

Bayesian Joint Models for Longitudinal and Multi-state Survival Data

Josua Mwanyekange¹, Samuel Musili Mwalili^{1,2} & Oscar Ngesa^{1,2,3}

¹ Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI), Department of Mathematics (Statistics), Kenya

² Jomo-Kenyatta University of Agriculture and Technology (JKUAT), Department of Statistics and Actuarial Science, Kenya

³ Taita Taveta University, Department of Mathematics and Informatics, Kenya

Correspondence: Josua Mwanyekange, Pan African University, Institute for Basic Sciences, Technology and Innovation (PAUSTI), Department of Mathematics (Statistics), Kenya. E-mail: josua@aims-cameroon.org

Received: December 2, 2018 Accepted: January 7, 2019 Online Published: January 15, 2019

doi:10.5539/ijsp.v8n2p34 URL: <https://doi.org/10.5539/ijsp.v8n2p34>

Abstract

Joint models for longitudinal and time to event data are frequently used in many observational studies such as clinical trials with the aim of investigating how biomarkers which are recorded repeatedly in time are associated with time to an event of interest. In most cases, these joint models only consider a univariate time to event process. However, many clinical trials of patients with cancer, involve multiple recurrences of a single event together with a single terminal event experienced by patients over time. Therefore, this article proposes joint modelling approaches for longitudinal and multi-state data. The approach considers two sub-models that are linked by a common latent random variable. The first sub-model is linear mixed effect model that defines the longitudinal process and the second sub-model is a proportional intensity function for the multi-state process. Furthermore, on the proportional intensity model, two different formulations are used to define dependence structure between longitudinal and multi-state processes. In this article, a semi-Markov process that consider the time spent in the current state is defined for the transitions between states. Moreover, the time spent in each transient state is assumed to have Gompertz distribution. A Bayesian method using Markov Chain Monte Carlo (MCMC) is developed for parameter estimation and inferences. The deviance information criterion (DIC) is also derived for Bayesian model selection and comparison. Finally, our proposed joint modeling approach is evaluated through a simulation study and is applied to real datasets (colorectal and colorectal.Longi) which present a random selection of 150 patients from a multi-center randomized phase III clinical trial FFCD 2000-05 of patients diagnosed with metastatic colorectal cancer.

Keywords: joint modelling, longitudinal data, multi-state data, semi-Markov process, MCMC, DIC, sojourn times

1. Introduction

In many biomedical studies such as HIV/AIDS studies, subjects' biomarkers are collected repeatedly over time together with the time to occurrence of a clinical event. The clinical event can be defined as the occurrence of the same type of event more than once or different types among treatment centres. For instance, in prostate cancer, a subject can be followed after cancer treatment and at each clinical visit, biomarker (prostate specific antigen) called longitudinal measurements are recorded together with the time to re-occurrence of cancer (Yu et al., 2008). Another example include the study of time to re-hospitalization by Dendale et al. (2012). In their study, they stated that in order to improve the clinical outcome (re-hospitalization event times) for chronic heart failure patients, patients' blood pressure, body weight and heart rate were measured repeatedly after their first discharge from the hospital along with the time to re-hospitalization. This process (re-hospitalization) is considered as multi-state survival data since a patient could be re-hospitalized more than once over time after the initial discharge.

In the past decade, there have been tremendous works on simultaneously analysis of longitudinal and time to event data. The ultimate idea of these studies, is to account for dependence between these two processes and provide valid and efficient inferences. Traditionally, these two processes are analysed separately, which in most cases leads to inefficient inferences. Classical examples are given by Rizopoulos. (2011) and Guo, & Carlin. (2004) on the HIV/AIDS progression in which CD4 counts were collected repeatedly over time together with the time to AIDS or death. Faucett and Thomas, (1996), Wuffsohn & Tsiatis. (1997), also gave comprehensive reviews of joint models for longitudinal and time to event data. In all these studies, the same principle was used, that is two sub-models were defined, one for longitudinal process and one for time to event process linked by a common latent structure.

Several extensions of joint models have been proposed in recent studies (Ibrahim et al., 2010, McCrink et al., 2013, Lawrence Gould et al., 2015; Król et al., 2016). Another attractive modelling approach in the joint modelling framework

is the shared frailty models. These models attempts to account for unobserved heterogeneity which may occur due to the fact that some subjects are more prone to the event than others subjects (Hougaard, 2012). The frailty term is imposed on the hazard function to account for unmeasured covariates. In early development of frailty models, Vaupel et al. (1979) explained that different subjects are considered to be at risk of a clinical event even though some measurements attributes to the risk make no difference in the model. This makes the joint models more efficient and less biased. However, the frailty models have received some criticisms in current development of joint models. One of such criticisms is in a case of multi-state data (recurrent events and terminal event), frailty models assumes that the joint distributions of both events are completely independent.

Under the joint modelling framework, several authors have developed joint models for repeated measurements and multivariate time to events. One of such authors is Chi & Ibrahim (2006), who proposed joint models for multivariate longitudinal and multivariate survival data. A joint model of longitudinal and competing risks survival data was proposed by Huang et al. (2010). However, there have been insufficient studies on the joint models for longitudinal outcomes with multi-state data. The most recently study on such joint models was proposed by Ferrer et al. (2016). Their approach introduces a shared random effects variable for longitudinal measurements and times of transitions between states. To the best of our knowledge, there exist no joint model for longitudinal and multi-state data (recurrent events with death as terminal event) under the framework of Bayesian inference. Therefore, we address this problem by jointly modelling longitudinal and multi-state data under the framework of continuous time Markov process with transient states representing recurrent events and a single absorbing state for terminal event. As pointed out by de Castro et al. (2015), many issues remain unaddressed in the existing literature. One of the issues they pointed out is the dependence structure of the gap times between two states which they say is not always explicitly accounted for. Hence, this paper seek to develop a joint multi-state semi-Markov model that accommodates the time spent in each transient state. Our model comprises two sub-models linked by shared random effects. The first sub-model is linear mixed effect model for modelling evolution of longitudinal measurements where the interrelation between measurements and subject specific effects are accounted by random intercept and slope. The second sub-model is for transition intensities between states which includes covariates history of the process.

