Biometrics DOI: 10.1111/biom.12490

Joint Model for Left-Censored Longitudinal Data, Recurrent Events and Terminal Event: Predictive Abilities of Tumor Burden for Cancer Evolution with Application to the FFCD 2000–05 Trial

Agnieszka Król,[1]* Loïc Ferrer,[1] Jean-Pierre Pignon,[2] Cécile Proust-Lima,[1] Michel Ducreux,[3] Olivier Bouché,[4] Stefan Michiels,[2] and Virginie Rondeau[1]

[1]University of Bordeaux, INSERM U1219, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France
[2]INSERM U1018 CESP, Service de Biostatistique et d'Épidémiologie Gustave Roussy, U. Paris-Sud,
114 rue Édouard-Vaillant, 94805 Villejuif Cedex, France
[3]Medical Oncology, Gustave Roussy, U. Paris-Sud, 114 rue Édouard-Vaillant, 94805 Villejuif Cedex, France
[4]University Hospital, Hôpital Robert Debré, Avenue du Général Koenig, 51092 Reims Cedex, France
**email:* agnieszka.krol@isped.u-bordeaux2.fr

Summary. In oncology, the international WHO and RECIST criteria have allowed the standardization of tumor response evaluation in order to identify the time of disease progression. These semi-quantitative measurements are often used as endpoints in phase II and phase III trials to study the efficacy of new therapies. However, through categorization of the continuous tumor size, information can be lost and they can be challenged by recently developed methods of modeling biomarkers in a longitudinal way. Thus, it is of interest to compare the predictive ability of cancer progressions based on categorical criteria and quantitative measures of tumor size (left-censored due to detection limit problems) and/or appearance of new lesions on overall survival. We propose a joint model for a simultaneous analysis of three types of data: a longitudinal marker, recurrent events, and a terminal event. The model allows to determine in a randomized clinical trial on which particular component treatment acts mostly. A simulation study is performed and shows that the proposed trivariate model is appropriate for practical use. We propose statistical tools that evaluate predictive accuracy for joint models to compare our model to models based on categorical criteria and their components. We apply the model to a randomized phase III clinical trial of metastatic colorectal cancer, conducted by the Fédération Francophone de Cancérologie Digestive (FFCD 2000–05 trial), which assigned 410 patients to two therapeutic strategies with multiple successive chemotherapy regimens.

Key words: Colorectal cancer; Joint model; Longitudinal data; Predictive accuracy; Recurrent events; Tumor measurement.

1. Introduction

In phase II trials in oncology a patient's response to a treatment is usually evaluated using categorical criteria such as the Response Evaluation Criteria in Solid Tumors (RECIST) (Eisenhauer et al., 2009) and the World Health Organization criteria (WHO, 1979). These sets of rules evaluate target, non-target, and new lesions in order to classify a tumor response to a given therapy into one of four categories: complete response, partial response, stable disease, and progressive disease. In phase III trials the progression-free survival (PFS) is often chosen as the clinical endpoint with progression being defined mainly by an increase of the tumor size (sum of products of the two longest diameters in perpendicular dimensions in WHO and sum of the longest diameters (SLD) in RECIST) and/or appearance of any new lesions. Given the continuous tumor size changes, the categorization performed within these criteria results in a loss of information on patient's tumor burden and risk of misclassification.

There have been several alternatives to RECIST and WHO criteria proposed for modeling the tumor response. Karrison et al. (2007) considered continuous change in tumor size burden as the primary endpoint in phase II clinical trials. Alternative categorizations of tumor response were proposed by An et al. (2011) (total sum of measurements, relative change from baseline) and evaluated in terms of association with overall survival (OS). Claret et al. (2013) proposed the individual time to tumor size nadir (or time to tumor growth) predicted from a tumor growth inhibition model, and used it as a covariate in a survival model for OS.

For a simultaneous analysis of repeated measurements of tumor size, appearance of new lesions and survival, we propose to use joint models. Joint models for longitudinal and time-to-event data (Schluchter, 1992; Wulfsohn and Tsiatis, 1997) consider dependency and association between repeated measurements of a biomarker and a terminal event through a shared latent structure. They overcome the time-dependent Cox model's limiting assumptions of external time-varying covariates not related to the failure mechanism and diminish the linear mixed-effects models' bias related to ignoring the association between longitudinal and time-to-event processes. There are serveral extensions of joint models for longitudinal data and survival, for example, including competing risk survival data (Elashoff et al., 2008) or multiple longitudinal outcomes (Ibrahim et al., 2004). In the analysis of recurrent and terminal events, joint frailty models enabled dependence between repeated and terminal events processes via a common term called "frailty" (Liu et al., 2004; Rondeau et al., 2007). An example of joint analysis of recurrent events and longitudinal data is a marginal model considering overdispersion (Efendi et al., 2013). Few approaches were proposed for the joint modeling of the three processes (Liu et al., 2008; Liu and Huang, 2009). In particular, Liu and Huang (2009), motivated by a study on HIV, focused on the associations between the CD4 cell counts and the intensity of opportunistic disease and the effect of the processes on death. The estimation was carried out through a Gaussian quadrature technique, assuming piecewise constant baseline hazard functions.

Tumor size metrics are often skewed due to the fact that lesions under a treatment tend to diminish to such an extension that an investigator is not able to determine their size. In order to include such a quantification limit, left-censoring can be taken into account in a model for a biomarker. Indeed, it has been shown that consideration of a detection limit reduces the bias of estimated parameters in comparison to "naïve" methods (e.g., replacing undetected values by the detection limit)(Lyles et al., 2000; Jacqmin-Gadda et al., 2000).

The aim of this work is to give a general formulation of the trivariate joint model for a left-censored longitudinal outcome, recurrent events and survival data, and to propose a semi-parametric penalized likelihood method for the smooth estimation of the hazard functions. We propose a statistical tool for marginal dynamic predictions of death and adapt measures of predictive accuracy. In a simulation study, we compare this model to standard bivariate joint models: for longitudinal data and a terminal event (Wulfsohn and Tsiatis, 1997) and for recurrent events and a terminal event (Rondeau et al., 2007) in terms of estimation and prediction accuracy. This work was motivated by a phase III clinical trial FFCD 2000–05 (Fédération Francopohone de Cancérologie Digestive) of 410 patients with advanced colorectal cancer (Ducreux et al., 2011). We compare the models applied to the data in terms of predictive abilities in order to ascertain whether continuous tumor size metric and appearances of new lesions can better predict OS than the categorical criteria.

The rest of this article is organized as follows. Section 2 describes the trivariate joint model. Section 3 introduces dynamic predictions for joint models and Section 4 presents predictive accuracy measures. In Section 5, we provide a simulation study and in Section 6 we apply the method for the clinical trial FFCD 2000–05. Finally, Section 7 concludes the article.

2. Proposed Trivariate Joint Model for Left-Censored Longitudinal Data, Recurrent Events and a Terminal Event

2.1. *Model*

The proposed model combines the joint model for recurrent events and a terminal event with the joint model for longitudinal data and a terminal event (see Web Appendix A).

For subject i ($i \in \{1, \ldots, N\}$) we define the time of terminal event (death) T_i^* and the censoring time C_i with $\delta_i = I_{\{T_i^* \leq C_i\}}$ indicating if the event time was observed before the censoring time (I denotes indicator function). Let T_{ij}^* denote the jth recurrent time for subject i ($j \in \{1, \ldots, r_i\}$) (appearance of new lesions). The recurrent follow-up times are $T_{ij} = \min(T_{ij}^*, C_i, T_i^*)$ and $\delta_{ij} = I_{\{T_{ij} = T_{ij}^*\}}$ is the indicator of recurrent events. Similarly, the last follow-up time, time of terminal event, is defined by $T_i = \min(T_i^*, C_i)$. The n_i-vector of left-censored repeated measurements $\boldsymbol{y}_i = \{y_i(t_{ik}), k = 1, \ldots, n_i\}$ (tumor size) consist of censored outcomes $\boldsymbol{y}_i^c$ of length n_i^c (if all measures are observed, then $n_i^c = 0$), measures that are below a threshold s and of completely observed measures, n_i^o-vector $\boldsymbol{y}_i^o$. We assume continuous recurrent, terminating and censoring processes, and in a small interval $[t, t + dt]$ the terminal event occurs first. Measurements $y_i(t_{ik})$ and recurrent events T_{ij} can be possibly observed at the same moments, for example, patient visits. Neither T_{ij}^* nor $y_i(t_{ik})$ can be observed after T_i^*. However, the right-censoring does not interrupt the processes, they are simply no longer observed. We assume independent right-censoring, i.e., the intensities of recurrent events and terminal event processes of patient i do not change after C_i (Rondeau et al., 2007).

