Assessing importance of biomarkers: A Bayesian joint modelling approach of longitudinal and survival data with semi-competing risks

Fan Zhang¹, Ming-Hui Chen², Xiuyu Julie Cong³, and Qingxia Chen⁴

¹Pfizer Inc., Groton, CT, USA
²Department of Statistics, University of Connecticut, Storrs, CT, USA
³Everest Medicines, Shanghai, China
⁴Department of Biostatistics, Vanderbilt University, Nashville, TN, USA

Abstract: Longitudinal biomarkers such as patient-reported outcomes (PROs) and quality of life (QOL) are routinely collected in cancer clinical trials or other studies. Joint modelling of PRO/QOL and survival data can provide a comparative assessment of patient-reported changes in specific symptoms or global measures that correspond to changes in survival. Motivated by a head and neck cancer clinical trial, we develop a class of trajectory-based models for longitudinal and survival data with disease progression. Specifically, we propose a class of mixed effects regression models for longitudinal measures, a cure rate model for the disease progression time (T_P) and a Cox proportional hazards model with time-varying covariates for the overall survival time (T_D) to account for T_P and treatment switching. Under the semi-competing risks framework, the disease progression is the non-terminal event, the occurrence of which is subject to a terminal event of death. The properties of the proposed models are examined in detail. Within the Bayesian paradigm, we derive the decompositions of the deviance information criterion (DIC) and the logarithm of the pseudo-marginal likelihood (LPML) to assess the fit of the longitudinal component of the model and the fit of each survival component, separately. We further develop ΔDIC as well as ΔLPML to determine the importance and contribution of the longitudinal data to the model fit of the T_P and T_D data.

Key words: cure rate model, DIC decomposition, Markov chain Monte Carlo, patient-reported outcome, shared parameter model, time-varying covariates

Submitted September 2019; revised April 2020; accepted May 2020

1 Introduction

A patient-reported outcome (PRO) is any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else (U.S. Food and Drug Administration Guidance for Industry, 2009). PROs can help identify symptoms or problems that may be missed...
during clinical queries, guide policy and health care delivery, and help patients choose between two equally efficacious therapies. Joint modelling of PRO and survival data can provide a comparative assessment of patient-reported changes in specific symptoms or global measures that correspond to changes in survival.

In joint modelling literature, there are usually a longitudinal component, a survival component and a dependence structure connecting those two components. There has been much work on joint modelling of one longitudinal measure and a survival measure, including Schluchter (1992), Hogan and Laird (1997), Law et al. (2002), Brown and Ibrahim (2003), Chen et al. (2004), Ibrahim et al. (2004), Brown et al. (2005), Chi and Ibrahim (2006), Chi and Ibrahim (2007), Ibrahim et al. (2010), Ye et al. (2008), Brown (2009), Rizopoulos (2012), Pawitan and Self (1993), De Gruttola and Tu (1994), Lavallee and Degruutola (1996), De Gruttola and Tu (1994), Xu and Zeger (2001b), Xu and Zeger (2001a), Jacqmin-Gadda et al. (2010) and Proust-Lima et al. (2014). Tudur-Smith et al. (2016) provided a recent review on joint modelling. In the recent development, multiple longitudinal processes and spline-based methods have been used in the joint modelling (Rizopoulos and Ghosh, 2011). Joint modelling has also been applied to more complex data structures such as competing risks data (Williamson et al., 2008; Elashoff et al., 2008; Huang et al., 2011; Li et al., 2012; Proust-Lima et al., 2016) and semi-competing risks data (Li and Su, 2017). In this article, we propose a mixed effects regression model for longitudinal measures, a cure rate model for the disease progression time ($T_P$) and a Cox proportional hazards model with time-varying covariates for the overall survival time ($T_D$) to account for $T_P$ and treatment switching. Note that Chen et al. (2020) provide a detailed elaboration of treatment switching, and they also develop a method for estimating the intended treatment effect accounting for treatment switching. In addition, the trajectory function from the longitudinal model is substituted into the hazard function of each survival submodel, thereby serving as another time-varying covariate in the survival submodel. Thus, our model is the trajectory model (TM). Although Li and Su (2017) address a very similar context as semi-competing risks within the joint modelling framework, they assume a probit model for the discrete time hazard of dropout and another probit model for the discrete time hazard of death. Li and Su (2017) further assume the shared parameter model (SPM), where the longitudinal model and the survival submodels share common random effects, which then induce correlation between the longitudinal and survival components.

Zhang et al. (2017) developed a novel decomposition of DIC and LPML to assess the fit of the longitudinal and survival components of the joint model, separately. Based on this decomposition, they then proposed new Bayesian model assessment criteria, namely $\Delta$DIC and $\Delta$LPML, to determine the importance and contribution of the longitudinal data to the model fit of the survival data. In this article, we extend the decomposition of Zhang et al. (2017) to our joint model with semi-competing risks. We derive the decompositions of DIC and LPML (i.e., $\text{DIC} = \text{DIC}_{\text{Long}} + \text{DIC}_{\text{Pg|Long}} + \text{DIC}_{\text{Surv|Pg,Long}}$ and $\text{LPML} = \text{LPML}_{\text{Long}} + \text{LPML}_{\text{Pg|Long}} + \text{LPML}_{\text{Surv|Pg,Long}}$, respectively) to assess the fit of the longitudinal component of the model and the fit of each survival component, separately. Let $\Delta\text{DIC}_{\text{Surv}}$ be the difference of $\text{DIC}_{\text{Surv|Pg,Long}}$ and $\text{DIC}_{\text{Surv|Pg,Long}}$. 

Statistical Modelling 2021; 21(1–2): 30–55
Fan Zhang et al.

the overall survival model alone without joint modelling. $\Delta$DIC$_{\text{Surv}}$ measures the additional information gained by adding disease progression time and longitudinal components into modelling of overall survival data. Similarly, $\Delta$DIC$_{\text{Pg}}$ measures the additional information gained by adding longitudinal components into modelling of progression data. Similar ideas apply to $\Delta$LPML$_{\text{Surv}}$ and $\Delta$LPML$_{\text{Pg}}$. We also develop a simple but more attractive prior for the covariance matrix based on its Cholesky decomposition borrowing the idea of the Lewandowski, Kurowicka, and Joe (LKJ) prior for the correlation matrix $C$ which has a density function of $f(C|\tau) \propto |C|^{\tau-1}$ (Lewandowski et al., 2009).

The rest of the article is organized as follows. Section 2 gives a detailed description of the motivating longitudinal and survival data from a head and neck clinical cancer trial. In Section 3, we propose a mixed effects regression model for the longitudinal measure, a cure rate model for $T_P$ and a time-varying covariates model for $T_D$. The properties of the proposed models are discussed. The likelihood functions under the proposed models are derived in Section 4. In Section 4, we also present the prior and posterior distributions, develop an efficient Markov chain Monte Carlo sampling algorithm for carrying out the posterior computations, and provide the formulation of the decomposition of the LPML and $\Delta$LPML's for model comparison. A comprehensive analysis of the longitudinal and survival data from a head and neck cancer trial is given in Section 5. We conclude the article with a discussion including some extensions in Section 6.

2 Motivating longitudinal and survival data from a head and neck cancer clinical trial

The motivating study was an open-label, randomized, active-controlled phase III trial comparing the efficacy and safety of afatinib versus methotrexate in patients with recurrent/metastatic squamous carcinoma cancer of head and neck (HNSCC Machiels et al., 2015). Patients were treated on either afatinib or methotrexate, according to randomization, until the occurrence of disease progression or intolerable adverse events. Upon discontinuation of randomized treatment, a patient was free to take other anti-cancer therapies per protocol (except that patients discontinued methotrexate cannot start afatinib). It was reported that more than half of the patients who discontinued study medication started another anti-cancer treatment.

Of the total 483 patients, 332 patients were randomized to the treatment arm (methotrexate) and 161 patients were randomized to the control arm (afatinib). Furthermore, 246 patients switched the treatments, 358 died and 125 were censored. The median follow-up of 6.7 months (interquartile range (IQR): 3.1 – 9.0). Figure 1 shows the counts of disease progression, treatment switch and death in each treatment arm. Three patient-reported measures: global health status (GH), head and neck pain scale (HNPA), and head and neck swallowing condition (HNSW) were collected during this phase III trial. The number of patients with at least one post-baseline values of the longitudinal measures were 481 for GH, 480 for HNPS and 473 for
Figure 1 Graphical representation of the disease progression and survival data from the head and neck cancer clinical trial
Table 1 A summary of the observed counts of longitudinal measures GH, HNPA and HNSW

| Number of observed longitudinal measures | GH       | HNPA     | HNSW     |
|----------------------------------------|----------|----------|----------|
| 0                                      | 6 (1.24%)| 7 (1.45%)| 14 (2.90%)|
| 1                                      | 78 (16.15%)| 77 (15.94%)| 82 (16.98%)|
| 2                                      | 86 (17.81%)| 87 (18.01%)| 94 (19.46%)|
| 3                                      | 111 (22.98%)| 113 (23.40%)| 104 (21.53%)|
| 4                                      | 100 (20.70%)| 96 (19.88%)| 91 (18.84%)|
| 5                                      | 52 (10.77%)| 52 (10.77%)| 50 (10.35%)|
| ≥ 6                                    | 50 (10.35%)| 51 (10.56%)| 48 (9.94%)|

HNSW as shown in Table 1. In this article, we investigate the associations between changes in the longitudinal measures and risks of disease progression ($T_P$) and death ($T_D$), in which $T_D$ was the key secondary endpoint of the trial.

We consider the following baseline covariates: patients’ prior treatment with an epidermal growth factor receptor (EGFR)-targeted antibody, cetuximab (PEGFRA: 1 if yes and 0 if no), age more than 65 years (AGE: 1 if yes and 0 if no), gender (SEX: male/female), prior treatment with plantinum-based chemotherapy and radiotherapy (PCRTDC: 1 if yes and 0 if no), primary tumour site (TMSITED01: 1 if oral cavity, oropharynx, hypopharynx, larynx and 0 if others), metastatic sites of lymph node (METLOC2: 1 if yes and 0 if no), metastatic sites of lung (METLOC3: 1 if yes and 0 if no), liver (METLOC4: 1 if yes and 0 if no), or bone (METLOC5: 1 if yes and 0 if no), Eastern Cooperative Oncology Group (ECOG) performance status at baseline (ECOG: 0/1) and BSDTARG (Baseline target lesion: mm). We exclude two patients from the analysis because of missing BSDTARG measures.

Median survival time in days and its 95% confidence interval are 434 and (386, 470), median switch time, and its 95% confidence interval are 166 and (147, 190), and median progression time and its 95% confidence interval are 84 and (81, 86).