Computation for this proposed modelling approach can be awkward. Therefore, in this work, we adapt Bayesian Monte Carlo Markov chain (MCMC) technique that requires both likelihood function which defines the random process that generates the data and the prior probability distributions for the parameters. In particular, we carry out the computational analysis using an existing software package OpenBugs. The rest of this paper is organized as follows: In section 2, we presented the formulation of the joint multi-state semi-Markov model. In section 3, we presented estimation and inference; this section includes joint likelihood, Bayesian parameter estimation procedure which consists of prior and joint posteriors specification. Section 4 presents Monte Carlo simulation study to assess the performance of the proposed joint modeling approach. Applications of the proposed joint modeling approach to real data are given in section 5. Then, the last section presents the discussion.

2. Joint Multi-state Semi-Markov Model Framework

The joint multi-state semi-Markov model comprised two sub-models: 1) a longitudinal sub-model(linear mixed model) and multi-state semi-Markov sub-model. These sub-models are linked through any function of longitudinal sub-model parameters or shared random effects.

2.1 Longitudinal Sub-model

Suppose that we have a sample of n subjects under study, denoted by $i(i = 1, 2, \dots, n)$. Let $y_i(t_{ij})$ be $n_i \times 1$ column vector of random variables representing the observed longitudinal outcomes for subject i measured at time points $t_{ij}(t_{i1}, t_{i2}, \dots, t_{in_i})$, where $j = (1, 2, \dots, n_i)$. Here n_i represents the numbers of repeated measurements for subject i which varies among subjects. In practice, we may have missing observations for some subjects. This happens due to the fact that some subjects may decide to drop out of the study for reasons not related to the occurrence of event of interest. Therefore, in this study, we assume that these missing values in longitudinal measurements trajectory are missing independently of the unobserved measurements.

To analyse the longitudinal process, we define the distribution of y_{ij} by a Linear Mixed Effects model,

$$y_{ij} = \mu_i^*(t_{ij}) + \epsilon_{ij}, \quad (1)$$

where

$$\mu_i^*(t_{ij}) = x_i'(t_{ij})\beta + \eta_i(t_{ij})$$

is the true value of y_{ij} and the outcome variable x_i is the $(n_i \times p)$ design matrix of fixed effects which includes possible time dependent covariates; β is a $p \times 1$ corresponding column vector of the fixed effect coefficients; $\eta_i(t_{ij}) = z_i'(t_{ij})\mathbf{w}_i$, where z_i

denotes $(q \times 1)$ design matrix for the random effects; $w_i \sim \mathcal{MVN}(0, A)$ and ϵ_i is a $(n_i \times 1)$ column vector of the residuals which represents the part of y_{ij} which is not accounted by the model: $x'(t_{ij})\beta + \eta_i(t_{ij})$ such that $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2 I_{n_i})$. It should be noted here that A represents the variance co-variance matrix in which correlations among the repeated measurements and within subject correlation measurement values are represented. On other hand residuals among subjects and random effects w_i are independent of each other. The variance covariance matrix is defined as

$$A = \begin{pmatrix} \sigma_{w_{i0}}^2 & \rho\sigma_{w_{i0}}\sigma_{w_{i1}} \\ \rho\sigma_{w_{i0}}\sigma_{w_{i1}} & \sigma_{w_{i1}}^2 \end{pmatrix},$$

where ρ denotes correlation between random intercept and slope.

2.2 Multi-state Semi-Markov Sub-model

A multi-state semi-Markov model is a stochastic process in which individuals occupy one of the set of discrete state at any time and their transition probabilities to the next possible states only depends on history process through the time spent in the current transient state (Foucher et al., 2007).

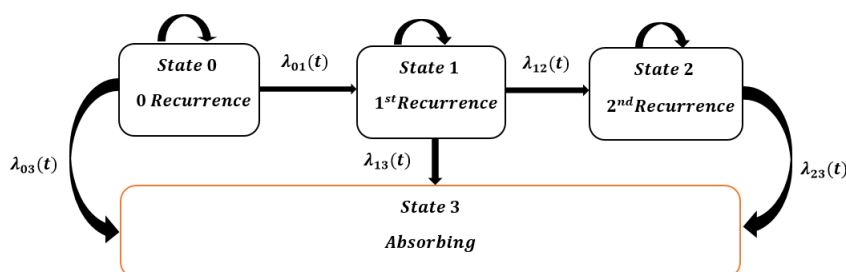


Figure 1. Multi-state model with three transient states (recurrent events) and one absorbing state (terminal event)

Let $X_i = \{X_i(t), t \geq 0\}$ be the right continuous random process for subject i with finite space $E = \{0, 1, 2, \dots, M - 1, M\}$, where the integers $0, 1, 2, \dots, M - 1$ represents the cumulative number of recurrences that subject experienced over the study period. In addition, M represent the absorbing state(e.g., death). In general, $X_i(t)$ is defined as the state in which the process is at time t . This process includes the value of possible time dependent covariates up to time t .

We assume that all subjects enter the study at the same time through initial state $X_i(t_{i,0}) = X_i(0)$ with zero recurrence. Subjects' markers, time to recurrence and death time are recorded periodically. This allows the data to be stratified in such a way that, a subject move directly from state 0 to an absorbing state or alternatively move to state 1 immediately after experiencing the first recurrence. It follows that a subject is expected to stay in state 1 until it experience its second recurrence then move to state 2. If death was observed, then it move to an absorbing state and so on. This process is said to have a semi-Markov property. That is, individual intensity of moving between states $X_i(t_{i,r+1}) = g$ and $X_i(t_{i,r+1}) = h$ depends strongly on history process through the duration spent in the current state $X_i(t_{i,r}) = g$.