To analyze jointly a left-censored longitudinal outcome Y, recurrences of an event and a terminal event, we define a trivariate model as follows:

$$\begin{cases} Y_i(t) = m_i(t) + \epsilon_i(t) = \boldsymbol{X}_{li}(t)^\top \boldsymbol{\beta}_l + \boldsymbol{Z}_i(t)^\top \boldsymbol{b}_i + \epsilon_i(t) \\ r_{ij}(t|v_i, \boldsymbol{b}_i) = r_0(t)\ \mathrm{e}^{v_i + \boldsymbol{X}_{ri}{}^\top \boldsymbol{\beta}_r + g(\boldsymbol{b}_i, \boldsymbol{\beta}_l, \boldsymbol{Z}_i(t), \boldsymbol{X}_{li}(t))^\top \boldsymbol{\eta}_r} \\ \lambda_i(t|v_i, \boldsymbol{b}_i) = \lambda_0(t)\ \mathrm{e}^{\alpha v_i + \boldsymbol{X}_{ti}{}^\top \boldsymbol{\beta}_t + h(\boldsymbol{b}_i, \boldsymbol{\beta}_l, \boldsymbol{Z}_i(t), \boldsymbol{X}_{li}(t))^\top \boldsymbol{\eta}_t} \end{cases} \tag{1}$$

where $r_{ij}(\cdot)$ and $\lambda_i(\cdot)$ are the hazard functions of recurrent and terminal events, respectively, with $r_0(\cdot)$ and $\lambda_0(\cdot)$ their baseline hazard functions. The covariates for the random effects are $\boldsymbol{Z}_i(t)$ and the covariates for the fixed effects are $\boldsymbol{X}_{li}(t)$, $\boldsymbol{X}_{rij}$, and $\boldsymbol{X}_{ti}$. The regression coefficients $\boldsymbol{\beta}_l$, $\boldsymbol{\beta}_r$, and $\boldsymbol{\beta}_t$ correspond to the fixed effects of the respective parts of the model. The observed longitudinal outcome $Y_i(t)$ is represented by its true value $m_i(t)$ and the measurement error $\epsilon_i(t)$. The measurements errors are independent and follow $\mathcal{N}(0, \sigma_\epsilon^2)$.

Let $\boldsymbol{u}_i$ represent the vector of the random effects $\boldsymbol{u}_i = (\boldsymbol{b}_i^\top, v_i)^\top$ of dimensions $q = q_1 + 1$ and we assume that its distribution is a multivariate normal $\mathcal{N}(\mathbf{0}, \mathbf{B})$ such that $\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \sigma_v^2 \end{pmatrix}$. The q_1-vector of the random effects $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{B}_1)$ represents the individual variability of the biomarker's trajectory as well as the association with recurrent events and the terminal event. The link between the longitudinal biomarker and the recurrent events, and the biomarker and death is explained partly by the random effects $\boldsymbol{b}_i$ in the link functions $g(\cdot)$ and $h(\cdot)$ and by frailty terms v_i. The within-subject correlation of the recurrent events is described by v_i assumed to be independent from the random effects $\boldsymbol{b}_i$.

The functions $g(\boldsymbol{b}_i, \boldsymbol{\beta}_l, \boldsymbol{Z}_i(t), \boldsymbol{X}_{li}(t))$ and $h(\boldsymbol{b}_i, \boldsymbol{\beta}_l, \boldsymbol{Z}_i(t), \boldsymbol{X}_{li}(t))$ can be, for example, directly $\boldsymbol{b}_i$, the biomarker's current level $m_i(t)$ and/or slope $\partial m_i(t)/\partial t$. The coefficients $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_r$ determine the association's strength. The structure of the dependence is chosen a priori and may influence the model in terms of fit and predictive accuracy. The choice should be done with caution and be relevant to the interpretation it imposes on the relationships. In the framework of the joint models for longitudinal and survival data this issue has already been discussed (Séne et al., 2013; Rizopoulos, 2012).

The time of recurrence is usually considered as the time from the beginning of the study (calendar timescale) or the time from the previous recurrence (gap timescale). As the models considered here are motivated by a randomized clinical trial, we adopt the calendar timescale for recurrent and terminal events and the time of origin is the time of randomization.

2.2. *Maximum Likelihood Estimation (MLE)*

For subject i the observed data is $\{\boldsymbol{y}_i, \boldsymbol{T}_i^r, \boldsymbol{\delta}_i^r, T_i, \delta_i\}$, where $\boldsymbol{T}_i^r = \{T_{ij}, j = 1, \ldots, r_i\}$ and $\boldsymbol{\delta}_i^r = \{\delta_{ij}, j = 1, \ldots, r_i\}$. Using the assumption that the three processes are independent of each other given $\boldsymbol{u}_i$, the joint individual marginal likelihood can be written as:

$$L_i(\boldsymbol{\theta}) = \int_{\boldsymbol{u}_i} \prod_{k=1}^{n_i} [f_{Y|\boldsymbol{u}_i}(y_i(t_{ik})|\boldsymbol{b}_i;\boldsymbol{\theta})] \prod_{j=1}^{r_i} [f_{T^r|\boldsymbol{u}_i}(T_{ij}, \delta_{ij}|\boldsymbol{u}_i;\boldsymbol{\theta})]$$
$$\times f_{T^t|\boldsymbol{u}_i}(T_i, \delta_i|\boldsymbol{u}_i;\theta) f_{\boldsymbol{u}_i}(\boldsymbol{u}_i;\boldsymbol{\theta})\mathrm{d}\boldsymbol{u}_i$$

where the parameters to estimate are $\boldsymbol{\theta} = (\boldsymbol{\beta}_l, \boldsymbol{\beta}_r, \boldsymbol{\beta}_t, \sigma_\epsilon, r_0(\cdot), \lambda_0(\cdot), \alpha, \mathbf{B})^\top$. Functions $f_{Y|\boldsymbol{u}_i}(\cdot|\cdot)$, $f_{T^r|\boldsymbol{u}_i}(\cdot|\cdot)$, $f_{T^t|\boldsymbol{u}_i}(\cdot|\cdot)$ are density functions of longitudinal outcomes, recurrent events and the terminal event, respectively, conditional on $\boldsymbol{u}_i$ of density $f_{\boldsymbol{u}_i}(\cdot)$. To consider the left-censoring, the cumulative distribution function $F_{Y|\boldsymbol{u}_i}(\cdot)$ is used (Lyles et al., 2000):

$$\begin{aligned} f_{Y|\boldsymbol{u}_i}(\boldsymbol{y}_i|\boldsymbol{b}_i;\boldsymbol{\theta}) &= f_{Y|\boldsymbol{u}_i}(\boldsymbol{y}_i^o|\boldsymbol{b}_i;\boldsymbol{\theta}) \times \mathbb{P}(\boldsymbol{y}_i^c < s|\boldsymbol{b}_i;\boldsymbol{\theta}) \\ &= \prod_{k=1}^{n_{io}} f_{Y|\boldsymbol{u}_i}(y_i(t_{ik})|\boldsymbol{b}_i;\boldsymbol{\theta}) \prod_{k=1}^{n_{ic}} F_{Y|\boldsymbol{u}_i}(s|\boldsymbol{b}_i;\boldsymbol{\theta}). \end{aligned}$$