3 The proposed models

Let $Y(t)$ denote the longitudinal measure observed at time $t$. Also, let $T_P$ denote the time from the study entry until disease progression and let $T_D$ denote the time from the study entry until death. In addition, we let $x$ be the $p$-dimensional vector of baseline covariates including the randomized treatment but not the intercept.

3.1 Longitudinal component of the joint model

We first assume a mixed effects regression model for the longitudinal measure $Y(t)$, which is given by

$$Y(t) = (\theta_0^* + \theta)^{g(t)} + \gamma^x + \epsilon(t),$$

(3.1)
where \( \mathbf{g}(t) \) is a \((q + 1)\)-dimensional vector of functions of \( t \) including the intercept, \( \mathbf{\theta}^*_0 \) is a \((q + 1)\)-dimensional vector of subject-dependent random effects, \( \mathbf{\theta} \) is the \((q + 1)\)-dimensional vector of the overall effects, \( \mathbf{y} \) is a \( p \)-dimensional vector of regression coefficients, and \( \epsilon(t) \) is a measurement error. We further assume that 
\[
\mathbf{\theta}^*_0 = (\theta^*_{00}, \ldots, \theta^*_{0q})' \quad \text{and} \quad \epsilon(t) \quad \text{for} \quad t \geq 0 \quad \text{are independent;}
\]
\[
\mathbf{\theta}^*_0 \sim N(0, \Omega), \tag{3.2}
\]
where \( \Omega \) is a \((q + 1) \times (q + 1)\) positive definite covariance matrix; and for \( t \geq 0 \),
\[
\epsilon(t) \sim N(0, \sigma^2), \tag{3.3}
\]
where \( \sigma^2 > 0 \) is a variance parameter.

We note that in (3.1), if \( q = 1 \), \( \mathbf{g}(t) = (1, t)' \) and \((\mathbf{\theta}^*_0 + \mathbf{\theta})' \mathbf{g}(t) \) represents a linear trajectory, and if \( q = 2 \), \( \mathbf{g}(t) = (1, t, t^2)' \) and \((\mathbf{\theta}^*_0 + \mathbf{\theta})' \mathbf{g}(t) \) leads to a quadratic trajectory. A more general form of \( \mathbf{g}(t) \) is given by \( \mathbf{g}(t) = (1, b_1(t), \ldots, b_q(t))' \), where \( b_1(t), \ldots, b_q(t) \) are a set of base functions. We also note that in (3.1), \( \mathbf{\theta}^*_0 \) captures the patient-level dependence of the \( Y(t) \) over time \( t \). However, the covariance matrix \( \Omega \) depends on the data only via random effects \( \mathbf{\theta}^*_0 \), which can lead to slow convergence of Gibbs sampling. To overcome this issue, we consider the reparametrization: \( \mathbf{\theta}^*_0 = \Gamma\mathbf{\theta}^* \), where \( \Gamma \) is the Cholesky decomposition of \( \Omega \) with all diagonal elements being positive such that \( \Omega = \Gamma\Gamma' \). Under this reparametrization, we have
\[
\mathbf{\theta}^* \sim N(0, \mathbf{I}_{q+1}) \tag{3.4}
\]
and a reparametrized version of (3.1) is written as
\[
Y(t) = (\Gamma\mathbf{\theta}^* + \mathbf{\theta})' \mathbf{g}(t) + \mathbf{\gamma}' \mathbf{x} + \epsilon(t). \tag{3.5}
\]

**Remark 3.1:** Write
\[
\Gamma = \begin{bmatrix}
\exp(b_1) & 0 & 0 & \ldots & 0 \\
0 & \exp(b_2) & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & \exp(b_{q+1})
\end{bmatrix}.
\]

Then for any values of \( b_1, \ldots, b_{q+1} \) and \( a_{iq}, i = 2, \ldots, q + 1, j = 2, \ldots, i - 1 \), \( \Gamma \) is a full-rank matrix. Thus, there are no constraints on \( b_1, \ldots, b_{q+1} \) and \( a_{ij}, i = 2, \ldots, q + 1, j = 2, \ldots, i - 1 \). Consequently, a multivariate normal prior can be specified for these parameters. Therefore, in this sense, our parametrization approach may be more attractive than the one discussed in Lewandowski et al. (2009). In addition, we have \( \mathbf{g}(t)'\Gamma\mathbf{\theta}^* = \sum_{i=1}^{q+1} \sum_{i=1}^{q+1} g_i(t)a_{ij}\theta^*_j \), where \( a_{jj} = \exp(b_j) \) for \( j = 1, 2, \ldots, q + 1 \). We may view \( \theta^*_j \)'s as ‘known data points’ and \( a_{ij} \)'s as the ‘regression coefficients’. Under some mild conditions of the longitudinal data, the parameters \( a_{ij} \)'s are stochastically identifiable.
3.2 Survival component of the joint model

Let $\delta$ denote the disease progression status of subjects such that $\delta = 1$ if the subject has disease progression and $\delta = 0$ if the subject never has disease progression. We assume a logistic regression model for $\delta$ as

$$\logit(P(\delta = 1|\phi, x)) = \log \left\{ \frac{P(\delta = 1|\phi, x)}{1 - P(\delta = 1|\phi, x)} \right\} = \phi_0 + \phi_1'x,$$

(3.6)

where $\phi = (\phi_0, \phi_1)'$ is a $(p + 1)$-dimensional vector of regression coefficients. Following Berkson and Gage (1952) and Chen et al. (1999), we assume a cure rate model for $T_p$ with

$$P(T_p = \infty|x, \theta^*, \delta = 0) = 1$$

(3.7)

and a Cox-type model when $\delta = 1$ with

$$P(T_p > t|\lambda_{10}, \beta_1, \alpha_1, x, \theta^*, \theta, \delta = 1) = \exp \left[ -\int_0^t \lambda_{10}(u) \exp\{\beta_1(\Gamma\theta^* + \theta)g(u) + \alpha_1'x\}du \right].$$

(3.8)

where $\lambda_{10}(t)$ is a proper baseline hazard function such that $\int_0^\infty \lambda_{10}(t)dt = \infty$, $\beta_1$ is a regression coefficient corresponding to the longitudinal trajectory function $(\Gamma\theta^* + \theta)g(t)$ and $\alpha_1$ is a $p$-dimensional vector of regression coefficients corresponding to the baseline covariates. From (3.7) and (3.8), we see that $T_p$ can be written as

$$T_p = \infty \times 1\{\delta = 0\} + T_p^* \times 1\{\delta = 1\},$$

(3.9)

where $\infty \times 0 = 0$, $T_p^*$ is a proper random survival time that follows the distribution specified in (3.8) and $1\{A\}$ denotes the indicator function, which takes a value of 1 if $A$ is true and 0 otherwise.

We assume a time-varying covariates Cox model for $T_D$ with hazard function defined as

$$\lambda_2(t|\lambda_{20}, \beta_2, \theta^*, \theta, \alpha_2, \eta, x, z(t)) = \lambda_{20}(t) \exp\{\beta_2(\Gamma\theta^* + \theta)g(t) + \alpha_2'x + \eta'z(t)\},$$

(3.10)

where $\beta_2$ is a regression coefficient corresponding to the longitudinal trajectory function $(\Gamma\theta^* + \theta)g(t)$, $\alpha_2$ is a $p$-dimensional vector of regression coefficients and $z(t)$ is the history of time-dependent covariate up to $t$ with $\eta$ being a vector of the corresponding regression coefficients.

Remark 3.2: We note that although we use $x$ to denote a vector of baseline covariates, different subsets of baseline covariates may be used in (3.1), (3.6), (3.8) and (3.10), respectively. Also $x$ can be different in the longitudinal and survival models. For the notational simplicity, we use the same $x$ for baseline covariates throughout the article.
Remark 3.3: Let \( V_{TP}(t) = 1 \{ T_P \leq t \} \). The variable \( V_{TP}(t) \) essentially tracks the status of disease progression so that \( V_{TP}(t) = 0 \) if the patient has not had disease progression at time \( t \) and \( V_{TP}(t) = 1 \) if the patient has had disease progression before time \( t \). Similarly, we denote \( V_{TSW}(t) = 1 \) if the subject has switched treatment at time \( t \), and \( V_{TSW}(t) = 0 \) if the subject has switched treatment at time \( t \). Suppose that \( R \) denotes the indicator for the randomized treatment such that \( R = 1 \) if the subject is on the treatment arm and \( R = 0 \) if the subject is on the control arm. In (3.10), we take \( z(t) = (V_{TP}(t), V_{TSW}(t), R \times V_{TSW}(t)) \).

Remark 3.4: From the proposed joint model for the longitudinal data and survival data, we see that \( \theta_p \) captures the dependence between the \( Y(t) \) and \( (T_P, T_D) \), while the joint distribution of \( T_P \) and \( T_D \) is specified by the product of the marginal distribution \( [T_P] \) and the conditional distribution \( [T_D|T_P] \). Since the event ‘\( \delta = 0 \)’ is not observed, \( \phi \) is only partially identifiable as shown in Chen et al. (1999). Thus, in our analysis of the head and neck cancer data and the simulation study, we specify a moderate informative prior for \( \phi \).

Remark 3.5: The regression coefficient \( \beta_1 \) in (3.8) (\( \beta_2 \) in (3.10)) captures the association between the longitudinal measure \( Y(t) \) and \( T_P \) (\( T_D \)). If \( \beta_1 = 0 \) and \( \beta_2 = 0 \), \( Y(t) \) is independent of \( T_P \) and \( T_D \). If \( \beta_1 > 0 \), an increase in the longitudinal measure \( Y(t) \) is associated with a shorter disease programme time \( T_P \). Similarly, if \( \beta_2 > 0 \), an increase in \( Y(t) \) is associated with a worse overall survival mortality. In addition, the longitudinal trajectory function \( (\Gamma \theta_p + \theta)^T g(t) \) is directly included in the hazard function for \( T_P \) in (3.8) and the hazard function for \( T_D \) in (3.10) as a time-dependent covariate. Therefore, the proposed joint model is the TM.