Let $d_{i,r}$ denote the holding time for subject i in the state $X_i(t_{i,r})$ before moving to state $X_i(t_{i,r+1})$, represented by $d_{i,r} = t_{i,r+1} - t_{i,r}$, where $t_{i,r}$ is the time of entry into state $X_i(t_{i,r})$ for $r = 0, \dots, m_i$. Also let $D_{i,r}$ denote the corresponding random variable called sojourn time. Then, borrowing some ideas from Foucher et al. (2005), we can express

$$p_{i,gh} = P(X_i(t_{i,r+1}) = h | X_i(t_{i,r}) = g), \quad g, h \in E, \quad r \geq 0 \tag{2}$$

as the probability of transition from state g to state h with constant $\sum_{h \neq g} p_{i,gh} = 1$. Note that if state g is a transient state then, $p_{i,gh} \leq 0$ and $p_{gh} = 0$ for $g = h$. If state g is an absorbing state, the $p_{i,gh} = 0$ for $h \neq g$ and $p_{i,gh} = 1$ for $h = g$. The semi-Markov property assumption works in such a way that, time is rest to zero at each transition. Since our main interest is on the time elapsed between consecutive transitions, we can define the conditional distribution function (CDF) of that time as

$$F_{i,gh}(d_{i,r}) = P(D_{i,r} \leq d_{i,r} | X_i(t_{i,r}) = g, X_i(t_{i,r+1}) = h), \quad d_{i,r} > 0. \tag{3}$$

As usual in survival analysis, survival function s and the transition intensity $\theta_{i,gh}$ associated to probability distribution function of sojourn time can be deduced from the probability density function (pdf) $f_{i,gh}$. If we assume that state g is transient state, that is state 0 to state $M - 1$ in our case, we can define the pdf of the sojourn time in transient states as

$$f_{i,gh}(d_{i,r}) = \lim_{\Delta d_{i,r} \downarrow 0} \frac{1}{\Delta d_{i,r}} P(d_{i,r} \leq D_{i,r} < d_{i,r} + \Delta d_{i,r} | X_i(t_{i,r+1}) = h, X_i(t_{i,r}) = g). \tag{4}$$

Survival function can now be realized from equation (4) as

$$S_{i,gh}(d_{i,r}) = P(D_{i,r} \geq d_{i,r} | X_i(t_{i,r+1}) = h, X_i(t_{i,r}) = g) = \int_{d_{i,r}}^{\infty} f_{i,gh}(s) ds \tag{5}$$

and the conditional transition intensities of the semi-Markov process corresponds to the probabilities of transition from state g to state h between $d_{i,r}$ and $d_{i,r} + \Delta d_{i,r}$ given that the process is in state g for a duration $d_{i,r}$ is defined by

$$\theta_{i,gh}(d_{i,r}) = \lim_{\Delta d_{i,r} \downarrow 0} \frac{1}{\Delta d_{i,r}} P(d_{i,r} \leq D_{i,r} < d_{i,r} + \Delta d_{i,r} | D_{i,r} > d_{i,r}, X_i(t_{i,r+1}) = h, X_i(t_{i,r}) = g) = \frac{f_{i,gh}(d_{i,r})}{S_{i,gh}(d_{i,r})}. \tag{6}$$

By using the total probability theorem, marginal survival function of sojourn times in state g is defined from transition probability p_{gh} and survival function S_g as

$$S_{i,g}(d_{i,r}) = 1 - P(D_{i,r} \leq d_{i,r} | X_i(t_{i,r}) = g) = \sum_{h \neq g} p_{i,gh} (1 - F_{i,gh}(d_{i,r})) = \sum_{h \neq g} p_{i,gh} S_{i,gh}(d_{i,r}). \tag{7}$$

The transition intensity function of the semi-Markov process is said to be related to the transition intensity of the sojourn time, the survival function of the sojourn times and the probabilities of the process. Hence, the corresponding transition intensity function for any subject moving from state g to state h given that the process stays in state g for duration $d_{i,r}$ is defined by

$$\begin{aligned} \lambda_{i,gh}(d_{i,r}) &= \lim_{\Delta d_{i,r} \downarrow 0} \frac{1}{\Delta d_{i,r}} P(d_{i,r} \leq D_{i,r} < d_{i,r} + \Delta d_{i,r}, X_i(t_{i,r+1}) = h | D_{i,r} > d_{i,r}, X_i(t_{i,r}) = g) \\ &= \frac{p_{i,gh} S_{i,gh}(d_{i,r}) \theta_{i,gh}(d_{i,r})}{S_{i,g}(d_{i,r})}, \end{aligned} \tag{8}$$

where $g \neq h, g, h \in S$ and $\lambda_{i,gg}(d_{i,r}) = -\sum_{h \neq g} \lambda_{i,gh}(d_{i,r})$. Hence, equation (7) corresponds to the instantaneous joint probability of moving to state h from state g after staying in state g for duration $d_{i,r}$.

2.3 Distribution of Sojourn Times

We assume that the sojourn times ($d_{i,r}$) follow a Gompertz distribution (GD) (Gompertz,1824) with parameter α_{gh} and γ_{gh} such that, $d_{i,r} \sim Gompertz(\alpha_{gh}, \gamma_{gh})$. This distribution is considered to be one of the classical mathematical models that represent survival function based on laws of mortality. It plays an important role in modelling human mortality. El-Gohary et al. (2013) stated that, Gompertz distribution has been used as a growth model to fit the tumor growth due to its flexibility. It's applications are mainly in the area of actuarial and demography and biology (Ahuja & Nash, 1967). In absence of covariates, the Gompertz distribution has the following probability density function,

$$f_{i,gh}(d_{i,r}) = \alpha_{gh} \exp\{\gamma_{gh} d_{i,r}\} \exp\left\{-\frac{\alpha_{gh}}{\gamma_{gh}} [\exp\{\gamma_{gh} d_{i,r}\} - 1]\right\}. \tag{9}$$