For simplicity, we denote $g(t) = g(\boldsymbol{b}_i, \boldsymbol{\beta}_l, \boldsymbol{Z}_i(t), \boldsymbol{X}_{li}(t)))$, $h(t) = h(\boldsymbol{b}_i, \boldsymbol{\beta}_l, \boldsymbol{Z}_i(t), \boldsymbol{X}_{li}(t)))$, $r_{ij}(t)$=$r_0(t)\mathrm{e}^{\boldsymbol{X}_{rij}{}^\top \boldsymbol{\beta}_r}$ and $\lambda_i(t)$=$\lambda_0(t)\mathrm{e}^{\boldsymbol{X}_{ti}{}^\top \boldsymbol{\beta}_t}$. The individual contribution to the likelihood can be written as (details in Web Appendix A):

$$\begin{aligned} L_i(\boldsymbol{\theta}) = \; & \frac{1}{(\sqrt{2\pi}\sigma_\epsilon)^{n_i^o}} (\lambda_i(T_i))^{\delta_i} \prod_{j=1}^{r_i} (r_{ij}(T_{ij}))^{\delta_{ij}} \\ & \int_{\boldsymbol{u}_i} \Bigg\{ \prod_{k=1}^{n_i^c} \Phi\left(\frac{s - m_i^c(t_{ik})}{\sigma_\epsilon}\right) \exp\Bigg[-\frac{||\boldsymbol{y}_i^o - \boldsymbol{m}_i^o||^2}{2\sigma_\epsilon^2} \\ & - \mathrm{e}^{\alpha v_i} \int_0^{T_i} \lambda_i(t)\mathrm{e}^{h(t)^\top \boldsymbol{\eta}_t}\mathrm{d}t + \delta_i h(t)^\top \boldsymbol{\eta}_t\Bigg] \\ & \times \prod_{j=1}^{r_i} \exp\left[-\mathrm{e}^{v_i} \int_{T_{i(j-1)}}^{T_i} r_{ij}(t)\mathrm{e}^{g(t)^\top \boldsymbol{\eta}_r}\mathrm{d}t + \delta_{ij} g(t)^\top \boldsymbol{\eta}_r\right] \\ & \times \mathrm{e}^{(N_i^r(T_i)+\delta_i\alpha)v_i} \frac{1}{(2\pi)^{q/2}} |\mathbf{B}|^{-1/2} \mathrm{e}^{-\frac{\boldsymbol{u}_i{}^\top \mathbf{B}^{-1} \boldsymbol{u}_i}{2}} \Bigg\}\mathrm{d}\boldsymbol{u}_i \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cdf, $\boldsymbol{m}_i^o = \{m_i^o(t_{ik}), j = 1, \ldots, n_i^o\}$ and $\boldsymbol{m}_i^c = \{m_i^c(t_{ik}), j = 1, \ldots, n_i^c\}$ are the true biomarker levels for observed and censored measurements, respectively. Finally, $N_i^r(t)$ is the observed number of recurrent events until t.

The inference of the model is based on the penalized maximum likelihood estimation using the Marquardt algorithm (Marquardt, 1963). The baseline risk functions are approximated by m cubic M-splines with Q knots:

$$\tilde{r}_0(\cdot) = \sum_{i=1}^{m} \zeta_{ri} M_i(\cdot),\; \tilde{\lambda}_0(\cdot) = \sum_{i=1}^{m} \zeta_{ti} M_i(\cdot),\; m = Q + 2.$$

Penalization of the log-likelihood is performed to obtain smooth estimation of $r_0(t)$ and $\lambda_0(t)$: spline coefficients, ζ_{ri} and ζ_{ti} for the baseline hazard for recurrent events and for death, respectively. The point-wise 95% confidence intervals are approximated on the basis of splines the penalized log-likelihood. The penalization of the log-likelihood is performed to obtain smooth estimation of baseline hazard functions:

$$pl(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - \kappa_1 \int_0^\infty r_0''(t)^2 \mathrm{d}t - \kappa_2 \int_0^\infty \lambda_0''(t)^2 \mathrm{d}t,$$

where $l(\boldsymbol{\theta}) = \sum_{i=1}^N \mathrm{log} L_i(\boldsymbol{\theta})$ and κ_1 and κ_2 two positive smoothing parameters chosen using an approximate cross-validation criterion applied separately to the submodels of recurrent events (shared frailty model) and death (Cox model). The multivariate integrals are approximated using the non-adaptive Gauss–Hermite quadrature with 20 nodes. Once the model reaches convergence, the standard errors are calculated from the diagonal elements of the inverse Hessian matrix of the penalized log-likelihood.

3. Dynamic Predictions for Joint Models

In dynamic predictions the interest is to the calculate probability that an event occurs in a finite horizon $[t, t + w]$ given the covariates and that the individual did not experience the event before prediction time t. In the joint model approach this probability is additionally conditioned on the individual's history, i.e., the past recurrent events and/or repeated measurements. Mauguen et al. (2013) developed predictions in the framework of joint models for recurrent and terminal events and Proust-Lima and Taylor (2009) and Rizopoulos (2011) for the joint models for longitudinal data and a terminal event.

In the context of model (1), the complete recurrence history given that subject i had J events until time t is given by $\mathcal{H}_i^J(t) = \{N_i^r(t) = J, T_{i1}^* < \ldots < T_{iJ}^* \le t\}$ and the complete biomarker history given that he/she had K repeated measurements is $\mathcal{Y}_i(t) = \{y_i(t_{ik}), t_{i1} < \ldots < t_{iK} < t\}$. Let define $\mathcal{F}_i(t) = \{\mathcal{H}_i^J(t), \mathcal{Y}_i(t)\}$ and $\boldsymbol{X}_i$ all the covariates included in the model. We use marginal predictions, where the conditional probability of the event in $[t, t + w]$ is integrated over the distribution of random effects:

$$\begin{aligned}
P_{[t,t+w]}(\boldsymbol{\theta}) &= \mathbb{P}(T_i^* \le t + w | T_i^* > t, \mathcal{F}_i(t), \boldsymbol{X}_i; \boldsymbol{\theta}) \\
&= \int_{\boldsymbol{u}_i} \mathbb{P}(T_i^* \le t + w | T_i^* > t, \mathcal{F}_i(t), \boldsymbol{X}_i, \boldsymbol{u}_i; \boldsymbol{\theta}) \times f(\boldsymbol{u}_i | T_i^* > t, \mathcal{F}_i(t), \boldsymbol{X}_i; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{u}_i \\
&= \frac{\int_{\boldsymbol{u}_i} [S_i^t(t|\boldsymbol{X}_{ti}, \boldsymbol{u}_i; \boldsymbol{\theta}) - S_i^t(t + w|\boldsymbol{X}_{ti}, \boldsymbol{u}_i; \boldsymbol{\theta})] \exp(J(v_i + g(t)^\top \boldsymbol{\eta}_r)) S^r_{i(J+1)}(t|\boldsymbol{X}_{riJ}, v_i; \boldsymbol{\theta}) f(\mathcal{Y}_i(t)|\boldsymbol{X}_{liK}, \boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{u}_i) \mathrm{d}\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i} S_i^t(t|\boldsymbol{X}_{ti}, \boldsymbol{u}_i; \boldsymbol{\theta}) \exp(J(v_i + g(t)^\top \boldsymbol{\eta}_r)) S^r_{i(J+1)}(t|\boldsymbol{X}_{riJ}, v_i; \boldsymbol{\theta}) f(\mathcal{Y}_i(t)|\boldsymbol{X}_{liK}, \boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{u}_i) \mathrm{d}\boldsymbol{u}_i}
\end{aligned}$$

The estimates of the predicted probabilities can be computed by replacing the parameters $\boldsymbol{\theta}$ by their estimates $\widehat{\boldsymbol{\theta}}$ and the confidence intervals by the Monte Carlo method. Percentiles of the distributions of the predicted probabilities are calculated from large number of parameter vectors drawn from the asymptotic normal distribution of $\boldsymbol{\theta}$, $\mathcal{N}(\widehat{\boldsymbol{\theta}}, \widehat{\mathrm{Var}}(\widehat{\boldsymbol{\theta}}))$. These intervals consider the uncertainty from the parameters estimates, including the variances of the random effects. Additional uncertainty from the random effects via the empirical Bayes estimates can be incorporated as in Rizopoulos, 2011.