3.3 The likelihood function

Suppose that there are \( n \) subjects. For the \( i \)th subject, let \( y_i(t) \) denote the longitudinal measure, which is observed at time \( t \in \{ a_{i1}, a_{i2}, \ldots, a_{im_i} \} \), where \( 0 \leq a_{i1} < a_{i2} < \cdots < a_{im_i} \) and \( m_i \geq 0 \). Note that \( y_i(0) \) corresponds to the baseline value. Also let \( x_i \) denote a \( p \)-dimensional vector of covariates, which may include the randomized treatment. Let \( t_i \) and \( v_i \) denote the overall survival time and the right-censored indicator, respectively, where \( v_i = 1 \) if \( t_i \) is a death time and \( 0 \) if \( t_i \) is right-censored for the \( i \)th subject. Also let \( t_{Pi} \) be the disease progression time and \( \delta_i \) be the status of disease progression. If disease progression has been observed, then \( t_{Pi} < t_i \) and \( \delta_i = 1 \). However, if disease progression has not yet been observed at time \( t_i \), \( t_{Pi} \) and \( \delta_i \) are not observed. Let \( t_{SW_i} \) denote treatment switch time, and if treatment switch has not been observed yet at time \( t_i \), \( t_{SW_i} \) is not observed. Let \( z_i(t) = (V_{TP_i}(t), V_{TSW_i}(t), R \times V_{TSW_i}(t)) \), where \( t \leq t_i \).

Write \( y_i = (y_i(a_{i1}), \ldots, y_i(a_{im_i}))' \) and \( W_i = (W_{i1}, W_{i2}) \), which is an \( m_i \times (q + 1 + p) \) matrix with \( m_i \times (q + 1) \) matrix \( W_{i1} = (g(a_{ij})', j = 1, \ldots, m_i) \) and \( m_i \times p \) matrix \( W_{i2} = (x_j', j = 1, \ldots, m_i) \). Then, the complete data likelihood function for the longitudinal
data $y_i$ is given by
\[
f(y_i | \gamma, \sigma^2, \Gamma, \theta^*_i, \theta, x_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - W_{i1}^* \theta^*_i - W_i(\theta', y'))' (y_i - W_{i1}^* \theta^*_i - W_i(\theta', y')) \right\},
\]
where $\theta^*_i$ is a $(q + 1)$-dimensional vector of random effects, $W_{i1}^* = W_{i1}\Gamma$ and, $W_i(\theta', y') = W_{i1}\theta + W_{i2}y$.

Using (3.2), the density of $\theta^*_i$ is given by
\[
f(\theta^*_i) = \frac{1}{(2\pi)^{(q+1)/2}} \exp \left\{ -\frac{1}{2}(\theta^*_i)'(\theta^*_i) \right\},
\]
for $i = 1, \ldots, n$.

Based on the definition of $z(t)$, we see that $z(t)$ depends on $t_{P_i}$ and $t$ only through $V_{t_{P_i}}(t) = 1\{t_{P_i} \leq t\}$. Write $p(\phi, x_i) = \frac{\exp(\phi + \phi', x_i)}{1 + \exp(\phi + \phi', x_i)}$. We also let $z_i$ denote the group indicator.

Define the survival function for $T_D$ $P(T_D > t_i | \phi_3, x_i, z_i(t), \Gamma, \theta_i^*, \theta) = \exp \left\{ -\int_0^{t_i} \lambda_{20}(u) \exp\{\beta_2((\Gamma \theta^*_i + \theta)'g(u)) + \alpha'_2 x_i + \eta'_i z_i(u)\} du \right\}$, and the density function $f_2(t_i | \phi_3, x_i, z_i(t), \Gamma, \theta_i^*, \theta) = \lambda_{20}(t_i) \exp\{\beta_2((\Gamma \theta^*_i + \theta)'g(t_i)) + \alpha'_2 x_i + \eta'_i z_i(t_i)\} P(T_D > t_i | \phi_3, x_i, z_i(t), \Gamma, \theta_i^*, \theta)$. Define the survival function for $T_P$ $P(T_P > t_i | \phi_{21}, x_i, \Gamma, \theta_i^*, \theta, \delta = 1) = \exp \left\{ -\int_0^{t_i} \lambda_{10}(u) \exp\{\beta_1((\Gamma \theta^*_i + \theta)'g(u)) + \alpha'_1 x_i\} du \right\}$, and the density function $f_1(t_{P_i} | \phi_{21}, x_i, \Gamma, \theta_i^*, \theta, \delta = 1) = \lambda_{10}(t_{P_i}) \exp\{\beta_1((\Gamma \theta^*_i + \theta)'g(t_{P_i})) + \alpha'_1 x_i\} P(T_P > t_{P_i} | \phi_{21}, x_i, \Gamma, \theta_i^*, \theta, \delta = 1)$.

Under the proposed cure rate and time-varying covariates model, the observed data can be divided into four groups of observations. We derive the likelihood function for each of the four groups as follows.

**Group 1.** Subjects are observed to die at time $t_i$ and no disease progressions have been incurred. The survival data consists of $D_{i1} = \{t_i, v_i = 1, x_i, z_i(t)\}$, group indicator $\zeta_i = 1$ and $\delta_i$ is not observed. Since $t_{P_i} = \infty$ when $\delta_i = 0$ and $t_{P_i} \geq t_i$ when $\delta_i = 1$. The likelihood function is given by
\[
L(\phi, \lambda_{10}, \beta_1, \alpha_1, \lambda_{20}, \beta_2, \alpha_2, \eta | D_{i1}, \theta^*_i) = f_2(t_i | \phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta) \times \left( 1 - p(\phi, x_i) + p(\phi, x_i)P(T_P > t_i | \phi_{21}, x_i, \Gamma, \theta_i^*, \theta, \delta = 1) \right).
\]

**Group 2.** Subjects are observed to have disease progression at time $t_{P_i}$ and then die at $t_i$. The survival data consists of $D_{i2} = \{t_i, v_i = 1, t_{P_i}, \delta_i = 1, x_i, z_i(t)\}$, $t_i > t_{P_i}$, and
\( \xi_i = 2 \). The likelihood function is given by

\[
L(\phi, \lambda_{10}, \beta_1, \alpha_1, \lambda_{20}, \beta_2, \alpha_2, \eta|D_{i2}, \theta_i^*) = f_2(t_i|\varphi_3, x_i, z_i(t), \Gamma, \theta_i^* \theta) p(\phi, x_i) \\
	imes f_1(t_{Pi}|\varphi_{21}, x_i, \Gamma, \theta_i^*, \theta, \delta = 1)
\]

(3.14)

**Group 3.** Subjects are observed to have disease progression at time \( t_{Pi} \) and then censor at \( t_i \). The survival data consists of \( D_{i3} = \{ t_i, v_i = 0, t_{Pi}, \delta_i = 1, x_i, z_i(t) \} \), \( t_i > t_{Pi} \), and \( \xi_i = 3 \). The likelihood function takes the form

\[
L(\phi, \lambda_{10}, \beta_1, \alpha_1, \lambda_{20}, \beta_2, \alpha_2, \eta|D_{i3}, \theta_i^*) = P(T_D > t_i|\varphi_3, x_i, z_i(t), \Gamma, \theta_i^* \theta) \\
	imes p(\phi, x_i)f_1(t_{Pi}|\varphi_{21}, x_i, \Gamma, \theta_i^*, \theta, \delta = 1).
\]

(3.15)

**Group 4.** Subjects are censored at \( t_i \) and no disease progression occurs before \( t_i \). The survival data consists of \( D_{i4} = \{ t_i, v_i = 0, x_i, z_i(t) \} \), \( \xi_i = 4 \) and \( \delta_i \) is not observed. The likelihood function is written as

\[
L(\phi, \lambda_{10}, \beta_1, \alpha_1, \lambda_{20}, \beta_2, \alpha_2, \eta|D_{i4}, \theta_i^*) = P(T_D > t_i|\varphi_3, x_i, z_i(t), \Gamma, \theta_i^* \theta) \\
	imes (1 - p(\phi, x_i) + p(\phi, x_i)P(T_P > t_i|\varphi_{21}, x_i, \Gamma, \theta_i^* \theta, \delta = 1)).
\]

(3.16)

**Remark 3.6:** The disease progression is a non-terminal event while the death is a terminal event. The overall survival time \( T_D \) censors the disease progression time \( T_P \), but not vice versa. Thus, the disease progress and the death constitute the semi-competing risks. Under the proposed model, these semi-competing risks are factored into the constructions of the likelihood functions in Groups 1 and 4.

### 4 Bayesian inference

#### 4.1 Prior and posterior distributions

In (3.8) and (3.10), we assume a piecewise constant hazard function for \( \lambda_{j0}(t) \) defined as

\[
\lambda_{j0}(t) = \lambda_{jk}, \quad t \in (s_{j,k-1}, s_{jk}] \quad \text{for } k = 1, \ldots, K_j,
\]

(4.1)

where \( 0 = s_{j0} < s_{j1} < s_{j2} < \ldots < s_{j,k-1} < s_{jK_j} = \infty \) is a finite partition of the time axis, where \( K_j \geq 1 \) for \( j = 1, 2 \). Write \( \lambda_j = (\lambda_{j1}, \ldots, \lambda_{jK_j})' \) for \( j = 1, 2 \). We assume independent normal priors for \( \gamma, \theta, \phi, \beta_1, \alpha_1, \beta_2, \alpha_2 \) and \( \eta \) with \( \gamma \sim N(\mu_{01}, V_{01}), \theta \sim N(\mu_{02}, V_{02}), \phi \sim N(\mu_{03}, V_{03}), (\beta_1, \alpha_1)' \sim N(\mu_{04}, V_{04}), \) and \((\beta_2, \alpha_2, \eta)' \sim N(\mu_{05}, V_{05})\), where \( \mu_{0j} \) and \( V_{0j} \) (a positive definite matrix) for \( j = 1, \ldots, 5 \) are pre-specified hyperparameters. An inverse gamma (IG) prior is specified for \( \sigma^2 \), that is, \( \sigma^2 \sim IG(a_0, b_0) \) with mean \( b_0/(a_0 - 1) \), where \( a_0 \geq 0 \) and \( b_0 \geq 0 \) are pre-specified. When
40 Fan Zhang et al.

$q = 1$, we write $\Gamma = \begin{bmatrix} a & 0 \\ c & b \end{bmatrix}$ where $a = \exp(a^*)$ and $b = \exp(b^*)$. Then, we assume normal prior $N(\mu_{06}, V_{06})$ for $(a^*, b^*, c)$. We further assume independent gamma priors for $\lambda_1$ and $\lambda_2$ with $\lambda_{jk} \sim \text{Gamma}(a_j, b_j)$, where $a_j \geq 0$ and $b_j \geq 0$ for $k = 1, \ldots, K_j$ and $j = 1, 2$. In Section 5 (the real data analysis), the pre-specified hyperparameters are given by $\mu_{01} = \mu_{02} = \mu_{03} = \mu_{04} = \mu_{05} = \mu_{06} = 0$, $V_{01} = 1000I_{p_1}, V_{02} = 1000I_{q_1}, V_{03} = 10I_{p_1}, V_{04} = 100000I_{p_4}$ with $p_3 = \dim((\beta_1, \alpha_1)'), \ V_{05} = 10000I_{p_4}$ with $p_4 = \dim((\beta_2, \alpha_2', \eta'))$, $V_{06} = 10I_2$, $a_0 = 0.001, b_0 = 0.001, d_0 = q + 2$, $\Omega_0 = 0.001I_{q_{j+1}}$, and $a_j = 0.001$ and $b_j = 0.001$ for $j = 1, 2$, where $I_{p_{j+1}}, I_{p_{j+1}}, I_{p_3}, I_{p_4}$ and $I_2$ are the identity matrices. These choices of the hyperparameters yield noninformative priors for all the model parameters except for $\phi$.