The survival function of the GD is given by

$$\begin{aligned} S_{i,gh}(d_{i,r}) &= \exp\left\{-\int_0^{d_{i,r}} \alpha_{gh} \exp\{\gamma s\} ds\right\} \\ &= \exp\left\{-\frac{\alpha_{gh}}{\gamma_{gh}} [\exp\{\gamma_{gh} d_{i,r}\} - 1]\right\} \end{aligned} \tag{10}$$

and its corresponding transition intensity function is defined by

$$\begin{aligned} \lambda_{i,gh}(d_{i,r}) &= \frac{f_{i,gh}(d_{i,r})}{S_{i,gh}(d_{i,r})} = \frac{\alpha_{gh} \exp\{\gamma_{gh} d_{i,r}\} \exp\left\{-\frac{\alpha_{gh}}{\gamma_{gh}} [\exp\{\gamma_{gh} d_{i,r}\} - 1]\right\}}{\exp\left\{-\frac{\alpha_{gh}}{\gamma_{gh}} [\exp\{\gamma_{gh} d_{i,r} - 1]\right\}} \\ &= \alpha_{gh} \exp\{\gamma_{gh} d_{i,r}\}, \end{aligned} \tag{11}$$

When $\gamma_{gh} > 0$, the hazards increases and when $\gamma_{gh} < 0$, the hazards decreases. While $\gamma_{gh} = 0$, the transition intensity function is equivalent to the one from an exponential distribution with constant hazard rate α_{gh} .

The cumulative function ($F_{i,gh}(d_{i,r})$) of GD is given by

$$F_{i,gh}(d_{i,r}) = P(d_{i,r} \leq d_{i,r}) = 1 - S_{i,gh}(d_{i,r}). \tag{12}$$

From equation 12, we have that the quantiles of the GD can be obtained by using the following expression.

$$F^{-1}(u_{i,gh}) = \frac{1}{\gamma_{gh}} \log \left(-\frac{\gamma_{gh}}{\alpha_{gh}} \log(1 - u_{i,gh}) \gamma_{gh} + 1 \right), \tag{13}$$

where $u_{i,gh} \sim U(0, 1)$, $\gamma_{gh} > 0$ and $\alpha > 0$.

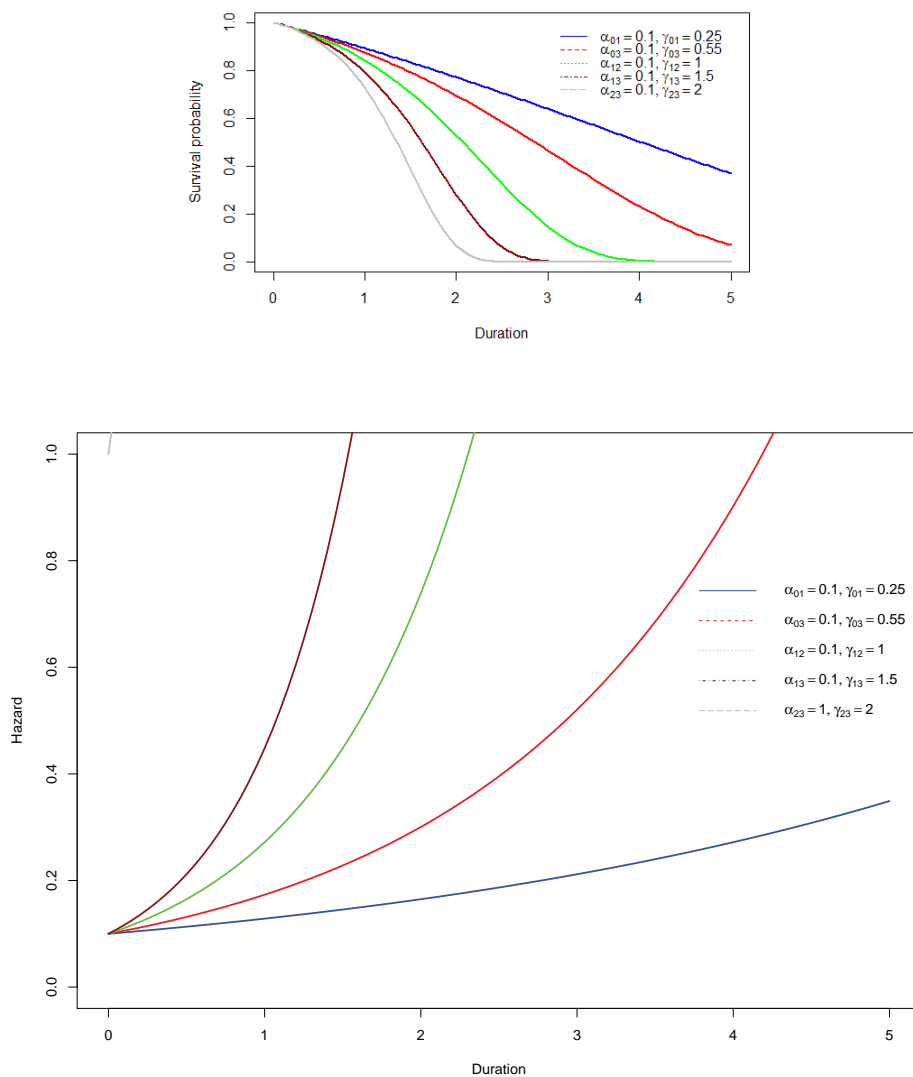


Figure 2. The survival (left) and hazard (right) function of the Gompertz distribution $d_{i,r} \sim GD(\alpha_{gh}; \gamma_{gh})$ with constant α_{gh} and different values of γ_{gh}