4. Predictive Accuracy Measures

4.1. *EPOCE*

Predictive accuracy of joint models can be evaluated using the expected prognostic observed cross-entropy (EPOCE), a measure of prognostic information based on prognostic density and adapted to right-censored data (Commenges et al., 2012). It represents the average measure of the loss function defined by the Kullback–Leibler distance between the true prognostic density and the prognostic density derived from the joint model. The lower this risk, the better the predictive model. The prognostic density $f_{T|\mathcal{F}(t),T^* \ge t}$ is the conditional density of the time of event $T = \min(T^*, C)$, given the history of repeated measures $\mathcal{F}(t)$ until t and that the event did not occur before t. The Approximated Cross-Validated Prognosis Observed Loss (CVPOL$_a$) used as an estimator of EPOCE is defined as:

$$\mathrm{CVPOL}_a(t) = -\frac{1}{N_t} \sum_{i=1}^{N_t} F_i(\widehat{\theta}, t) + N \mathrm{Trace}(\boldsymbol{H}^{-1} \boldsymbol{K}_t) \qquad (2)$$

where N_t is the number of subjects still at risk at time t, $\boldsymbol{H}$ is the hessian matrix of the joint log-likelihood, $\boldsymbol{K}_t = \frac{1}{N_t(N-1)} \sum_{i=1}^N I_{\{T_i \ge t\}} \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{d}}_i^\top$, $\widehat{\boldsymbol{v}}_i$ being the gradient of the individual contribution to the conditional log-likelihood using the individual's history until t and $\widehat{\boldsymbol{d}}_i^\top$, the gradient of the individual contribution to the conditional log-likelihood using all the history. The individual contribution to the conditional log-likelihood of a terminal event is:

$$\begin{aligned}
F_i(\widehat{\boldsymbol{\theta}}, t) = \ln\Bigg[\int_{\boldsymbol{u}_i} & f(\mathcal{F}(t)|\boldsymbol{u}_i; \widehat{\boldsymbol{\theta}}) \lambda_i(T_i|\boldsymbol{X}_{ti}, \boldsymbol{u}_i; \widehat{\boldsymbol{\theta}})^{\delta_i} \\
& \times S_i^t(T_i|\boldsymbol{X}_{ti}, \boldsymbol{u}_i; \widehat{\boldsymbol{\theta}}) f(\boldsymbol{u}_i) \mathrm{d}\boldsymbol{u}_i \Bigg] - \ln\Bigg[\int_{\boldsymbol{u}_i} f(\mathcal{F}(t)|\boldsymbol{u}_i; \widehat{\boldsymbol{\theta}}) \\
& \times S_i^t(T_i|\boldsymbol{X}_{ti}, \boldsymbol{u}_i; \widehat{\boldsymbol{\theta}}) f(\boldsymbol{u}_i) \mathrm{d}\boldsymbol{u}_i \Bigg]
\end{aligned}$$

The estimator has the advantage that it may be computed on the same data as used in the model, as the second term of the sum corrects for the over-optimism. Thus, it can be easily calculated without reestimating the model k times as for the calculation of prediction error (see Section 4.2). Moreover, a 95% tracking interval can be computed so that joint models can be statistically compared for a relevant inference. The term "tracking" is used here rather than "confidence" as the difference in CVPOL$_a$ changes with N.

4.2. *Prediction Error*

Let $\widehat{S}(t + w|t) = 1 - P_{[t,t+w]}(\widehat{\boldsymbol{\theta}})$ be the predicted value of the conditional survival function at time $t + w$. The mean squared error of prediction is the expectation of squared residuals $\mathbb{E}[I_{\{T_i^* > t+w\}} - \widehat{S}(t + w|t)]^2$ and its estimation that considers the independent right-censoring can be performed by weighting the observations according to the probability that they are observed. The inverse probability of censoring weighted error estimator (data-based Brier score), is defined by (Gerds and Schumacher, 2006; Mauguen et al., 2013):

$$\begin{aligned}
\widehat{\mathrm{BS}}(t, w) = \frac{1}{N_t} \sum_{i=1}^{N_t} & \left[I_{\{T_i^* > t+w\}} - (1 - P_{[t,t+w]}(\widehat{\boldsymbol{\theta}})) \right]^2 \\
& \times \left(\frac{I_{\{T_i^* \le t+w\}} \delta_i}{\widehat{G}(T_i^*)/\widehat{G}(t)} + \frac{I_{\{T_i^* > t+w\}}}{\widehat{G}(t + w)/\widehat{G}(t)} \right) \qquad (3)
\end{aligned}$$

with $\widehat{G}(t)$ being the Kaplan–Meier estimate of the survival function of the censoring distribution at t. The second term of the product considers censored observations as a particular case of missing data. The internal validation can be performed using a k-fold cross-validation to correct for over-optimism. For each left-out partition the predictions are based on the estimates resulted from the joint model built on the remaining $k - 1$ partitions.

5. Simulations

5.1. *Scenario*

The proposed trivariate joint model was evaluated in terms of performance of the estimators by a simulation study reflecting the FFCD 2000–05 trial. The datasets were generated using the trivariate model with random intercept and slope with the link functions $h(t)$ and $g(t)$ being directly $\boldsymbol{b}_i$:

$$\begin{cases}
Y_i(t) = \beta_0 + \beta_1 t + \beta_2 X_{1i} + b_{0i} + b_{1i} t + \epsilon_i(t) \\
r_{ij}(t|v_i, \boldsymbol{b}_i) = r_0(t)\ \mathrm{e}^{v_i + \beta_3 X_{1i} + \eta_{r1} b_{0i} + \eta_{r2} b_{1i}} \\
\lambda_i(t|v_i, \boldsymbol{b}_i) = \lambda_0(t)\ \mathrm{e}^{\alpha v_i + \beta_4 X_{2i} + \eta_{t1} b_{0i} + \eta_{t2} b_{1i}}
\end{cases} \qquad (4)$$

The relationship between the process of recurrent events and terminal event was positive ($\sigma_v = 0.8,\ \alpha = 2.6$), the random intercept of the biomarker had higher variance than the slope and they were negatively correlated. Thus, the variance–covariance matrix as follows:

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} = \begin{pmatrix} \sigma_0^2 & \sigma_{01} & 0 \\ \sigma_{10} & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{pmatrix} = \begin{pmatrix} 1.5^2 & -0.5 & 0 \\ -0.5 & 0.8^2 & 0 \\ 0 & 0 & 0.8^2 \end{pmatrix}$$

with equal strength of association both for the relationships of the biomarker and recurrent events ($\eta_{r1} = 0.2,\ \eta_{r2} = 0.2$) and the terminal event ($\eta_{t1} = 0.5,\ \eta_{t2} = 0.5$).

The purpose of this study was to verify if the proposed method estimated well the parameters for data generated within the trivariate setting. Secondly, we wanted to compare it in terms of performance and predictive accuracy to the standard joint models for the same datasets. Finally, we compared the trivariate model with the left-censored biomarker with a corresponding "naïve" trivariate model in which the undetectable measures were replaced by the value of the threshold s (limit of detection).