Let $\varphi_1 = (\gamma, \sigma^2, \theta, \Gamma), \varphi_2 = (\phi, \varphi_{21})$, where $\varphi_{21} = (\lambda_1, \beta_1, \alpha_1)$, and $\varphi_3 = (\lambda_2, \beta_2, \alpha_2, \eta)$ denote the collections of the model parameters for the longitudinal component, the survival component for the disease progression time and the survival component for the overall survival time, respectively, under the joint model. We further let $\varphi = (\varphi_1, \varphi_2, \varphi_3)$ be the collection of all the model parameters. Then, the joint prior of $\varphi$ is given by

$$\pi(\varphi) = \pi(\gamma)\pi(\theta)\pi(\sigma^2)\pi(a^*, b^*, c)\pi(\phi)\pi(\beta_1, \alpha_1)\pi(\beta_2, \alpha_2, \eta)\pi(\lambda_1)\pi(\lambda_2). \quad (4.2)$$

Let $D_{\text{obs}} = \{y_i, D_{i1}\{\zeta_i = 1\}, D_{i2}\{\zeta_i = 2\}, D_{i3}\{\zeta_i = 3\}, D_{i4}\{\zeta_i = 4\}, i = 1, \ldots, n\}$ denote the observed data, where the indicator function $1(\zeta_i = j) = 1$ if $\zeta_i = j$ and 0 if $\zeta_i \neq j$ for $j = 1, \ldots, 4$. The observed likelihood function is

$$L(\varphi|D_{\text{obs}}) = \prod_{i=1}^{n} \int \left[ f(y_i, \gamma, \sigma^2, \theta, \Gamma, \theta_i^*, x_i) \prod_{j=1}^{4} \left[ L(\varphi_1, \varphi_2, \theta_i^*|D_{\theta_i^*}) 1(\zeta_i = j) \right] f(\theta_i^*)d\theta_i^* \right]. \quad (4.3)$$

where $f(y_i, \gamma, \sigma^2, \theta, \Gamma, \theta_i^*), f(\theta_i^*)$, and $L(\varphi_1, \varphi_2, \theta_i^*)$, $j = 1, \ldots, 4$, are given in (3.11), (3.12), (3.13) to (3.15) and (3.16), respectively, and $\pi(\varphi)$ is defined by (4.2). The joint posterior distribution of $\varphi$ given the observed data is written as

$$\pi(\varphi|D_{\text{obs}}) = \frac{L(\varphi|D_{\text{obs}})\pi(\varphi)}{c(D_{\text{obs}})}, \quad (4.4)$$

where

$$c(D_{\text{obs}}) = \int L(\varphi|D_{\text{obs}})\pi(\varphi)d\varphi. \quad (4.5)$$

4.2 Computational development

To ease exposition, we present the computational algorithm only for $\beta_2(\Gamma\theta_i^* + \theta)g(t)$ and $\beta_1(\Gamma\theta_i^* + \theta)g(t)$ in the survival functions. We can derive a similar
function when assuming the term $\beta_1((\Gamma^*_i + \theta)^T g(T_{max,i}))$ and $\beta_2(\Gamma^*_i + \theta)^T g(T_{max,i})$ after certain time point $T_{max,i}$. Models as a similar algorithm can be developed for the SPM. We introduce a latent vector $\delta^* = (\delta^*_i : \xi_i = 1$ or $\xi_i = 4, 1 \leq i \leq n)$. 

\[
P(\delta^*_i = 1) = \frac{P(\phi, x_i)_{i} P(T_P > t_i|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta) [1 - P(\phi, x_i)]^{(1 - \delta^*_i)}}{1 - P(\phi, x_i)_{i} P(T_P > t_i|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta, \delta = 1)} \quad \text{and} \quad P(\delta^*_i = 0) = 1 - P(\delta^*_i = 1).
\]

Let $\theta^*_R = (\theta^*_1, \ldots, \theta^*_n)'$ denote the vector of the random effects. Then, the augmented posterior distribution of $\pi(\phi|D_{obs})$ in (4.4) can be written as

\[
\pi(\phi, \delta^*, \theta^*_R|D_{obs}) \propto \prod_{i=1}^{n} \left\{ f_2(t_i|\phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta) + P(T_P > t_i|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta, \delta = 1) \right\}^{1(\xi_i = 1)}
\]

\[
\times \left\{ f_2(t_i|\phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta) P(\phi, x_i)|f_1(t_P|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta, \delta = 1) \right\}^{1(\xi_i = 2)}
\]

\[
\times \left\{ P(T_D > t_i|\phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta) P(\phi, x_i)|f_1(t_P|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta, \delta = 1) \right\}^{1(\xi_i = 3)}
\]

\[
\times \left\{ P(T_D > t_i|\phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta) [1 - P(\phi, x_i)]^{(1 - \delta^*_i)} \right\}^{1(\xi_i = 4)}
\]

\[
\times f(y_i|\gamma, \sigma^2, \theta, \Gamma, \theta^*_i, x_i)|f(\theta^*_i) \times \pi(\phi), \quad (4.6)
\]

where $f_2(t_i|\phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta), P(T_D > t_i|\phi_3, x_i, z_i(t), \Gamma, \theta^*_i, \theta), f_1(t_P|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta, \delta = 1)$ and $P(T_P > t_i|\phi_{21}, x_i, \Gamma, \theta^*_i, \theta, \delta = 1)$ are defined in Section 3.3. $f(y_i|\gamma, \sigma^2, \theta, \Gamma, \theta^*_i, x_i), f(\theta^*_i), \pi(\phi)$ are given in (3.11), (3.12) and (4.2). We note that

\[
\int \sum_{\delta^*, \theta^*_R} \pi(\phi, \delta^*, \theta^*_R|D_{obs}) d\theta^R = \pi(\phi|D_{obs}).
\]

Although an analytical evaluation of the posterior distribution in (4.4) is not possible, a convenient Gibbs sampling algorithm can be developed to sample from the augmented posterior distribution in (4.6). Specifically, we sample from the following full conditional distributions in turn: (i) $[y|\sigma^2, \Gamma, \theta^*_R, \theta, D_{obs}]$; (ii) $[\sigma^2|\theta^*_R, \theta, \Gamma, D_{obs}]$; (iii) $[(a^*, b^*, c)|\gamma, \sigma^2, \theta, \theta^*_R, D_{obs}]$; (iv) $[\delta^*|\phi_2, \Gamma, \theta^*_i, \theta, D_{obs}]$; (v) $[\phi|\delta^*, D_{obs}]$; (vi) $[\beta_1, \alpha_1|\lambda_1, \Gamma, \theta^*_R, \theta, \delta^*, D_{obs}]$; (vii) $[\lambda_1|\beta_1, \alpha_1, \Gamma, \theta^*_R, \theta, \delta^*, D_{obs}]$; (viii) $[\beta_2, \alpha_2, \eta|\lambda_2, \Gamma, \theta^*_R, \theta, D_{obs}]$; (ix) $[\lambda_2|\alpha_2, \beta_2, \Gamma, \theta^*_R, \theta, D_{obs}]$; (x) $[\theta^*_R|\phi_1, \phi_2, \phi_3, \delta^*, D_{obs}]$; and (xi) $[\theta|\phi_1, \phi_2, \phi_3, \delta^*, \theta^*_R, D_{obs}]$. We briefly discuss the full conditional distributions in Section S.1 in the supplementary materials.
4.3 Bayesian model assessment

Let \( f(y_i|\varphi_1, \theta_i^*, x_i) = \int f(y_i|\varphi_1, \theta_1^*, x_i)f(\theta_1^*) d\theta_1^* \), where marginal distribution of \( y_i \) follows \( y_i|\varphi_1^*, x_i \sim N(W_i \theta', y') \), \( \sigma^2 I_m + W_i \Gamma' W_i \). Let \( f(t_{Pi}, \delta_i|\varphi_1, \theta_1^*, x_i) = \{1 - p(\varphi, x_i) + p(\varphi, x_i)P(T_p > t_{Pi}|\varphi_2, x_i, \Gamma, \theta_1^*, \theta, \delta = 1)\}^{1[\zeta = 1]+1[\zeta = 3]}\{p(\varphi, x_i)f_1(t_{Pi})f(t_{Pi}, \delta_i|\varphi_1, \theta_1^*, x_i) = f(t_{Pi}, \delta_i|\varphi_2, \Gamma, \theta_1^*, x_i)\}

Based on the joint distributions, define the following conditional distributions: the posterior mean of \( \varphi \) and \( \overline{\varphi} \) are the posterior means of \( \varphi \) and \( \overline{\varphi} \), respectively, with respect to the posterior distribution in (4.4).