2.4 Incorporation of Time-Varying Covariates on Sojourn Times Distribution

In order to take into account the influence of the random effects from longitudinal process and the covariates on the waiting time distribution, we used the assumption of the risks proportionality. These factors are included in the transition intensity

function of the semi-Markov process. Let $v_{i,gh} = (v_{i,gh}^1, \dots, v_{i,gh}^{m_i})'$ denotes the m_i vector of prognostic factor associated with p-vector of regression coefficient $\xi_{gh} = (\xi_{gh}^1, \dots, \xi_{gh}^{m_i})$ for transition $g \rightarrow h$. Let $\chi_{i,gh}(\mathbf{w}_i; d_{i,r})$ be the function that defines the dependent structure between the longitudinal and multi-state semi-Markov processes. Different formulations of random effects function on the transition intensity function can be used in order to propose how the intensity of experiencing a transition depends on the unobserved value of biomarker at $d_{i,r}$. For example,

$$\chi_{i,gh}(\mathbf{w}_i; d_{i,r}) = \begin{cases} \mu_i^*(d_{i,r}), & \text{the current value of biomarker} \\ \mathbf{w}_i, & \text{the shared random effects} \end{cases}$$

Hence, the transition intensity function is given by

$$\lambda_{i,gh}(d_{i,r}|v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r})) = \lambda_{gh,0}(d_{i,r}) \exp\{v'_{i,gh}\xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r})\psi_{gh}\}, \tag{14}$$

where $\lambda_{gh,0}(d_{i,r})$ denotes the baseline intensity function as defined by equation (11) and ψ_{gh} is a vector of parameters that quantifies the strength between longitudinal and survival process for each transition.

Thus, we have

$$\lambda_{i,gh}(d_{i,r}|v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i, d_{i,r})) = \begin{cases} \lambda_{gh,0}(d_{i,r}) \exp\{v'_{i,gh}\xi_{gh} + \mu_i^*(d_{i,r})\psi_{gh}\}, & \text{(Model I)} \\ \lambda_{gh,0}(d_{i,r}) \exp\{v'_{i,gh}\xi_{gh} + \psi_{gh}\mathbf{w}_i\}, & \text{(Model II)} \end{cases}$$

Note that, from the transition intensity function of the waiting times, the regression coefficients can be interpreted in terms of relative risk. Also, the time-dependent covariates can be considered. However, the effects of time depended covariates can be burdensome.

Using the transition intensity function (Model I), we can define the cumulative transition intensity function for subject i with single movement from state g to state h as follows

$$\begin{aligned} \Lambda_{i,gh}(d_{i,r}|v_{i,gh}, kd_{i,r}) &= \int_0^{d_{i,r}} \lambda_{gh}(s) ds, \quad \forall r \geq 0 \\ &= \int_0^{d_{i,r}} \lambda_{gh,0}(s) \exp\{v'_{i,gh}\xi_{gh} + \chi'_{i,gh}(s)\psi_{gh}\} ds \\ &= \int_0^{d_{i,r}} \alpha_{gh} \exp\{\gamma_{gh}s\} \exp\{v'_{i,gh}\xi_{gh} + \chi_{i,gh}(s)\psi_{gh}\} ds \\ &= \alpha_{gh} \exp\{v'_{i,gh}\xi_{gh}\} \int_0^{d_{i,r}} \exp\{\gamma_{gh}s + \psi_{gh}ks\} ds \\ &= \alpha_{gh} \exp\{v'_{i,gh}\xi_{gh}\} \left[\frac{\exp\{(\gamma_{gh} + \psi_{gh}k)d_{i,r}\} - 1}{\gamma_{gh} + \psi_{gh}k} \right], \end{aligned} \tag{15}$$

where the current value $\chi_{i,gh}(\mathbf{w}_i; d_{i,r})$ is assumed to be proportional to $d_{i,r}$ such that $\chi_{i,gh}(\mathbf{w}_i; d_{i,r}) = kd_{i,r}$, with $k > 0$. and the survival function is given by

$$S_{i,gh}(d_{i,r}|v_{i,gh}, kd_{i,r}) = \exp\left\{-\alpha_{gh} \exp\{v'_{i,gh}\xi_{gh}\} \left[\frac{\exp\{(\gamma_{gh} + \psi_{gh}k)d_{i,r}\} - 1}{\gamma_{gh} + \psi_{gh}k} \right]\right\}. \tag{16}$$

The probability density function of the waiting times is given by

$$\begin{aligned} f_{i,gh}(d_{i,r}|v_{i,gh}, kd_{i,r}) &= \alpha_{gh} \exp\{\gamma_{gh}d_{i,r}\} \exp\{v'_{i,gh}\xi_{gh} + \chi'_{i,gh}(d_{i,r})\psi_{gh}\} \\ &\times \exp\left\{-\alpha_{gh} \exp\{v'_{i,gh}\xi_{gh}\} \left[\frac{\exp\{(\gamma_{gh} + \psi_{gh}k)d_{i,r}\} - 1}{\gamma_{gh} + \psi_{gh}k} \right]\right\}. \end{aligned} \tag{17}$$

Now, consider $0 < d_{i,0}^* < d_{i,1}^* < d_{i,2}^* \leq t$ to be the true direct observed durations for subject i through state 0,1, 2 before state 3 between time 0 and t . The cumulative transition intensity for subject i at first recurrence duration $d_{i,0}^*$ is given by

$$\begin{aligned} \Lambda_{01}^i(d_{i,0}^*|v_{i,01}, kd_{i,0}^*) &= \int_0^{d_{i,0}^*} \alpha_{01} \exp\{\gamma_{01}s_{01} + \xi_{01}v_{i,01} + \psi_{01}ks_{01}\} ds_{01} \\ &= \alpha_{01} \exp\{\xi_{01}v_{i,01}\} \int_0^{d_{i,0}^*} \exp\{\gamma_{01}s_{01} + \psi_{01}ks_{01}\} ds_{01} \\ &= \alpha_{01} \exp\{\xi_{01}v_{i,01}\} \left[\frac{\exp\{(\gamma_{01} + \psi_{01}k)s_{01}\}}{\gamma_{01} + \psi_{01}k} \right]_0^{d_{i,0}^*} \\ &= \alpha_{01} \exp\{\xi_{01}v_{i,01}\} \left[\frac{\exp\{(\gamma_{01} + \psi_{01}k)d_{i,0}^*\} - 1}{\gamma_{01} + \psi_{01}k} \right] \end{aligned} \tag{18}$$

and the inverse cumulative function is given by

$$\Lambda_{01}^{-1}(d_{i,0}^*|v_{i,01}, kd_{i,0}^*) = \frac{1}{\gamma_{01} + \psi_{01}k} \log\left(\frac{(\gamma_{01} + \psi_{01}k)d_{i,0}^*}{\alpha_{01} \exp\{\xi_{01}v_{i,01}\}} + 1\right). \tag{19}$$