5.2. *Data Generation*

We generated 500 datasets of 400 subjects and for each subject, we firstly generated a vector of random effects (b_{0i}, b_{1i}, v_i) from $\mathcal{N}(\mathbf{0}, \mathbf{B})$. The time-independent binary variables X_1 and X_2 were generated from Bernoulli distribution with $p = 0.5$. The measurement error $\epsilon_i(t_{ij})$ followed $\mathcal{N}(0, \sigma_\epsilon^2 = 1.25)$. We fixed the biomarker intercept $\beta_0 = 3.0$, the slope $\beta_1 = 0.5$ and the coefficient for X_{1i}, $\beta_2 = 0.5$, the effect X_{1i} on recurrent events, $\beta_3 = 0.5$ and the effect of X_{2i} on the terminal event, $\beta_4 = 0.1$. The exponential death times T_i were generated with $\lambda_0(t) = 1.5$. If $T_i \leq C_i$ then the terminal event is the death time ($\delta^*_{i,t} = 1$ and $T^*_i = T_i$) and the censoring time otherwise ($\delta^*_{i,t} = 0$ and $T^*_i = C_i$). A fixed right-censoring variable $C_i = 5.5$ was set to obtain around 20% of censored subjects. The exponential recurrent gap times S_{ij} representing times of a subject's visits were generated with $r_0(t) = 2.0$. As the models were estimated using calendar time, the recurrent times were $T_{ij} = \min(T_i, C_i, \sum_{k=1}^{j} S_{ik})$ and the data generation continued until $T_{ij} < T^*_i$ or if the number of recurrent events exceeded 6. The period between two consecutive biomarker measurements was fixed at 0.2. The repeated measurements stopped when the calendar time t exceeded T_i or the number of measurements reached 20. We set $s = -0.4$ in order to obtain around 4% of left-censored measures.

5.3. *Results*

The results are presented in detail in the Web Appendix Table C.1. There were around 6.5 observed repeated measurements of biomarker and around 1.7 recurrent events per subject. Using the trivariate model with the left-censored biomarker, the regression coefficients were well estimated with their coverage probabilities close to the nominal level of 95%. The association parameters estimations were as well satisfying and only small biases (around 5%) were observed. For the random effects, the model estimated the variance–covariance well and only the standard errors of the covariance σ_{01} were underestimated. In comparison, the bivariate joint model with the left-censored biomarker resulted in less accurate estimations, especially for the coefficient β_1 related to the fixed effect of time and the link parameters η_{t1} and η_{t2}. On the other hand, the bivariate model with the recurrent events resulted in good estimations of the parameters. However, it should be noted that the coefficients β_3 and β_4 in the standard and the trivariate joint models are conditional on different number of random effects and thus the interpretation of their results should be made with caution. In this example, the comparison of the estimation aimed to show expected differences in estimations to illustrate the potential advantages of the trivariate model for the analyzed framework and not to validate the standard models.

To compare the left-censoring methods, using the same datasets as those generated with model (4), we also performed estimations within the "naïve" trivariate model (the last column of the Table C.1). We found that using this approach we obtained more biased estimations of most of the association parameters compared to the proposed trivariate model, in particular for the variance of the measurement error.

Web Appendix Figure C.1 shows the results of CVPOL$_a$ for the four models. The boxplots of CVPOL$_a$ distributions for different prediction times indicate that the trivariate model had better predictive accuracy than the bivariate model with the recurrent events and only slightly better predictive accuracy than the bivariate model with the left-censored biomarker. We conclude, that in this scenario, the information from the biomarker history seems to bring more predictive value for the survival than the information from the recurrent events history. The difference in CVPOL$_a$ between the "naïve" trivariate model and the model with the left-censoring was negligible probably due to small number of affected measurements (4%).

6. Application

6.1. *Data Analysis*

The randomized phase III trial, FFCD 2000–05 included 410 patients with metastatic colorectal cancer. Patients were recruited between February 2002 and February 2006 from 53 centers in France. The cut-off date for the analysis was January 1, 2007. The aim of the trial was to investigate whether a combination arm was better than the sequential administration of the same drugs in terms of repeated progressions, toxicities, and survival. The sequential arm S consisted of 5-fluorouracil and leucovorin (LV5FU2) followed by FOLFOX (LV5FU2 + oxaliplatin) and then FOLFIRI (LV5FU2 + irinotecan) while the combination arm C began directly with FOLFOX followed by FOLFIRI (for details see Web Appendix D).

Among 407 (99%) patients who started their treatment, 321 (79%) died during the follow-up and among the censored patients, the majority (61, i.e., 71%) were censored due to the end of the study. There were 426 observed occurrences of new lesions (on average 1.05 per patient) and 741 progressions according to WHO criteria (on average 1.82 per patient). There were up to 4 lesions measured during each patient's visit, followed in two dimensions (maximal diameter and its perpendicular). The longitudinal biomarker, SLD, for patient

Table 1

Parameter estimates for Model 1 (tumor progressions as recurrent events and death) and Model 2 (appearance of new lesions as recurrent events and death) fitted to the FFCD 2000–05 data set using calendar time scale

	Model 1		Model 2	
	Progressions	Terminal event	New Lesions	Terminal event
Covariate	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Treatment(C/S)	0.89 (0.73 − 1.07)	1.11 (0.65 − 1.91)	0.99 (0.79 − 1.24)	1.15 (0.68 − 1.93)
Age (60–69/<60 years)	0.92 (0.72 − 1.18)	1.22 (0.58 − 2.56)	0.77 (0.58 − 1.03)	1.08 (0.55 − 2.10)
Age (≥70/<60 years)	0.96 (0.76 − 1.20)	1.69 (0.92 − 3.11)	0.84 (0.64 − 1.09)	1.59 (0.91 − 2.78)
Sex (female/male)	0.98 (0.81 − 1.19)	0.99 (0.57 − 1.70)	0.85 (0.67 − 1.07)	0.99 (0.60 − 1.63)
Center (10–29/≥30 patients)	1.13 (0.88 − 1.44)	2.29 (1.19 − 4.41)*	1.21 (0.90 − 1.62)	2.02 (1.11 − 3.69)
Center (<10/≥30 patients)	1.19 (0.94 − 1.51)	2.73 (1.37 − 5.43)**	1.23 (0.92 − 1.63)	2.27 (1.17 − 4.40)
Baseline WHO PS (1/0)	1.18 (0.96 − 1.45)	1.60 (0.91 − 2.81)	1.15 (0.90 − 1.62)	1.61 (0.93 − 2.79)
Baseline WHO PS (2/0)	2.43 (1.84 − 3.22)***	30.80 (13.71 − 69.18)***	2.40 (1.69 − 3.41)***	26.45 (12.37 − 56.56)***
Previous resection	0.73 (0.60 − 0.90)**	0.38 (0.22 − 0.66)***	-	0.46 (0.28 − 0.73)***
Metastases localization:				
Abdominal lymphad.	1.36 (1.08 − 1.73)**	2.45 (1.27 − 4.72)**	1.55 (1.17 − 2.04)*	2.44 (1.32 − 4.53)**
Association Parameters	Estimation (SE)		Estimation (SE)	
σ_v^2	0.41 (0.06)***		0.43 (0.08)***	
α	4.08 (0.32)***		3.77 (0.41)***	

σ_v^2, variance of the frailty term; α, association parameter between the processes

$^{*}p$-value ≤ 0.05, $^{**}p$-value ≤ 0.01, $^{***}p$-value ≤ 0.001

HR, hazard ratio; CI, confidence interval; SE, standard error; PS, performance status

i during visit k ($k = 0, 1, \cdots, n_i$, 0 for baseline measurement) was defined as $\text{SLD}_{ik} = \sum_{l=1}^{n_{ik}} d_{ikl}$, where n_{ik} is the number of lesions measured and d_{ikl} is the longest diameter of lesion l. In total, there were 2517 SLD measurements observed and on average, 6.18 per patient (range [1–17]) with an average SLD 9.78 cm. On average, patients in the S arm had a slightly larger tumor size and more progressions than patients in the C arm, but the average number of appearances of new lesions was similar in both arms (Web Appendix Table D.1).

Biomarker values equal to zero correspond to lesions that were too small to measure (<1 mm). Among all the observed repeated measurements there were 4.21% of "zero" measurements which resulted in the violation of the response variable normality assumption. Thus, we considered the left-censoring for SLD and set the zero values as below the limit of detectability. Also for the normality assumption, we applied the Box–Cox transformation and using the profile likelihood of the univariate linear mixed-effects model we chose $\text{SLD}^* = (\text{SLD}^{0.3} - 1)/0.3$. The heterogeneity in the transformed longitudinal profiles is similar in both arms, significant at baseline and in time (Web Appendix Figure D.2). Thus, we decided to apply random intercept and slope $\boldsymbol{b}_i = (b_{i0}, b_{i1})^\top$ in the longitudinal part of the models.