To assess the fit of the longitudinal component of the model and the fit of each survival component, we decompose DIC into three parts:

\[
\text{DIC} = \text{DIC}_{\text{Long}} + \text{DIC}_{\text{Pg|Long}} + \text{DIC}_{\text{Surv|Pg,Long}},
\]

where \( \text{DIC}_{\text{Long}} = \text{Dev}_{\text{Long}}(\overline{\varphi}) + 2p_D\), \( \text{DIC}_{\text{Pg|Long}} = \text{Dev}_{\text{Pg|Long}}(\overline{\varphi}) + 2p_D\), and \( \text{DIC}_{\text{Surv|Pg,Long}} = \text{Dev}_{\text{Surv|Pg,Long}}(\overline{\varphi}) + 2p_D\).

\[
\text{DIC}_{\text{Long}} = \text{Dev}_{\text{Long}}(\overline{\varphi}) + 2p_D, \quad \text{DIC}_{\text{Pg|Long}} = \text{Dev}_{\text{Pg|Long}}(\overline{\varphi}) + 2p_D, \quad \text{DIC}_{\text{Surv|Pg,Long}} = \text{Dev}_{\text{Surv|Pg,Long}}(\overline{\varphi}) + 2p_D.
\]

\[
\text{Dev}_{\text{Long}}(\overline{\varphi}) = -2 \sum_{i=1}^n \log f(y_i|\overline{\varphi}_1, x_i), \quad p_D = \overline{\text{Dev}(\varphi)} - \overline{\text{Dev}(\overline{\varphi})},
\]

where \( \overline{\varphi} \) and \( \overline{\text{Dev}(\varphi)} \) are the posterior means of \( \varphi \) and \( \text{Dev}(\varphi) \), respectively, with respect to the posterior distribution in (4.4).
Assessing importance of biomarkers: A Bayesian joint modelling

\( (y_i | \tilde{\phi}_i, \theta^*_i, x_i) f(\theta^*_i) d\theta^*_i \). Then, we have \( \text{Dev}_{\text{Long}}(\phi) + \text{Dev}_{\text{Pg|Long}}(\phi) + \text{Dev}_{\text{Surv|Pg, Long}}(\phi) = \text{Dev}(\phi) \) and \( p_D[\text{Long}] + p_D[\text{Pg|Long}] + p_D[\text{Surv|Pg, Long}] = p_D. \)

When we fit the survival data alone, assuming \( \beta_2 = 0 \) and \( \eta = 0 \). Define \( \lambda^*_2(t|\lambda_{20}, \alpha_2, x) = \lambda_{20}(t) \exp[\alpha'_2 x], \) and \( f_0(t_i|\lambda_{20}, \alpha_2, x_i, v_i) = [\lambda^*_2(t_i|\lambda_{20}, \alpha_2, x)]^{v_i} \exp[- \int_0^{t_i} \lambda^*_2(u|\lambda_{20}, \alpha_2, x) du]. \)

\[ \text{DIC}_{\text{Surv0}} = \text{Dev}_{\text{Surv0}}(\lambda^*_2, \alpha_2) + 2p_D[\text{Surv0}], \]

where \( \text{Dev}_{\text{Surv0}}(\lambda^*_2, \alpha_2) = -2 \sum^n_{i=1} \log f_0(t_i|\lambda^*_2, \alpha_2, x_i, v_i) \) and \( p_D[\text{Surv0}] = E \left[ -2 \sum^n_{i=1} \log \left\{ f_0(t_i|\lambda^*_2, \alpha_2, x_i, v_i) | D_{\text{Surv, obs}} \right\} \right] + 2 \sum^n_{i=1} \log f_0(t_i|\lambda^*_2, \alpha_2, x_i, v_i). \) Define model assessment criterion:

\[ \Delta \text{DIC}_{\text{Surv}} = \text{DIC}_{\text{Surv0}} - \text{DIC}_{\text{Surv|Pg, Long}}. \]

\( \Delta \text{DIC}_{\text{Surv}} \) measures the additional information gained by adding time to disease progression and longitudinal components into modelling of overall survival data. Large values of \( \Delta \text{DIC}_{\text{Surv}} \) correspond to more gain in the fit of the overall survival data. We can also fit survival data alone with time-dependent covariates, assuming \( \beta_2 = 0. \)

Similarly, denote

\[ \Delta \text{DIC}_{\text{Pg}} = \text{DIC}_{\text{Pg0}} - \text{DIC}_{\text{Pg|Long}}, \]

where \( \text{DIC}_{\text{Pg0}} \) is obtained by assuming \( \beta_1 = 0 \) (fitting the time to progression model alone). Define \( \lambda^*_1(t|\lambda_{10}, \alpha_1, x) = \lambda_{10}(t) \exp[\alpha'_1 x], \) \( P^*(T_P > t_{P_i}|\lambda_{10}, \alpha_1, x_i, \delta = 1) = \exp[- \int_0^{t_{P_i}} \lambda^*_1(u|\lambda_{10}, \alpha_1, x) du], \) and \( f_0(t_{P_i}|\lambda_{10}, \alpha_1, \phi, x_i) = \{ 1 - p(\phi, x_i) + p(\phi, x_i) P^*(T_P > t_{P_i}|\lambda_{10}, \alpha_1, x_i, \delta = 1) \}^{1(\xi_1 = 1) + 1(\xi_2 = 1)} \cdot \{ p(\phi, x_i, \lambda^*_1(t|\lambda_{10}, \alpha_1, x)) P^*(T_P > t_{P_i}|\lambda_{10}, \alpha_1, x_i, \delta = 1) \}^{1(\xi_2 = 1) + 1(\xi_3 = 1)} \). We have

\[ \text{DIC}_{\text{Pg0}} = \text{Dev}_{\text{Pg0}}(\lambda^*_1, \alpha_1, \phi) + 2p_D[\text{Pg0}], \]

where \( \text{Dev}_{\text{Pg0}}(\lambda^*_1, \alpha_1, \phi) = -2 \sum^n_{i=1} \log f_0(t_{P_i}, \delta_i|\lambda^*_1, \alpha_1, \phi, x_i) \) and

\[ p_D[\text{Pg0}] = E \left[ -2 \sum^n_{i=1} \log f_0(t_{P_i}, \delta_i|\lambda^*_1, \alpha_1, \phi, x_i) | D_{\text{Pg, obs}} \right] + 2 \sum^n_{i=1} \log f_0(t_{P_i}, \delta_i|\lambda^*_1, \alpha_1, \phi, x_i), \]

and \( D_{\text{Pg, obs}} = (t_{P_i}, \delta_i, x_i \ i = 1, \ldots, n). \) \( \Delta \text{DIC}_{\text{Pg}} \) measures the additional information gained by longitudinal components into modelling of progression data. Large values of \( \Delta \text{DIC}_{\text{Pg}} \) correspond to more gain in the fit of progression data.

As pointed out in Zhang et al. (2017), when the penalty for the additional parameters in the overall survival model (3.10) or the time to progression model (3.8) outweighs the gain of the fit in the respective survival component, \( \Delta \text{DIC}_{\text{Surv}} \) or \( \Delta \text{DIC}_{\text{Pg}} \) can be negative. When \( \Delta \text{DIC}_{\text{Surv}} \) (\( \Delta \text{DIC}_{\text{Pg}} \)) is zero or negative, the
longitudinal measure is not associated with the overall survival (the time to disease progression). This phenomenon is also applied to the $\Delta$ measures developed based on the CPO statistics below.

### 4.3.2 CPO decomposition and LPML decomposition

In this subsection, we first establish the conditional predictive ordinate (CPO) identity which will lead to the decomposition of CPO. CPO (Geisser and Eddy, 1979; Chen et al., 2000; Zhang et al., 2017) for the $i$th subject has the following equivalent definition and identities:

\[
\text{CPO}_i = \int f(y_i, t_{pi}, \delta_i, t_i | \phi, x_i, z_i(t), \nu_i) \pi(\phi | D^{(-i)}_{\text{obs}}) d\phi,
\]

\[
\text{CPO}_i = \frac{c(D_{\text{obs}})}{c(D^{(-i)}_{\text{obs}})} \frac{f(y_i, t_{pi}, \delta_i, t_i | \phi, x_i, z_i(t), \nu_i) \pi(\phi | D^{(-i)}_{\text{obs}})}{\pi(\phi | D_{\text{obs}})},
\]

\[
\text{CPO}_i = \frac{1}{\int \frac{f(y_i, t_{pi}, \delta_i, t_i | \phi, x_i, z_i(t), \nu_i)}{\pi(\phi | D_{\text{obs}})} \pi(\phi | D_{\text{obs}}) d\phi}.
\]

where $c(D_{\text{obs}})$ is defined in (4.5), $c(D^{(-i)}_{\text{obs}}) = \int \prod_{j \neq i} f(y_j, t_{pj}, \delta_j, t_j | \phi, x_j, \nu_j) \pi(\phi) d\phi$, and $f(y_i, t_{pi}, \delta_i, t_i | \phi, x_i, \nu_i)$ is defined in (4.8).

Let $\phi_1^*$, $\phi_2^*$ and $\phi_3^*$ denote the posterior means $\phi_1$, $\phi_2$ and $\phi_3$, and $\phi^* = (\phi_1^*, \phi_2^*, \phi_3^*)$. We have

\[
f(y_i, t_{pi}, \delta_i, t_i | \phi^*, x_i, z_i(t), \nu_i) = f(t_i | \phi^*, x_i, z_i(t), \delta_i, \nu_i, y_i) f(t_{pi}, \delta_i | y_i, \phi_1^*, \phi_2^*, x_i) \times f(y_i | \phi_1^*, x_i).
\]

We also observe that $\pi(\phi^* | D^{(-i)}_{\text{obs}}) = \pi(\phi_1^* | D^{(-i)}_{\text{obs}})\pi(\phi_2^* | \phi_1^*, D^{(-i)}_{\text{obs}})\pi(\phi_3^* | \phi_2^*, \phi_1^*, D^{(-i)}_{\text{obs}})$, $\pi(\phi^* | D_{\text{obs}}) = \pi(\phi_1^* | D_{\text{obs}})\pi(\phi_2^* | \phi_1^*, D_{\text{obs}})\pi(\phi_3^* | \phi_2^*, \phi_1^*, D_{\text{obs}})$. We can propose the CPO decomposition:

\[
\text{CPO}_i = \text{CPO}_{i,\text{Long}} \ast \text{CPO}_{i,\text{PgLong}} \ast \text{CPO}_{i,\text{Surv}} \ast \text{CPO}_{i,\text{Surv|PgLong}}
\]

where $\text{CPO}_{i,\text{Long}} = \frac{f(y_i | \phi_1^*, x_i) \pi(\phi_1^* | D^{(-i)}_{\text{obs}}; \phi_1^*)}{\pi(\phi_1^* | D_{\text{obs}})}$, $\text{CPO}_{i,\text{PgLong}} = \frac{f(t_{pi}, \delta_i | y_i, \phi_1^*, \phi_2^*, x_i) \pi(\phi_1^* | \phi_1^*, D^{(-i)}_{\text{obs}})}{\pi(\phi_2^* | \phi_1^*, D_{\text{obs}})}$,

$\text{CPO}_{i,\text{Surv|PgLong}} = \frac{f(t_i | \phi_1^*, \phi_2^*, x_i, t_{pi}, \delta_i, \nu_i) \pi(\phi_1^* | \phi_2^*, \phi_1^*, D^{(-i)}_{\text{obs}})}{\pi(\phi_2^* | \phi_1^*, D_{\text{obs}})}$.

Let $D^{(-i)}_{\text{Long,obs}} = \{(y_j, x_j), j = 1, \ldots, i - 1, \ldots, n\}$, $D^{(-i)}_{\text{Pg,obs}} = \{(t_{pj}, \delta_j, x_j), j = 1, \ldots, i - 1, \ldots, n\}$. When we assume that $D_{\text{Long,obs}}$, $D_{\text{Pg,obs}}$ and $D_{\text{Surv,obs}}$ are mutually independent. Assume $\pi(\phi) = \frac{1}{\int \pi(\phi)}$. 

