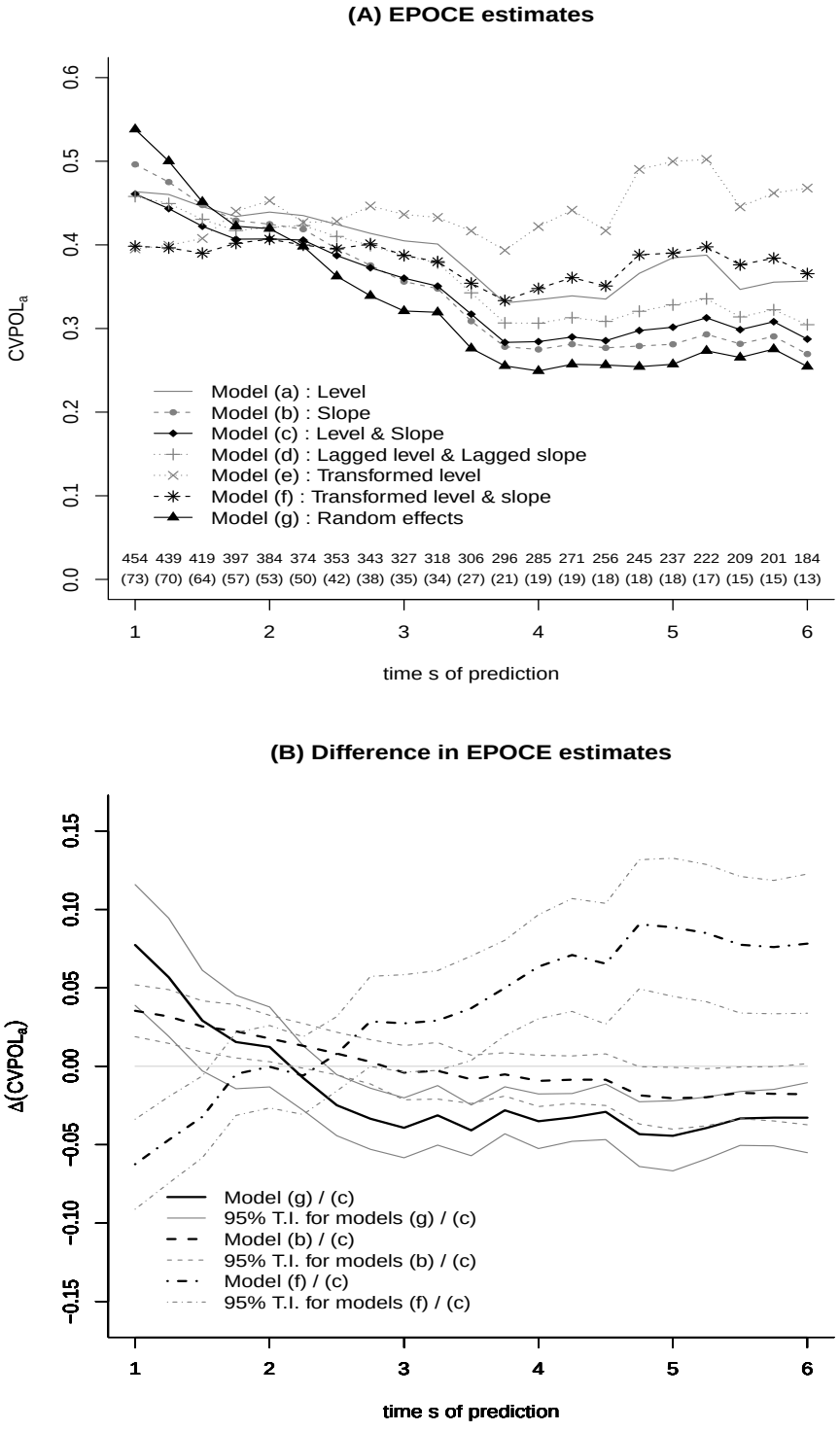

Figure 6: (A) CVPOL$_a$ computed from the 7 shared random-effect models from times of prediction from 1 to 6 years after the end of radiation therapy. For each time of prediction, the number of subjects still at risk (and the number of clinical recurrence) are indicated. (B) Differences in EPOCE estimate with 95% tracking interval (T.I.) between 4 joint models from times of prediction from 1 to 6 years after the end of radiation therapy.