6.2. *Joint Models*

To compare the predictive abilities on death of categorized criteria, tumor size, and appearance of new lesions we built four joint models : a reference model for recurrent events (tumor progressions according to WHO criteria) and death (Model 1), a model for recurrent events (appearance of new lesions) and death (Model 2), a model for left-censored longitudinal biomarker (SLD) and death (Model 3), and a trivariate model for left-censored longitudinal biomarker (SLD), recurrent events (appearance of new lesions) and death (Model 4).

Baseline covariates included patient's age (<60, 60–69 and ≥70 years old), sex, treatment arm (C vs. S arm), WHO Performance Status (PS) (0–2), center size (≥30 patients, 10–29 patients, <10 patients), previous adjuvant chemotherapy, previous adjuvant radiotherapy, previous resection of the primate tumor, number of metastatic sites (1 site vs. ≥2 sites), binary variables for localization of the metastases: liver, peritoneum, lymphadenopathies (abdominal and others), lungs (Web Appendix Table D.2). The selection of covariates was performed using global backward elimination. We decided to force the selection to keep treatment, age, sex, WHO PS, and center size variables given their importance in terms of interpretation. For the proposed trivariate model, given its computational complexity, we decided to begin the elimination selection with the covariates from the appropriate standard bivariate models. In all the joint models we applied 5 knots for the cubic M-splines.

6.3. *Results*

We considered 402 subjects as 5 patients did not have the measurements of the baseline covariates chosen for the models. Tables 1–3 show the results of the fitted Models 1–4.

In all the joint models the effect of the treatment arm was not significantly associated with the risk of death, progressions nor appearances of new lesions ($p > 0.05$, Wald test). However, there was a significant treatment by time interaction for SLD in both Model 3 (−0.42, SE = 0.13) and Model 4 (−0.42, SE = 0.15), adjusted on other effects. Indeed, the trajectory of the tumor size was decreasing more in the combination arm than in the sequential arm in the beginning of the trial. The effects of age and sex were not significantly

Table 2
Parameter estimates for Model 3 (transformed tumor size SLD as a longitudinal biomarker and death) fitted to the FFCD 2000–05 dataset

	SLD	Terminal event
Covariate	Estimation(SE)	HR (95% CI)
Intercept	3.10 (0.29)***	-
Time (t)	-0.32 (0.12)**	-
Treatment (C/S)	-0.25 (0.14)	0.89 (0.69 – 1.14)
Treatment (C/S)×t	-0.42 (0.15)**	-
Age (60–69/<60 years)	0.18 (0.17)	0.84 (0.61 – 1.16)
Age (≥70/<60 years)	-0.02 (0.16)	1.12 (0.83 – 1.52)
Sex (female/male)	0.24 (0.14)	0.91 (0.70 – 1.18)
Center (10–29/≥30 pat.)	-0.15 (0.17)	0.95 (0.69 – 1.31)
Center (<10/≥30 pat.)	-0.11 (0.16)	0.98 (0.73 – 1.32)
Baseline WHO PS (1/0)	-0.20 (0.15)	1.01 (0.76 – 1.33)
Baseline WHO PS (2/0)	0.40 (0.21)	3.11 (2.14 – 4.50)***
Previous resection	-0.72 (0.14)***	0.51 (0.39 – 0.67)***
Metastatic sites (≥2/1)	-0.49 (0.13)***	-
Metastases localization:		
Liver	0.80 (0.22)***	-
Peritoneum	-0.42 (0.18)*	-
Abdominal lymphad.	0.41 (0.17)*	1.48 (1.09 – 2.02)*
Association Parameters	Estimation(SE)	
η_{t1}	0.43 (0.06)***	
η_{t2}	0.05 (0.11)	
Matrix **B** components		
Var(Intercept)	1.56 (0.06)***	
Var(Slope)	1.01 (0.07)***	
cov(Intercept, Slope)	-0.26 (0.09)*	
σ_ϵ^2	0.95 (0.02)**	

η_{t1}, η_{t2}, link parameters between the processes
σ_ϵ^2, variance of the measurement error
$^{*}p$-value ≤ 0.05, $^{**}p$-value ≤ 0.01, $^{***}p$-value ≤ 0.001
HR, hazard ratio; CI, confidence interval; SE, standard error; PS, performance status

influencing the response variables of the models. Smaller centers had an increased risk of death in the joint models together with progressions and new lesions. In all the models, we observed a strong effect of the WHO performance status 2 on the risk of death, progressions and new lesions. In terms of SLD, the performance status 2 was associated with larger tumor size only in Model 4 (0.45, SE = 0.21), adjusted for other effects. There was a strong positive association between tumor progressions and death in Model 1 ($\sigma_v^2 = 0.41$, $\alpha = 4.08$) and appearances of new lesions and death in Model 2 ($\sigma_v^2 = 0.43$, $\alpha = 3.77$). The individual variability explained by the random intercept was significantly associated with death in Model 3 ($\eta_{t1} = 0.43$). Similar associations were found in Model 4, but the association between the biomarker and death via the random intercept was more important than in Model 3 with $\eta_{t1} = 0.75$. In Model 4, the association explained by the random effects between the biomarker and appearance of new lesions was not significant, this was not possible to investigate in the chosen bivariate joint models. A sensitivity analysis for the link functions in Model 4 showed robustness for the explanatory variables but also the importance of choice of these functions.

Finally, we compared the models in terms of predictive abilities. EPOCE (2), estimated in a window of 2.2 years (0.3–2.5 years after randomization), is summarized in Figure 2 with the values of CVPOL$_a$ (top graph), the differences in CVPOL$_a$ and the 95% tracking intervals (bottom graph). Model 4 seems to have a better predictive accuracy than the standard joint models until around 2 years. After this time, Model 1, Model 2, and Model 4 had similar predictive abilities, while Model 3 performed worse. Based on the tracking interval graph of the differences between Model 4 and the other models we observed that it was substantially better than Model 2 in the first 2 years and better than Model 1 until around 1 year of treatment, since 0 was below the boundaries of the 95% tracking intervals. However, Model 3 did not perform significantly worse than Model 4 in the beginning of the trial.

A 10-fold cross-validation error of prediction was calculated and is displayed in Figure 1. For the same time points as in Figure 2, we performed dynamic predictions for a prediction time $t = 1.0$ with varying window w from 0.1 to 1.5 and for each time point we calculated the prediction error according to the equation (3). The trivariate Model 4, again, had the best prediction ability and the Model 1 based on progressions had higher prediction error than the other models. We performed additional evaluations of prediction error for different prediction times $t = 0.5$, $t = 1.5$, and $t = 2$ with the horizon up to 2.5 years (the results are presented in the Web Appendix Figures E.1–4). For all the different prediction times we obtained similar results, with the smallest prediction error for Model 4. However, for a short time of prediction ($t = 0.5$) we observed that Model 2 had the prediction error comparable to Model 4 but for a longer time of prediction ($t = 2$), it was Model 3 that had the prediction error similar to Model 4. Thus, in a period of time close to the randomization, the history of tumor size seems to be of high predictive value for OS, whereas after a certain time, the history of new lesions occurrence seems to carry more predictive information for OS.

7. Discussion

Tumor response criteria (RECIST, WHO) allow a standardized assessment and comparison of treatments among patients and trials. However, their limitations due to the categorization of continuous process of change in tumor size provide the arguments to improve these tools (Royston et al., 2006). Joint models can give better assessment of the treatment effect by including the association between the processes of tumor response (appearance of new lesions, tumor size) and OS. For this purpose, we proposed a trivariate joint model for left-censored longitudinal data, recurrent events and a terminal event and validated its usefulness by a simulation study in which we compared our method with reduced joint models in terms of results accuracy and predictive abilities. For the trivariate model we extended the dynamic probabilities of death, the measure of predictive accuracy (CVPOL$_a$), and prediction error (Brier Score). For both measures, the cross-validation corrects for over-optimism (approximated in CVPOL$_a$ and k-fold in the Brier Score), but CVPOL$_a$ has the asset to be obtained directly without repeating k times the estimation of the model.