Statistical Modelling 2021; 21(1–2): 30–55
\( \pi(\varphi_1)\pi(\varphi_2)\pi(\varphi_3) \). We have

\[
CPO_{i,\text{Surv}|\text{Pg,Long}} = CPO_{i,\text{Surv}0} = \int f_0(t_i|\lambda_2, \alpha_2, x_i, v_i)\pi(\varphi_3|D^{(-i)}_{\text{Surv,obs}})d\varphi_3,
\]

\[
CPO_{i,\text{Pg|Long}} = CPO_{i,\text{Pg}0} = \int f_0(t_{Pi}, \delta_i|\lambda_1, \alpha_1, \phi, x_i)\pi(\varphi_2|D^{(-i)}_{\text{Pg,obs}})d\varphi_2.
\]

From Section S.2 in the supplementary materials, we have the following identities for \( CPO_i, CPO_{i,\text{Long}}, CPO_{i,\text{Pg|Long}} \) and \( CPO_{i,\text{Surv}|\text{Pg,Long}} \):

\[
CPO_{i,\text{Long}} = CPO_i f(y_i|\varphi_1^*, x_i)
\times \int f^{-1}(y_i, t_{Pi}, \delta_i, t_i|\varphi_1^*, \varphi_2, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_2, \varphi_3|\varphi_1^*, D_{\text{obs}})d\varphi_2d\varphi_3,
\]

\[
CPO_{i,\text{Pg|Long}} = f(t_{Pi}, \delta_i|y_i, \varphi_1^*, \varphi_2^*, x_i)
\times \frac{[\int f^{-1}(y_i, t_{Pi}, \delta_i, t_i|\varphi_1^*, \varphi_2^*, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_3|\varphi_1^*, \varphi_2^*, D_{\text{obs}})d\varphi_3]}{[\int f^{-1}(y_i, t_{Pi}, \delta_i, t_i|\varphi_1^*, \varphi_2, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_2, \varphi_3|\varphi_1^*, D_{\text{obs}})d\varphi_2d\varphi_3]},
\]

\[
CPO_{i,\text{Surv}|\text{Pg,Long}} = \left[\int f^{-1}(t_i|\varphi_1^*, \varphi_2^*, \varphi_3, x_i, t_{Pi}, \delta_i, v_i, y_i)\pi(\varphi_3|\varphi_1^*, \varphi_2^*, D_{\text{obs}})d\varphi_3\right]^{-1}.
\]

The logarithm of the pseudo-marginal likelihood (LPML; Ibrahim et al., 2001) is defined as

\[
\text{LPML} = \sum_{i=1}^{n} \log(CPO_i).
\]

Then LPML can be decomposed as

\[
\text{LPML} = \text{LPML}_{\text{Long}} + \text{LPML}_{\text{Pg|Long}} + \text{LPML}_{\text{Surv}|\text{Pg,Long}},
\]

where \( \text{LPML}_{\text{Long}} = \sum_{i=1}^{n} \log(CPO_{i,\text{Long}}), \text{LPML}_{\text{Pg|Long}} = \sum_{i=1}^{n} \log(CPO_{i,\text{Pg|Long}}) \) and \( \text{LPML}_{\text{Surv}|\text{Pg,Long}} = \sum_{i=1}^{n} \log(CPO_{i,\text{Surv}|\text{Pg,Long}}) \).

Under the same assumptions for the calculation of \( \text{DIC}_{\text{Surv}0} \), define \( \text{LPML}_{\text{Surv}0} = \sum_{i=1}^{n} \log(CPO_{i,\text{Surv}0}) \). We propose the model assessment criterion:

\[
\Delta\text{LPML}_{\text{Surv}} = \text{LPML}_{\text{Surv}|\text{Pg,Long}} - \text{LPML}_{\text{Surv}0}.
\]

\( \Delta\text{LPML}_{\text{Surv}} \) measures the additional information gained by adding time to disease progression and longitudinal components into modelling of overall survival data. Large values of \( \Delta\text{LPML}_{\text{Surv}} \) correspond to more gain in the fit of the overall survival data. Similarly, denote \( \Delta\text{LPML}_{\text{Pg}} = \text{LPML}_{\text{Pg|Long}} - \text{LPML}_{\text{Pg}0} \). \( \text{LPML}_{\text{Pg}0} \) is from assuming \( \beta_1 = 0 \) and fit time to progression model alone. \( \Delta\text{LPML}_{\text{Pg}} \) measures the additional information gained by longitudinal components into modelling of progression data. Large values of \( \Delta\text{LPML}_{\text{Pg}} \) correspond to more gain in the fit of progression data.
Algorithm I. For \( f(y_i, t_{pi}, \delta_i, t_i|\varphi_1, \varphi_2, \varphi_3, x_i, z_i(t), v_i) \) and \( f(t_{pi}, \delta_i, y_i, \varphi_1, \varphi_2, x_i) \) can be approximated by adaptive Gaussian quadrature (AGQ).

Now, let \( \{\varphi_b, b = 1, \ldots, B\} \) denote a Gibbs sample of \( \varphi \). An estimate of \( \text{CPO}^{-1}_i \) is given by

\[
\hat{\text{CPO}}^{-1}_i = \frac{1}{B} \sum_{b=1}^{B} f^{-1}(y_i, t_{pi}, \delta_i, t_i|\varphi_b, x_i, z_i(t), v_i).
\]

Next, fix \( \varphi_1 = \varphi_1^* \), and let \( \{\varphi_{2b}, \varphi_{3b}, b = 1, \ldots, B\} \) denote a Gibbs sample of \( (\varphi_2, \varphi_3) \).

\[
\hat{\text{CPO}}_{i,\text{Long}} = \frac{1}{B} \sum_{b=1}^{B} f^{-1}(y_i, t_{si}, \delta_i, t_i|\varphi_1^*, \varphi_{2b}, \varphi_{3b}, x_i, z_i(t), v_i).
\]

Next, fix \( \varphi_1 = \varphi_1^* \) and \( \varphi_2 = \varphi_2^* \), and let \( \{\varphi_{3b}, b = 1, \ldots, B\} \) denote a Gibbs sample of \( (\varphi_3) \). We can calculate

\[
\hat{\text{CPO}}_{i,\text{PgL}} = \frac{1}{B} \sum_{b=1}^{B} f^{-1}(y_i, t_{si}, \delta_i, t_i|\varphi_1^*, \varphi_2^*, \varphi_{3b}, x_i, z_i(t), v_i).
\]

Finally, \( \hat{\text{CPO}}_{i,\text{SurvPgL}} = \frac{\hat{\text{CPO}}_{i,\text{PgL}}}{\hat{\text{CPO}}_{i,\text{Long}} \hat{\text{CPO}}_{i,\text{PgLLong}}}. \)

Algorithm II. Following Section S.2 of the supplementary materials, let \( w_i(\theta^*_i) \) be a normalized weight function such that \( \int w_i(\theta^*_i)d\theta^*_i = 1 \).

\[
\hat{\text{CPO}}^{-1}_i = \int w_i(\theta^*_i)f^{-1}(y_i, t_{si}, \delta_i, t_i|\varphi, x_i, z_i(t), v_i)\pi(\varphi, \theta^*_R|D_{obs})d\theta^*_Rd\varphi.
\]

The optimal choice for \( w_i(\theta^*_i) \) in minimizing the variance of the Monte Carlo estimator for \( \hat{\text{CPO}}_{i,\text{PgL}} \) is

\[
\int w_i(\theta^*_i)f^{-1}(y_i, t_{si}, \delta_i, t_i, \theta^*_i|\varphi_1^*, \varphi_2^*, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_3, \theta^*_R|\varphi_1^*, \varphi_2^*, D_{obs})d\theta^*_Rd\varphi_3 \quad \text{Similarly,} \quad \int f^{-1}(y_i, t_{pi}, \delta_i, t_i|\varphi_1^*, \varphi_2^*, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_3, \theta^*_R|\varphi_1^*, \varphi_2^*, D_{obs})d\theta^*_Rd\varphi_3 = \int w_i(\theta^*_i)f^{-1}(y_i, t_{pi}, \delta_i, t_i, \theta^*_i|\varphi_1^*, \varphi_2^*, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_3, \theta^*_R|\varphi_1^*, \varphi_2^*, D_{obs})d\theta^*_Rd\varphi_3.
\]

The optimal choice for \( w_i(\theta^*_i) \) in minimizing the variance of the Monte Carlo estimator is

\[
\int w_i(\theta^*_i)f^{-1}(y_i, t_{pi}, \delta_i, t_i, \theta^*_i|\varphi_1^*, \varphi_2^*, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_2, \varphi_3|\theta^*_R, \varphi_1^*, \varphi_2^*, D_{obs})d\theta^*_R d\varphi_2 d\varphi_3 = \int w_i(\theta^*_i)f^{-1}(y_i, t_{pi}, \delta_i, t_i, \theta^*_i|\varphi_1^*, \varphi_2, \varphi_3, x_i, z_i(t), v_i)\pi(\varphi_2, \varphi_3|\theta^*_R, \varphi_1^*, D_{obs})d\theta^*_R d\varphi_2 d\varphi_3.
\]
Next, fix \( \phi_1 = \phi^*_1 \), and let \( \{ \phi_{2b}, \phi_{3b}, \theta^R_b, b = 1, \ldots, B \} \) denote a Gibbs sample of \( (\phi_2, \phi_3, \theta^R) \). Set the weight function to be \( w_{i1}(\theta^*_i) \).

\[
\text{CPO}_{i,\text{Long}} = \text{CPO}_i f(y_i | \phi^*_1, x_i) \times \frac{1}{B} \sum_{b=1}^{B} w_{i1}(\theta^*_i) f^{-1}(y_i, t_{Si}, \delta_i, t_i, \theta^i_{ib} | \phi^*_1, \phi_{2b}, \phi_{3b}, x_i, z_i(t), v_i).
\]

Next, fix \( \phi_1 = \phi^*_1 \) and \( \phi_2 = \phi^*_2 \), and let \( \{ \phi_{3b}, \theta^R_b, b = 1, \ldots, B \} \) denote a Gibbs sample of \( (\phi_3, \theta^R) \). Set the weight function to be \( w_{i2}(\theta^*_i) \). We can calculate \( \sum_{b=1}^{B} w_{i2}(\theta^i_{ib}) f^{-1}(y_i, t_{Si}, \delta_i, t_i, \theta^i_{ib} | \phi^*_1, \phi^*_2, \phi_{3b}, x_i, z_i(t), v_i) \). Then we get

\[
\text{CPO}_{i,\text{Pg|Long}} = \frac{\text{CPO}_{i,\text{Long}}}{{\text{CPO}_{i,\text{Long}}}^{-1} \times \text{CPO}_{i,\text{Pg|Long}}^{-1} \times \sum_{b=1}^{B} w_{i2}(\theta^i_{ib}) f^{-1}(y_i, t_{Si}, \delta_i, t_i, \theta^i_{ib} | \phi^*_1, \phi^*_2, \phi_{3b}, x_i, z_i(t), v_i) .}
\]