Consequently, the survival function for subject i in state 0 between time 0 and $d_{i,0}^*$ is given by

$$S_{i,01}(d_{i,0}^*|v_{i,01}, kd_{i,0}^*) = \exp\left\{-\left(\alpha_{01} \exp\{\xi_{01}v_{i,01}\} \left[\frac{\exp\{(\gamma_{01} + \psi_{01}k)d_{i,0}^*\} - 1}{\gamma_{01} + \psi_{01}k} \right]\right)\right\}. \tag{20}$$

Now, suppose that subject i has experienced second recurrence at $d_{i,1}^*$, then its cumulative transition intensity is expressed as

$$\begin{aligned} \Lambda_{i,12}(d_{i,1}^*|v_{i,01}, kd_{i,1}^*) &= \int_{d_{i,0}^*}^{d_{i,1}^*} \alpha_{12} \exp\{\gamma_{12}s_{12} + \xi_{12}v_{i,12} + \psi_{12}ks_{12}\} ds_{12} \\ &= \alpha_{12} \exp\{\xi_{12}v_{i,12}\} \left[\frac{\exp\{(\gamma_{12} + \psi_{12}k)s_{12}\}}{\gamma_{12} + \psi_{12}k} \right]_{d_{i,0}^*}^{d_{i,1}^*} \\ &= \alpha_{12} \exp\{\xi_{12}v_{i,12}\} \left[\frac{\exp\{(\gamma_{12} + \psi_{12}k)d_{i,1}^*\} - \exp\{(\gamma_{12} + \psi_{12}k)d_{i,0}^*\}}{\gamma_{12} + \psi_{12}k} \right] \end{aligned} \tag{21}$$

and its inverse cumulative transition intensity is defined by

$$\Lambda_{i,12}^{-1}(d_{i,1}^*|v_{i,01}, kd_{i,1}^*) = \frac{1}{\gamma_{12} + \psi_{12}k} \log\left(\frac{(\gamma_{12} + \psi_{12}k)d_{i,1}^*}{\alpha_{12} \exp\{\xi_{12}v_{i,12}\}} + A_1\right), \tag{22}$$

where

$$A_1 = \exp\{(\gamma_{12} + \psi_{12}k)d_{i,0}^*\}$$

The survival function for subject i in state 1 between time $d_{i,0}^*$ and $d_{i,1}^*$ is

$$S_{i,12}(d_{i,1}^*|v_{i,01}, kd_{i,1}^*) = \exp\left\{-\left(\alpha_{12} \exp\{\xi_{12}v_{i,12}\} \left[\frac{\exp\{(\gamma_{12} + \psi_{12}k)d_{i,1}^*\} - A_1}{\gamma_{12} + \psi_{12}k} \right]\right)\right\}. \tag{23}$$

If subject i transit to an absorbing state, that is subject i dies at time $d_{i,2}^*$, then its cumulative transition intensities can be expressed as follows:

$$\begin{aligned} \Lambda_{23}^i(d_{i,2}^*|v_{i,01}, kd_{i,2}^*) &= \int_{d_{i,1}^*}^{d_{i,2}^*} \alpha_{23} \exp\{\gamma_{23}s_{23} + \xi_{23}v_{i,23} + \psi_{23}ks_{23}\} ds_{23} \\ &= \alpha_{23} \exp\{\xi_{23}v_{i,23}\} \left[\frac{\exp\{(\gamma_{23} + \psi_{23}k)s_{23}\}}{\gamma_{23} + \psi_{23}k} \right]_{d_{i,1}^*}^{d_{i,2}^*} \\ &= \alpha_{23} \exp\{\xi_{23}v_{i,23}\} \left[\frac{\exp\{(\gamma_{23} + \psi_{23}k)d_{i,2}^*\} - \exp\{(\gamma_{23} + \psi_{23}k)d_{i,1}^*\}}{\gamma_{23} + \psi_{23}k} \right] \end{aligned} \tag{24}$$

and its inverse cumulative transition intensities is

$$\Lambda_{i,23} \left(d_{i,2}^* | v_{i,01}, k d_{i,2}^* \right) = \frac{1}{\gamma_{23} + \psi_{23}k} \log \left(\frac{(\gamma_{23} + \psi_{23}k) d_{i,2}^*}{\alpha_{23} \exp \{ \xi_{23} v_{i,23} \}} + A_2 \right), \tag{25}$$

where

$$A_2 = \exp \{ (\gamma_{23} + \psi_{23}k) d_{i,1}^* \}$$

The survival function for subject i in state 2 between time $d_{i,1}^*$ and $d_{i,2}^*$ is

$$S_{i,22} \left(d_{i,2}^* | v_{i,01}, k d_{i,2}^* \right) = \exp \left\{ - \left(\alpha_{23} \exp \{ \xi_{23} v_{i,23} \} \left[\frac{\exp \{ (\gamma_{23} + \psi_{23}k) d_{i,2}^* \} - A_2}{\gamma_{23} + \psi_{23}k} \right] \right) \right\}. \tag{26}$$