Table 3

Parameter estimates for Model 4 (appearance of new lesion as recurrent events, transformed tumor size SLD as a longitudinal biomarker and death) fitted to the FFCD 2000–05 dataset

	SLD	New Lesions	Terminal event
Covariate	Estimation (SE)	HR (95% CI)	HR (95% CI)
Intercept	2.90 (0.29)***	-	-
Time (t)	−0.35 (0.13)**	-	-
Treatment (C/S)	−0.20 (0.14)	0.96 (0.75 – 1.21)	1.02 (0.64 – 1.61)
Treatment (C/S)×t	−0.42 (0.15)**	-	-
Age (60–69/<60 years)	0.23 (0.18)	0.75 (0.56 – 1.02)	1.04 (0.57 – 1.87)
Age (≥70/<60 years)	0.02 (0.16)	0.82 (0.61 – 1.09)	1.40 (0.79 – 2.49)
Sex (Female/Male)	0.27 (0.14)	0.86 (0.67 – 1.10)	1.02 (0.63 – 1.65)
Center (10–29/≥ 30 patients)	−0.10 (0.17)	1.06 (0.77 – 1.44)	1.26 (0.68 – 2.33)
Center (<10/≥ 30 patients)	−0.04 (0.16)	1.12 (0.84 – 1.49)	1.38 (0.77 – 2.51)
Baseline WHO PS (1/0)	−0.14 (0.15)	1.16 (0.89 – 1.51)	1.51 (0.85 – 2.68)
Baseline WHO PS (2/0)	0.45 (0.21)*	2.15 (1.44 – 3.21)***	10.22 (3.68 – 28.40)***
Previous resection	−0.64 (0.15)***	-	0.46 (0.31 – 0.70)***
Metastatic sites (≥2/1)	−0.45 (0.13)***	-	-
Metastases localization:			
Liver	0.78 (0.22)***	-	-
Peritoneum	−0.45 (0.18)*	-	-
Abdominal lymphad.	0.45 (0.17)**	1.62 (1.20 – 2.18)**	2.47 (1.28 – 4.75)**
Association Parameters*	Estimation (SE)		Estimation (SE)
η_{r1}	0.09 (0.06)	Matrix **B** components	
η_{r2}	−0.07 (0.09)	Var(Intercept)	1.56 (0.06)***
η_{t1}	0.75 (0.15)***	Var(Slope)	1.02 (0.09)*
η_{t2}	−0.07 (0.20)	cov(Intercept, Slope)	−0.26 (0.08)***
α	2.50 (0.47)***	σ_v^2	0.44 (0.09)***
σ_ϵ^2	0.95 (0.02)***		

η_{r1}, η_{r2}, link parameters (biomarker and recurrences); η_{t1}, η_{t2}, link parameters (biomarker and death);
σ_v^2, variance of the frailty term; α, link parameter between recurrences and death;
σ_ϵ^2, variance of the measurement error
$^{*}p$-value ≤ 0.05, $^{**}p$-value ≤ 0.01, $^{***}p$-value ≤ 0.001
HR, hazard ratio; CI, confidence interval; SE, standard error; PS, performance status

We applied the trivariate model to a colorectal phase III clinical trial FFCD 2000–05 and compared its results and predictive abilities to those obtained using reduced joint models. The purpose of the analysis was to verify whether tumor progressions evaluated using categorical criteria (WHO) had a worse predictive effect on death than the longitudinal tumor size measurements and/or appearance of new lesions. We found that the trivariate model was the best prognostic model for OS. Thus, we conclude that using jointly the biomarker and appearances of new lesions we increase the prediction of OS than using only the categorical criteria. Moreover, the trivariate joint model enabled a thorough analysis of the treatment effect on the disease evolution and death. In particular, it dissociates whether a treatment has an effect on tumor size, appearance of new lesions, or on mortality. In the application we found a significant treatment by time interaction on tumor size only, and no significant effect on appearance of new lesions or mortality.

The common detection limit in the measures of tumor size was treated here by incorporating the left-censoring for the biomarker. The differences between the models that considered the left-censoring (Model 3 and 4) and the corresponding "naïve" models that ignored it were found not to change importantly the interpretation of the covariates (results not shown).

In the proposed model the assumed normal distribution of the random effects could be relaxed by incorporating conjugate random effects (Molenberghs et al., 2010). It would be also of interest to distinguish the associations by introducing $\boldsymbol{b}_{i,1}$ and $\boldsymbol{b}_{i,2}$, independent of each other, with the following form of the model:

$$\begin{cases} Y_i(t) = \boldsymbol{X}_{li}(t)^\top \boldsymbol{\beta}_l + \boldsymbol{Z}_{i,1}(t)^\top \boldsymbol{b}_{i,1} + \boldsymbol{Z}_{i,2}(t)^\top \boldsymbol{b}_{i,2} + \epsilon_i(t) \\ r_{ij}(t|v_i, \boldsymbol{b}_{i,1}) = r_0(t)\ \mathrm{e}^{v_i + \boldsymbol{X}_{ri}{}^\top \beta_r + g(\boldsymbol{b}_{i,1}, \boldsymbol{\beta}_l, \boldsymbol{Z}_{i,1}(t), \boldsymbol{X}_{li}(t))^\top \boldsymbol{\eta}_r} \\ \lambda_i(t|v_i, \boldsymbol{b}_{i,2}) = \lambda_0(t)\ \mathrm{e}^{\alpha v_i + \boldsymbol{X}_{ti}{}^\top \boldsymbol{\beta}_t + h(\boldsymbol{b}_{i,2}, \boldsymbol{\beta}_l, \boldsymbol{Z}_{i,2}(t), \boldsymbol{X}_{li}(t))^\top \boldsymbol{\eta}_t} \end{cases}$$

However, this model may suffer from identifiability or estimation problems.

In the process of estimation the choice of the numerical integration method was important. The non-adaptive Gauss–Hermite quadrature enables for precise results but it is highly computationally demanding. One of the alternatives would be a multivariate Gaussian quadrature (Genz and Keister,

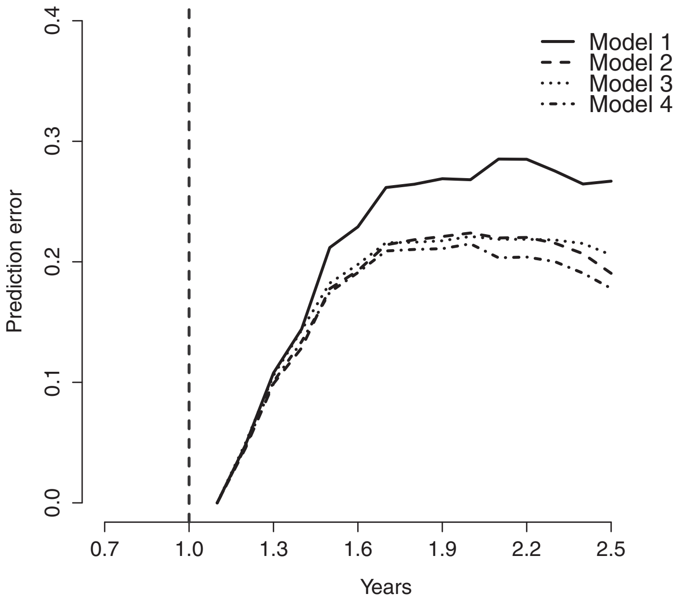

Figure 1. 10-fold cross-validated error of prediction at $t = 1$ year and varying window w from 0.1 to 1.5. Model 1—tumor progressions and death, Model 2—occurrence of new lesions and death, Model 3—SLD measure and death, Model 4—SLD measure, occurrence of new lesions and death.