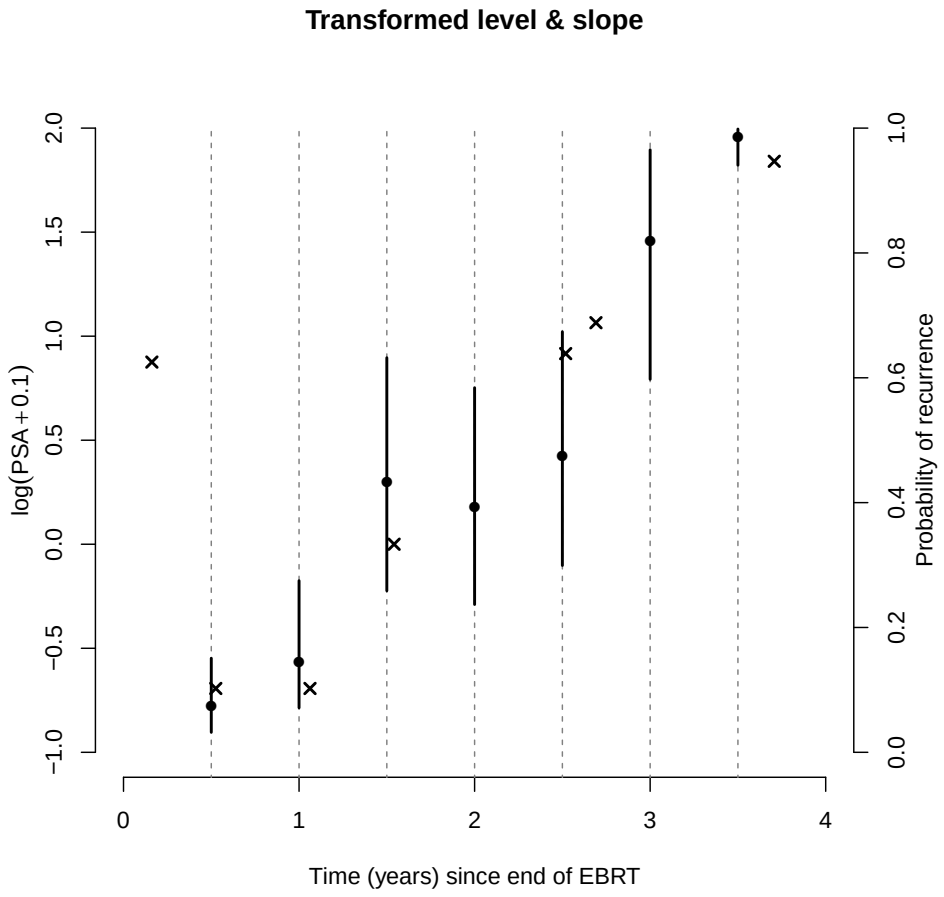

Figure 7: Individual dynamic predicted probability of clinical recurrence (•) within 3 years and updated every 6 months computed from model (f) and its 95% confidence bands (bold plain line) obtained by a Monte Carlo method with 2000 draws; repeated PSA measures are denoted by ×. The patient had a T-stage of 2, a Gleason score of 7, an initial PSA of 9.7ng/mL and had a clinical recurrence after 3.8 years following the end of EBRT.

Table 1: Goodness-of-fit statistics and parameter estimates (standard-deviation) of the association between PSA dynamics and risk of recurrence adjusted for other prognostic factors from joint models (a) to (g).

Model	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Dependency	$Y_i^*(t)$	$\frac{\partial Y_i^*(t)}{\partial t}$	$\begin{pmatrix} Y_i^*(t) \\ \frac{\partial Y_i^*(t)}{\partial t} \end{pmatrix}$	$\begin{pmatrix} Y_i^*(t-1) \\ \frac{\partial Y_i^*(t-1)}{\partial t} \end{pmatrix}$	$\Gamma\left(Y_i^*(t)\right)$	$\begin{pmatrix} \Gamma\left(Y_i^*(t)\right) \\ \frac{\partial Y_i^*(t)}{\partial t} \end{pmatrix}$	b_i
L	-2644.15	-2635.97	-2628.55	-2637.62	-2611.06	**-2605.57**	-2630.70
AIC	5356.29	5339.93	5327.09	5345.25	5290.11	**5281.14**	5333.41
# parameters	34	34	35	35	34	**35**	36
Level†	0.95 (0.08)	-	0,49 (0.12)	0.44 (0.15)	7.34 (0.78)	5.45 (0.84)	-
Slope‡	-	3.14 (0.35)	2.04 (0.39)	2.19 (0.37)	-	1.08 (0.32)	-
Random effects:							
b_{0i} (Intercept)	-	-	-	-	-	-	1.03 (0.30)
b_{1i} ($f_1(t)$)	-	-	-	-	-	-	-0.32 (0.12)
b_{2i} (t)	-	-	-	-	-	-	4.05 (0.55)

† Corresponds to the effect of the current true log PSA level in models (a) and (c); corresponds to the effect of the true log PSA level one year before in model (d), and
corresponds to the effect of the transformed true log PSA level in models (e) and (f).

‡ Corresponds to the effect of the current slope of log PSA in models (b), (c) and (f); corresponds to the effect of the slope one year before in model (d).