3. Parameter Estimation and Inference

The multi-state process defined in section 2 is continuous and only observed between the left truncated time $t_{i,0}$ and the right censoring C_i , such that the observed process is expressed as $X_i = \{ X_i(t), t_{i,0} \leq t \leq C_i \}$. During this observation time interval, the i^{th} subject may transit $m_i - 1$ times through all transient states at times $t_{i,1} < t_{i,2} < \dots < t_{i,m_i-1}$. At these time points, subject i occupies the state $X_i(t_{i,1}), X_i(t_{i,2}), \dots, X_i(t_{i,m_i-1})$, with $X_i(t_{i,r}) \neq X_i(t_{i,r+1})$ for all $r \geq 0$. At the last observation time t_{i,m_i} , subject i may move again or censored. Let $t_{i,r+1}^*$ denote the true observed transition for subject i and $\delta_{i,r+1} = I_{(t_{i,r+1} \leq t_{i,r+1}^*)}$, represents which event has occur at $t_{i,r+1}$ (transition or censoring), 1 if transition and 0 otherwise. For convenient interpretation, we denote $\delta_i = (\delta_{i,1}, \delta_{i,2}, \dots, \delta_{i,m_i})$ as the vector of observed transition indicators.

3.1 Likelihood Function

The likelihood function associated to semi-Markov process can be defined as

$$\mathcal{L} = \prod_{i=1}^n \prod_{r=0}^{m_i} \left[p_{X_i(t_{i,r}), X_i(t_{i,r+1})} f_{X_i(t_{i,r}), X_i(t_{i,r+1})} (t_{i,r+1} - t_{i,r}) \right]^{\delta_{i,r+1}} \times \left[S_{X_i(t_{i,r})} (t_{i,r+1} - t_{i,r}) \right]^{1 - \delta_{i,r+1}}. \tag{27}$$

Now, linking the longitudinal and multi-state sub-models, we first introduce some notation. Let $\phi = (\phi_Y, \phi_T, \phi_w)$ be the vector of all possible parameters to be estimated, where ϕ_Y denotes all parameters for the longitudinal outcome model $(\phi_Y = (\beta, \sigma_\epsilon^2, \sigma_\epsilon)^\top)$, ϕ_T includes all parameters for transition intensity $(\phi_T = (\alpha_{gh}, \gamma_{gh}, \xi_{gh}, \psi_{gh}))$ and ϕ_w represents all parameters from random effects density function. We denote $f(y_i, d_{i,r}, \mathbf{w}_i | \phi)$ the joint distribution of $(y_i, d_{i,r}, \mathbf{w}_i)$ written as

$$f(y_i, d_{i,r}, \mathbf{w}_i | \phi) = \int f_Y(y_i | \phi_Y; \mathbf{w}_i) f_T(d_{i,r} | \phi_T; \mathbf{w}_i) f_w(\mathbf{w}_i | \phi_w), \tag{28}$$

where $f_Y(y_i | \phi_Y; \mathbf{w}_i)$ is the probability density function of y_i defined by

$$f_Y(y_i | \phi_Y; \mathbf{w}_i) = \frac{1}{(2\pi\sigma_\epsilon^2)^{n_i/2}} \exp \left\{ - \frac{1}{2\sigma_\epsilon^2} \left(y_i - (x_i'(t_{ij})\beta + z_i'(t_{ij})\mathbf{w}_i) \right)^2 \right\}, \tag{29}$$

$f_T(d_{i,r} | \phi_T; \mathbf{w}_i)$ is probability density function of sojourn times $d_{i,r}$ defined as

$$f_T(d_{i,r} | \phi_T; \mathbf{w}_i) = \left[\lambda_{gh}(d_{i,r} | \phi_T, v_{i,gh}; \mathbf{w}_i) \right]^{\delta_{i,r+1}} \exp \left\{ - \int_0^{d_{i,r}} \lambda_{gh}(s | \phi_T, v_{i,gh}; \mathbf{w}_i) ds \right\} \tag{30}$$

and $f_w(\mathbf{w}_i | \phi_w)$ is the probability density function of random effects \mathbf{w}_i given by

$$f_w(\mathbf{w}_i | \phi_w) = (2\pi)^{-q/2} |A|^{-1/2} \exp \left\{ - \frac{1}{2} \mathbf{w}_i^\top A^{-1} \mathbf{w}_i \right\}. \tag{31}$$

Let $\mathcal{D}_{obs} = \{(y_i, d_{i,r}, \delta_{i,r+1}, x_i, v_{i,gh}), i = 1, \dots, n, r = 0, \dots, m_i, g, h \in E\}$ represents the observed data. Under the conditional independence assumption between the random effects and measurement errors of the longitudinal outcomes, the joint likelihood of the longitudinal and the transition intensity sub-models for all observed data is defined by

$$\mathcal{L}(\phi|\mathcal{D}_{obs}) = \prod_{i=1}^n \prod_{j=1}^{n_i} \prod_{r=0}^{m_i} \int_{\mathbf{w}_i} f_Y(y_{ij}|x_i, \mathbf{w}_i; \phi_Y) f_T(d_{i,r}|\mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \phi_T) f_w(\mathbf{w}_i|\phi_w) d\mathbf{w}_i. \tag{32}$$

Taking log of (32), gives

$$\begin{aligned} \log \mathcal{L}(\phi|\mathcal{D}_{obs}) &= \sum_{i=1}^n \int_{\mathbf{w}_i} \left[\left\{ \sum_{j=1}^{n_i} \log f_Y(y_{ij}|\mathbf{w}_i, x_i; \phi_Y) \right\} + \left\{ \sum_{r=0}^{m_i} \log f_T(d_{i,r}|\mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \phi_T) \right\} \right. \\ &\quad \left. + \log f_w(\mathbf{w}_i|\phi_w) \right] d\mathbf{w}_i. \end{aligned} \tag{33}$$

3.2 Bayesian Inference

Bayesian inference is an approach of statistical inference that uses the Bayes’ rule to update the estimates of ϕ after observing some data. This approach is one of the crucial techniques used in modern statistics, especially in mathematical statistics. In Bayesian framework, the unknown quantity, ϕ is treated as a random variable. More precisely, we assume that we have some available initial guess about the distribution of ϕ . This distribution is referred to as prior distribution. Within the Bayesian paradigm, this distribution can be incorporated into analysis. Hence, with the use of Bayes’ theorem, we can combine the joint likelihood function (32) with prior distribution to give posterior distribution for the parameters. The posterior summaries of the interest of this distribution of the parameters are obtained using simulation techniques. In the case of the joint model, the posterior distribution can be very complex leading to difficulty in using analytical methods (Zhang et al., 2017). To overcome such difficulties, we proposed Gibbs sampler technique. This is a well known Markov Chain Monte Carlo (MCMC) method that generates samples from posterior distribution of parameters by repeated sampling from the full-conditional densities of each components. More details are given in Robert and Casella (2004).