Finally, \( \text{CPO}_{i,\text{Surv|Pg|Long}} = \frac{\text{CPO}_{i,\text{Surv}}}{{\text{CPO}_{i,\text{Surv}}}^{-1} \times \text{CPO}_{i,\text{Pg|Long}}^{-1} \times \sum_{b=1}^{B} w_{i2}(\theta^i_{ib}) f^{-1}(y_i, t_{Si}, \delta_i, t_i, \theta^i_{ib} | \phi^*_1, \phi^*_2, \phi_{3b}, x_i, z_i(t), v_i) .}
\]

Due to the computational difficulty of the optimal weight functions, another possible choice is \( w_{i,\text{cond}}(\theta^*_i) = f(\theta^*_i | \phi_1, x_i, y_i) \). The conditional distribution of \( \theta^*_i | \phi_1, x_i, y_i \) is \( \theta^*_i | \phi_1, x_i, y_i \sim N(\mu^*, \Sigma^*) \), where \( \mu^* = \Gamma'W_i'\lambda_1, W_i'\Gamma'W_i \) and \( \Sigma^* = I_p - \Gamma'W_i'\lambda_1, W_i'\Gamma'W_i \).

### 5 Analysis of the head and neck cancer data

We apply our proposed method in Section 3 to analyse the head and neck cancer data introduced in Section 2. Out of the total 483 patients, we exclude two patients from the analysis because of missing BSDTARG0 measures. We are interested in identifying which of the three patient-reported measures: GH, HNPA and HNSW contributes most to the model fit of the survival data. Each of GH, HNPA and HNSW was measured on a continuous scale from 0 to 100. Following Ibrahim et al. (2001), we apply transformation to the original longitudinal measures: \( \sqrt{100 - \text{GH}}, \sqrt{100 - \text{HNPA}} \) and \( \sqrt{100 - \text{HNSW}} \), respectively, to achieve a better normal approximation. For each transformed longitudinal measure, we fit the joint model for the longitudinal measure, disease progression time and the overall survival, with a time-dependent covariate accounting for the model fit of the survival data. We include 11 covariates described in Section 2, namely ECOG, AGE, PEGFRA, FEMALE PCRTDC, TMSITED01, METLOC2, METLOC3, METLOC4, ETLOC5, BSDTARG, and treatment indicator R in the time to progression model (3.8) and the overall survival model (3.10). We include five covariates, namely ECOG, AGE,
PEGFRA, FEMALE, and PCRTDC, and treatment indicator R in the cure rate model (3.6). In the time to progression model the survival model, and the cure rate model, all of the covariates including those binary endpoints, except for the treatment indicator R, are standardized to have mean 0 and standard division 1. We use a linear trajectory in the longitudinal model. No standardization is done for any of these covariates in the longitudinal model. In the overall survival model, we also include time-dependent indicator for treatment switch \( V_{T_{SW}} \), time-dependent indicator for progression status \( V_{T_P} \) and an interaction term \( R \times V_{T_P} \).

We used DIC and LPML under the survival model alone for overall survival and the cure rate model alone for time to progression to determine \( K_j \), the number of pieces for the piecewise constant hazard function for \( j = 1, 2 \). As reported in the footnotes under Tables S.5 and S.6 in the supplementary materials, the ‘best choices’ were \( K_1 = 25 \) for time to progression and \( K_2 = 7 \) for overall survival according to both DIC and LPML. Under these choices of \( K_1 \) and \( K_2 \), we fit the models of the survival data jointly with each of GH, HNPA and HNSW longitudinal data and calculated the decompositions of DIC and LPML. The values of DIC, \( \Delta \text{DIC}_{Pg} \), \( \Delta \text{DIC}_{Surv,1} \), \( \Delta \text{DIC}_{Surv,2} \), \( \Delta \text{LPML}_{Pg} \), \( \Delta \text{LPML}_{Surv,1} \), and \( \Delta \text{LPML}_{Surv,2} \) are shown in Figure 2 for GH and reported in Table 2 for all three longitudinal measures.

Larger values of \( \Delta \text{DIC}_{Surv} \) and \( \Delta \text{LPML}_{Surv} \) correspond to more gain in the fit of the overall survival data by adding time to disease progression and different longitudinal components. All values of \( \Delta \text{DIC}_{Surv,1} \) and \( \Delta \text{LPML}_{Surv,1} \) in Table 2 were positive, implying that all three longitudinal measures together with the time-dependent covariates helped in the fit of the overall survival data. All values of \( \Delta \text{DIC}_{Surv,2} \) and \( \Delta \text{LPML}_{Surv,2} \) were positive, indicating that all three longitudinal measures helped in the fit of the overall survival data in addition to the time-dependent covariates. Comparing the magnitudes of \( \Delta \text{DIC}_{Surv,2} \) and \( \Delta \text{LPML}_{Surv,2} \) in Table 2, in terms of gain in the fit of overall survival, GH is more associated with the overall survival time than HNSW and HNPA, while HNSW has a stronger association with the overall survival time than HNPA. Similarly, larger values of \( \Delta \text{DIC}_{Pg} \) and \( \Delta \text{LPML}_{Pg} \) correspond to more gain in the fit of disease progression data by adding different longitudinal components. The results in Table 2 show that GH is helpful in the fit of the time to progression data, while there were weak associations between HNPA or HNSW and the time to progression data. After comparing the values of overall DIC for the models jointly with different longitudinal measures in Table 2, we see that HNPA < GH < HNSW in terms of the corresponding DIC values, while based on \( \text{DIC}_{Pg|Long} \) and \( \text{DIC}_{Pg|Long} \), GH has the smallest values. The reason for this is that there were different numbers of observations (1 589 for GH, 1 587 for HNPA and 1 525 for HNSW) for the three longitudinal measures. Thus, the values of the overall DIC were not comparable. The similar results of LPML and LPML decompositions were observed.

Table 3 shows the posterior estimates and 95% HPD intervals of the hazard ratios of variables/parameters for the overall survival model (3.10) and the time to progression model (3.8), and the odds ratios of variables for the cure rate logistic regression model (3.6) and the posterior estimates and the 95% HPD intervals of
Assessing importance of biomarkers: A Bayesian joint modelling

Figure 2 Diagram of the decompositions of DIC and LPML for the disease progression and survival data jointly with GH
parameters for the longitudinal model (3.1) under the joint model with GH. Posterior means, posterior standard deviations, and 95% HPD intervals of the parameters under the joint models with GH, HNPA and HNSW are reported in Tables S.2 to S.4 in the supplementary materials. We see from Table S.2 that both $\beta_1$ and $\beta_2$ with 95% HPD intervals (0.052, 0.246) and (0.195, 0.393), respectively, were significant, in the sense that their corresponding 95% HPD intervals do not contain zero. From Table 3, we also see that the 95% HPD intervals of the hazard ratios corresponding to $\beta_1$ and $\beta_2$ are (1.053, 1.279) and (1.215, 1.482), respectively, which do not contain 1. Thus, a covariate is significant if its 95% HPD interval of the hazard ratio does not contain 1. These results imply that an increase in the time-dependent longitudinal trajectory function is associated with a shorter disease progression time as well as a worse overall survival mortality, which is clinically appealing since a large value of the longitudinal trajectory function corresponds to a small GH value. Under the proposed joint model, $\beta_1$ constitutes relationship between the longitudinal measure and the time to progression, while $\beta_2$ induces correlation between the longitudinal measure and the overall survival time. In particular, $\beta_1 = 0$ or $\beta_2 = 0$ implies that the longitudinal measure does not influence the risk of disease progression or overall survival. From Tables S.3 and S.4, we see that only $\beta_2$ was significant for both HNPA with respective 95% HPD intervals ($-0.237$, $-0.087$) and ($-0.193$, $-0.095$). These results were consistent with our decompositions of DIC and LPML as the values of $\triangle$DIC$_{Surv,2}$ and $\triangle$LPML$_{Surv,2}$ were $34.75$ and $17.17$ for GH, $15.93$ and $7.96$ for HNPA, and $29.23$ and $14.88$ for HNSW, while the values of $\triangle$DIC$_{Pg}$ and $\triangle$LPML$_{Pg}$ were $7.48$ and $2.57$ for GH, $1.21$ and $-1.66$ for HNPA, and $-0.56$ and $-0.78$ for HNSW. Similarly, from Tables S.2 to S.4, we see that the 95% HPD for the coefficient of time-dependent covariate $V_{T_P}$ was significant for all of GH, HNPA and HNSW. Interestingly, the longitudinal measures could have a minor impact on the significance of covariates under the model for the overall survival. Specifically, METLOC5 was not significant under the joint model with GH, while METLOC5 was significant under the joint models with HNPA and HNSW. Also, the treatment effect (R) without treatment switch was significant under the joint models with GH and HNSW, while it was not significant under the joint model with HNPA.
### Table 3 Parameter estimates under the joint model with GH

| Parameter | EST Survival (HR) 95%HPD | Parameter | EST Progression (HR) 95%HPD |
|-----------|--------------------------|-----------|-----------------------------|
| ECOG      | 2.221 (1.669, 2.949)     | ECOG      | 1.062 (0.780, 1.442)        |
| AGE       | 0.937 (0.707, 1.207)     | AGE       | 0.833 (0.577, 1.190)        |
| PEGFRAD   | 1.174 (0.916, 1.500)     | PEGFRAD   | 1.548 (1.170, 2.043)        |
| Female    | 0.992 (0.707, 1.375)     | Female    | 0.964 (0.642, 1.446)        |
| PCRTDC    | 1.242 (0.978, 1.566)     | PCRTDC    | 1.028 (0.750, 1.349)        |
| TMSITED01 | 1.507 (1.150, 1.947)     | TMSITED01 | 0.853 (0.624, 1.123)        |
| METLOC2   | 1.072 (0.797, 1.408)     | METLOC2   | 0.942 (0.686, 1.229)        |
| METLOC3   | 0.755 (0.575, 0.949)     | METLOC3   | 1.046 (0.800, 1.389)        |
| METLOC4   | 1.699 (1.137, 2.499)     | METLOC4   | 1.983 (1.225, 3.007)        |
| METLOC5   | 1.397 (0.910, 2.055)     | METLOC5   | 1.127 (0.696, 1.758)        |
| BSDTARG   | 1.009 (1.006, 1.012)     | BSDTARG   | 1.003 (1.000, 1.006)        |
| R         | 0.705 (0.501, 0.945)     | R         | 0.973 (0.723, 1.288)        |
| $V_{TP}$  | 2.490 (1.904, 3.283)     | $\beta_1$ | 1.161 (1.053, 1.279)        |
| $V_{TSW}$ | 0.782 (0.515, 1.129)     | $\beta_2$ | 1.346 (1.215, 1.482)        |
| $R + V_{TSW}$ | 1.410 (0.861, 2.118) | $\ell_1$ | 4.863 (4.367, 5.397)        |
| $\beta_2$ | 1.346 (1.215, 1.482)     | $\ell_2$ | 0.126 (0.080, 0.173)        |