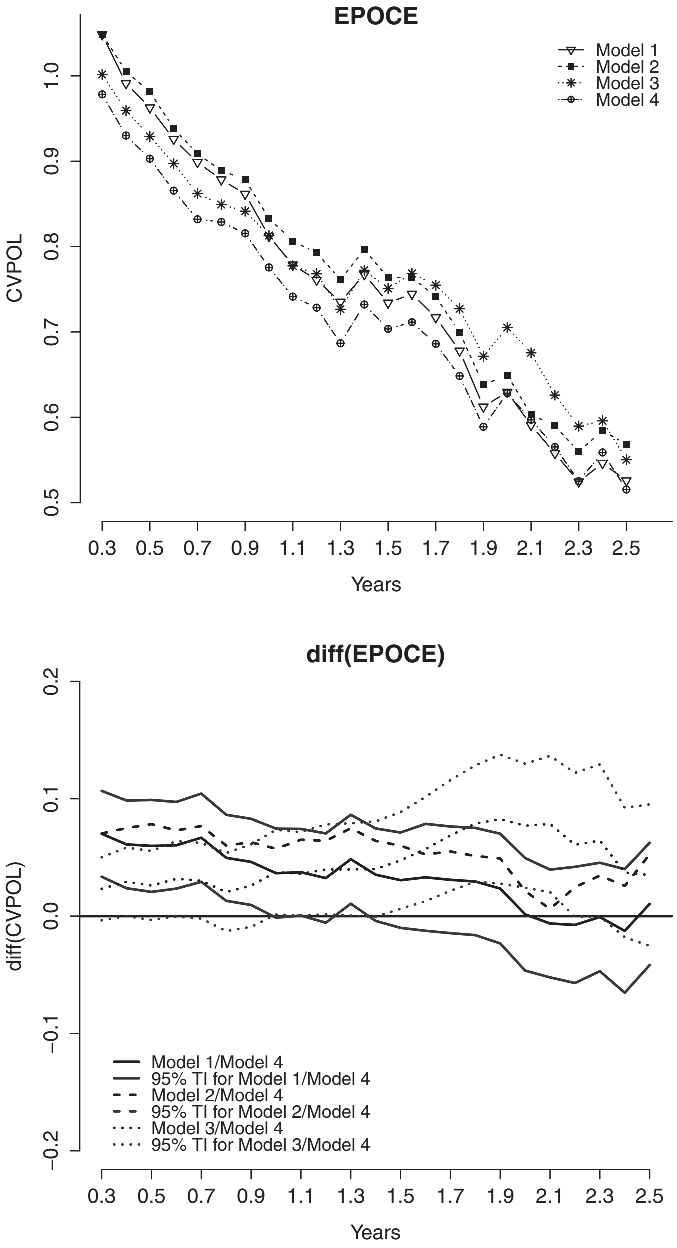

Figure 2. EPOCE computed from the joint models until 2.54 years (top picture) and differences in EPOCE with 95% tracking interval (TI) (bottom picture). Model 1—tumor progressions and death, Model 2—occurrence of new lesions and death, Model 3—SLD measure and death, Model 4—SLD measure, occurrence of new lesions and death.

1996) with a substantially shorter computational time but less stable estimation of standard errors. Another option would be to use a pseudo-adaptive Gaussian–Hermite quadrature which would result in smaller number of quadrature points and thus a shorter computational time (Rizopoulos, 2012).

During the analysis of the FFCD 2000–05 data, we encountered some limitations that are often present in cancer clinical trials. The measurements of the tumor size were given for maximum two dimensions and thus the choice for the summary was limited to uni- or bidimensional. In order to apply a volumetric measure, an extrapolation for the third dimension would be required. Moreover, the appearance of new lesions increases the tumor burden, thus it would be logical to add their measures to the sum of the tumor size but usually they are not registered. In future research on the categorical criteria, we would like to adjust the model for the change of treatment lines. Each line is identified by a respective treatment drug with its particular effects on tumor response and OS. Changes of therapy lines in some cases are related to the high level of toxicity, thus we could distinguish the causes of patient's change of therapy: due to a progression or due to an observed toxicity.

8. Supplementary Materials

All the models were estimated using the R package `frailtypack` (Rondeau et al., 2012). The bivariate model for a biomarker and a terminal event and the trivariate model are available since version 2.8. An extract of the data and R code is presented in Web Appendix E. Web Appendices, Tables, and Figures referenced in Sections 2, 5, 6, and 8 are available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgements

We thank la Fédération Francophone de Cancérologie Digestive for sharing the data of the FFCD 2000–05 trial supported by an unrestricted Grant from Sanofi. We thank IReSP for financing the project AAP 2012 Plan Cancer MULTIPLE and the École des Hautes Études en Santé Publique for funding the research of A. Król. Computer time was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) at the Université de Bordeaux and the Université de Pau et des Pays de l'Adour.

References

An, M.-W., Mandrekar, S. J., Branda, M. E., Hillman, S. L., Adjei, A. A., Pitot, H. C., et al. (2011). Comparison of continuous versus categorical tumor measurement-based metrics to predict overall survival in cancer treatment trials. *Clinical Cancer Research* **17**, 6592–9.

Claret, L., Gupta, M., Han, K., Joshi, A., Sarapa, N., He, J., et al. (2013). Evaluation of tumor-size response metrics to predict overall survival in Western and Chinese patients with first-line metastatic colorectal cancer. *Journal of Clinical Oncology* **31**, 2110–4.

Commenges, D., Liquet, B., and Proust-Lima, C. (2012). Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback-Leibler risks. *Biometrics* **68**, 380–7.

Ducreux, M., Malka, D., Mendiboure, J., Etienne, P.-L., Texereau, P., Auby, D., et al. (2011). Sequential versus combination chemotherapy for the treatment of advanced colorectal cancer (FFCD 2000-05): An open-label, randomised, phase 3 trial. *The Lancet Oncology* **12**, 1032–44.

Efendi, A., Molenberghs, G., Njagi, E. N., and Dendale, P. (2013). A joint model for longitudinal continuous and time-to-event outcomes with direct marginal interpretation. *Biometrical Journal* **55**, 572–88.

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D. J., Ford, R., et al. (2009). New response evaluation criteria in solid tumours : Revised RECIST guideline (version 1.1). *European Journal of Cancer* **45**, 228–247.

Elashoff, R. M., Li, G., and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64**, 762–771.

Genz, A. and Keister, B. (1996). Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight. *Journal of Computational and Applied Mathematics* **71**, 299–309.

Gerds, T. A. and Schumacher, M. (2006). Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal* **48**, 1029–1040.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica* **14**, 863–883.

Jacqmin-Gadda, H., Thiébaut, R., Chêne, G., and Commenges, D. (2000). Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* **1**, 355–68.

Karrison, T. G., Maitland, M. L., Stadler, W. M., and Ratain, M. J. (2007). Design of phase II cancer trials using a continuous endpoint of change in tumor size: Application to a study of sorafenib and erlotinib in non small-cell lung cancer. *Journal of the National Cancer Institute* **99**, 1455–61.

Liu, L. and Huang, X. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society* **58**, 65–81.

Liu, L., Huang, X., and O'Quigley, J. M. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64**, 950–8.

Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747–56.

Lyles, R. H., Lyles, C. M., and Taylor, D. J. (2000). Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative drop-outs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **49**, 485–497.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11**, 431–41.

Mauguen, A., Rachet, B., Mathoulin-Pélissier, S., MacGrogan, G., Laurent, A., and Rondeau, V. (2013). Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in Medicine* **32**, 5366–80.

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

Proust-Lima, C. and Taylor, J. M. G. (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**, 535–49.

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

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

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V., and Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation : application on cancer events. *Biometrics* **8**, 708–721.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2012). frailtypack : An r package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software* **47**.

Royston, P., Altman, D. G., and Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine* **25**, 127–141.

Schluchter, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861–70.

Séne, M., Bellera, C. A., and Proust-Lima, C. (2013). Joint modeling of longitudinal and time-to-event data with application to the prediction of prostate cancer recurrence. *Journal de la Société Française de Statistique* **155**, 134–155.

WHO (1979). WHO Handbook for reporting results of cancer treatment, Geneva (Switzerland): World Health Organization Offset Publication No. 48, 1979.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–9.

Received March 2015. *Revised December* 2015.
Accepted December 2015.