**Logistic (OR)**

| Parameter | EST 95%HPD | Parameter | EST 95%HPD |
|-----------|------------|-----------|------------|
| ECOG      | 0.893 (0.133, 7.473) | ECOG      | 0.187 (-0.144, 0.539) |
| AGE       | 0.569 (0.104, 16.586) | PEGFRAD   | 0.263 (-0.021, 0.578) |
| PEGFRAD   | 0.679 (0.056, 4.709)  | Female    | 0.113 (-0.326, 0.534) |
| Female    | 1.270 (0.114, 95.950) | PCRTDC    | 0.102 (-0.189, 0.415) |
| PCRTDC    | 2.715 (0.447, 24.536) | TMSITED01 | 0.204 (-0.131, 0.520) |
| R         | 1.796 (0.457, 9.381)  | METLOC2   | 0.169 (-0.196, 0.537) |
| METLOC3   | -0.028 (-0.342, 0.267) | METLOC4   | -0.604 (-1.182, -0.093) |
| METLOC4   | 0.339 (-0.197, 0.898) | BSDTARG   | 0.002 (-0.002, 0.006) |
| BSDTARG   | 0.028 (-0.007, 0.059) | $\ell_{11}$ | 2.036 (1.644, 2.463) |
| $\ell_{21}$ | -0.021 (-0.057, 0.059) | $\ell_{22}$ | 0.028 (0.000, 0.057) |

Finally, we note that (a) the treatment effect $R$ was not significant under the cure rate model alone for time to progression (Table S.5); (b) the survival model alone for overall survival without time-dependent covariates (Table S.6); and (c) the treatment effect ($R$) without treatment switch was significant under the survival model alone for overall survival with time-dependent covariates (Table S.7).

In all the Bayesian computations, we used 5 000 MCMC samples, which were taken from every third iteration, after a burn-in of 1 500 iterations for each model to compute all posterior estimates, including posterior means, posterior SDs, 95% HPD intervals, DIC and LPML. The convergence of the MCMC sampling algorithm was checked using several diagnostic procedures discussed in (Chen et al., 2000). The trace...
and autocorrelation plots of MCMC chains for the model parameters are shown in Figures S.1 to S.18. We see from these figures that the autocorrelations disappear even after lag 2 for some parameters and after lag 20 for most of the parameters. These trace and autocorrelation plots show good convergence and mixing of these MCMC chains. The HPD intervals were computed via the Monte Carlo method developed by (Chen and Shao, 1999). Computer code was written for the FORTRAN 95 compiler, and we used IMSL subroutines with double precision accuracy. The FORTRAN code is available from the authors upon request. In addition, an R interface of our Fortran code is developed, which can be downloaded from the journal website (http://www.statmod.org/smij/archive.html).

6 Discussion

In this article, we proposed a class of mixed effects regression models for longitudinal measures, a cure rate model for the disease progression time and a time-varying covariates model for the overall survival. In Section 5, we carried out a Bayesian analysis of the head and neck cancer data using the computational algorithm discussed in Section 4.2. To examine the empirical performance of the posterior estimates under the proposed joint model, we conducted a simulation study. The design and results of the simulation study are presented in Section S.3 of the supplementary materials. As shown in Table S.1, the posterior estimates are close to the true values for almost all of the parameters except for $\phi$, and the CP’s are close to 95% for most of the parameters. The simulation results empirically confirm convergence of the proposed Markov chain Monte Carlo sampling algorithm and identifiability of the model parameters except for $\phi$ under the proposed joint model.

Within the semi-competing risk setting, we developed the decompositions of DIC and LPML to determine which PRO is most associated with the survival outcome. As discussed in Section 4.3, the DIC and DIC decompositions require a numerical approximation of an integral over random effects $\theta^*$, while the LPML and LPML decompositions can be computed using the Monte Carlo methods. Thus, the LPML and LPML decompositions are more advantageous when the dimension of random effects becomes large. The proposed joint models and the corresponding decompositions of Bayesian model selection criteria can be extended to discrete longitudinal biomarkers such as Eastern Cooperative Oncology Group (ECOG) performance status and other non-normal longitudinal biomarkers. Other extensions include the joint models of the multiple longitudinal measures and the survival data with semi-competing risks. These extensions are currently under investigation.

Acknowledgements

This work was done when Dr Zhang was at University of Connecticut and when Dr Cong was at Boehringer Ingelheim (China) Investment Co. Ltd, Shanghai, China. We
would like to thank the guest editors and two anonymous reviewers for their very helpful and constructive comments, which have led to a much improved version of the article.

**Declaration of conflicting interests**

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

**Funding**

Dr M.-H. Chen’s research was partially supported by NIH grants #GM 70335 and #CA 74015.

**Supplementary materials**

The supplementary materials, which include Section S.1 The Full Conditional Posterior Distributions, Section S.2 Results and Formulas, Section S.3 A Simulation Study, Tables S.1 to S.7 and Figures S.1 to S.18, can be found through the link: [http://www.statmod.org/smij/archive.html](http://www.statmod.org/smij/archive.html).

**References**

Berkson J and Gage RP (1952) Survival curve for cancer patients following treatment. *Journal of the American Statistical Association, 47*, 501–515.

Brown ER (2009) Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. *The Annals of Applied Statistics, 3*, 1163–1182.

Brown ER and Ibrahim JG (2003) Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics, 59*, 686–693.

Brown ER, Ibrahim JG and DeGruttola V (2005) A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics, 61*, 64–73.

Chen M-H, Ibrahim JG and Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association, 94*, 909–919.

——— (2004) A new joint model for longitudinal and survival data with a cure fraction. *Journal of Multivariate Analysis, 91*, 18–34.

Chen M-H and Shao Q-M (1999) Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics, 8*, 69–92.

Chen M-H, Shao Q-M and Ibrahim JG (2000) *Monte Carlo Methods in Bayesian Computation*. New York, NY: Springer-Verlag.

Chen Q, Zhang F, Chen M and Cong XJ (2020) Estimation of treatment effects and model diagnostics with two-way
time-varying treatment switching: An application to a head and neck study. *Lifetime Data Analysis*. doi: 10.1007/s10985-020-09495-0.

Chi Y-Y and Ibrahim JG (2006) Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62, 432–445.

——— (2007) Bayesian approaches to joint longitudinal and survival models accommodating both zero and nonzero cure fractions. *Statistica Sinica*, 17, 445–462.

De Gruttola V and Tu XM (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50, 1003–1014.

Elasho RM, 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.

Geisser S and Eddy WF (1979) A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.

Hogan JW and Laird NM (1997) Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16, 239–257.

Huang X, Li G, Elasho RM and Pan J (2011) A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis*, 17, 80–100.

Ibrahim JG, Chen M and Lipsitz SR (2001) Trustmissing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, 88, 551–564.

Ibrahim JG, 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.

Ibrahim JG, Chu H and Chen LM (2010) Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28, 2796–2801.

Jacqmin-Gadda H, Proust-Lima C, Taylor JM and Commenges D (2010) Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics*, 66, 11–19.

Lavalley MP and Degroottola V (1996) Models for empirical Bayes estimators of longitudinal CD4 counts. *Statistics in Medicine*, 15, 2289–2305.

Law NJ, Taylor JM and Sandler H (2002) The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics*, 3, 547–563.

Lewandowski D, Kurowicka D and Joe H (2009) Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100, 1989–2001.

Li N, Elasho RM, Li G and Tseng C-H (2012) Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Statistics in Medicine*, 31, 1707–1721.

Li Q and Su L (2017). Accommodating informative dropout and death: A joint modelling approach for longitudinal and semicompeting risks data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67, 145–163.

Machiels J-PH, Haddad RI, Fayette J, Licitra LF, Tahara M, Vermorken JB, Clement PM, Gailer T, Cupissol D, Grau JJ, Guigay J, Caponigro F, de Castro Jr G, de Souza Viana L, Keilholz U, Del Campo JM, Cong XJ, Ehrnrooth E, Cohen EEW, LUX-H&N 1 investigators (2015) Afatinib versus methotrexate as second-line treatment in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck progressing on or after platinum-based therapy (lux-head & neck 1): An open-label, randomised phase 3 trial. *The Lancet Oncology*, 16, 583–594.

Pawitan Y, and Self S (1993) Modeling disease marker processes in AIDS. *Journal of the American Statistical Association*, 88, 719–26.
Proust-Lima C, Sene M, Taylor JM and Jacqmin-Gadda H (2014) Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research, 23*, 74–90.

Proust-Lima C, Dartigues J-F and Jacqmin-Gadda H (2016) Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: A latent process and latent class approach. *Statistics in Medicine, 35*, 382–398.

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

Rizopoulos D and Ghosh P (2011) A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine, 30*, 1366–1380.

Schluchter MD (1992) Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine, 11*, 1861–1870.

Spiegelhalter DJ, Best NG, Carlin BP and Van Der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*, 583–639.

Tudur-Smith C, Sudell M and Kolamunnage-Dona R (2016) Joint models for longitudinal and time-to-event data: A review of reporting quality with a view to meta-analysis. *BMC Medical Research Methodology, 16*, 168.

U.S. Department of Health and Human Services (2009) Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

Williamson P, Kolamunnage-Dona R, Philipson P and Marson A (2008) Joint modelling of longitudinal and competing risks data. *Statistics in Medicine, 27*, 6426–6438.

Xu J and Zeger SL (2001a) The evaluation of multiple surrogate endpoints. *Biometrics, 57*, 81–87.

——— (2001b) Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 50*, 375–387.

Ye W, Lin X and Taylor JM (2008) Semiparametric modeling of longitudinal measurements and time-to-event data: A two-stage regression calibration approach. *Biometrics, 64*, 1238–1246.

Zhang D, Chen M-H, Ibrahim JG, Boye ME and Shen W (2017) Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics, 26*, 121–133.