3.2.1 Prior Distributions

Let $p(\phi)$ represent prior distribution of a set of unknown parameters ϕ . Choosing prior distributions can be challenging in a case where prior information about the current data is unknown. In this work, we chose prior that will have little to no effect on our final results. This type of prior distribution is called non-informative prior distribution. The ultimate idea of using a non-informative prior distribution is to avoid some external information creeps into analysis. That is, we want observed data to speak for themselves. The prior distributions were specified as follows:

We assume independent normal prior for β , γ_{gh} , ξ_{gh} and ψ_{gh} such that $\beta \sim N(\mu_\beta, \tau_\beta)$, $\gamma_{gh} \sim N(\mu_{\gamma_{gh}}, \tau_{\gamma_{gh}})$, $\xi_{gh} \sim N(\mu_{\xi_{gh}}, \tau_{\xi_{gh}})$ and $\psi_{gh} \sim N(\mu_{\psi_{gh}}, \tau_{\psi_{gh}})$, where μ_{coef} and τ_{coef} are the mean and variance of the hyper-parameters. An independent gamma prior distribution for the Gompertz shape parameter α_{gh} was chosen such that $\alpha_{gh} \sim \Gamma(a_{gh}, b_{gh})$ where $a_{gh} > 0$ and $b_{gh} > 0$ are specified with low values such as 0.001 for each transition. Furthermore, we assume inverse gamma prior for residuals such that $\sigma_\epsilon^2 \sim \mathcal{IG}(a_1, b_1)$ where $a_1 > 0$ and $b_1 > 0$. For the covariance matrix \mathbf{A} , we take an inverse wishart prior, $\mathbf{A} \sim \mathcal{IW}(\mathbf{A}_0, \nu_0)$ where \mathbf{A}_0 denotes the $q \times q$ positive definite matrix with ν_0 degrees of freedom.

Hence, the joint prior distribution for ϕ is thus defined by

$$\begin{aligned} p(\phi) &= p(\beta) \times p(\sigma_\epsilon^2) \times p(\mathbf{A}) \times p(\gamma_{gh}) \times p(\xi_{gh}) \times p(\psi_{gh}) \times p(\alpha_{gh}) \\ &\propto \exp \left\{ -\frac{1}{2} (\beta - \mu_\beta)' \tau_\beta^{-1} (\beta - \mu_\beta) \right\} \times (\sigma_\epsilon^2)^{-a_1-1} \exp \left\{ -\frac{b_1}{\sigma_\epsilon^2} \right\} \\ &\quad \times |\mathbf{A}|^{-(\nu_0+q+1)/2} \exp \left\{ \frac{1}{2} tr(\mathbf{A}_0 \mathbf{A}^{-1}) \right\} \times \left[\prod_{h \neq g} \alpha_{gh}^{a_{gh}-1} \exp(-\alpha_{gh} b_{gh}) \right] \\ &\quad \times \exp \left\{ \frac{1}{2} (\psi_{gh} - \mu_{\psi_{gh}})' \tau_{\psi_{gh}}^{-1} (\psi_{gh} - \mu_{\psi_{gh}}) \right\} \times \exp \left\{ \frac{1}{2} (\gamma_{gh} - \mu_{\gamma_{gh}})' \tau_{\gamma_{gh}}^{-1} (\gamma_{gh} - \mu_{\gamma_{gh}}) \right\} \\ &\quad \times \exp \left\{ \frac{1}{2} (\xi_{gh} - \mu_{\xi_{gh}})' \tau_{\xi_{gh}}^{-1} (\xi_{gh} - \mu_{\xi_{gh}}) \right\}. \end{aligned} \tag{34}$$

3.2.2 Joint Posteriors

The inference about ϕ conditional on \mathcal{D}_{obs} can be done by means of the Bayes' theorem to construct a joint posterior distribution,

$$p(\phi|\mathcal{D}_{obs}) = \frac{\mathcal{L}(\phi|\mathbf{w}_i, \mathcal{D}_{obs}) p(\phi)}{\int \prod_{h \neq g} \prod_{i=1}^n f_{Y,T}(y_{ij}, d_{i,r}|\phi, \delta_{i,r+1}, \mathbf{w}_i) p(\phi) d\phi} \propto \mathcal{L}(\phi|\mathbf{w}_i, \mathcal{D}_{obs}) p(\phi), \tag{35}$$

where $p(\phi|\mathcal{D}_{obs})$ denotes the joint posterior probability distribution, which is proportional to joint likelihood contribution for each subject multiplied by the prior distribution. The samples are drawn from the posterior distribution using Gibbs sampler in order to obtain posterior estimates. Hence, the full joint posterior distribution is given the expression

$$\begin{aligned} p(\phi|\mathcal{D}_{obs}) &\propto \prod_{i=1}^n \int \left[\left\{ \prod_{j=1}^{n_i} f_y(y_i, d_{i,r}|\mathbf{w}_i; \phi) \right\} \times \left\{ \prod_{r=0}^{m_i} f_T(d_{i,r}|\mathbf{w}_i; \phi_T) \right\} \times f_w(\mathbf{w}_i|\phi_w) \right] \times p(\phi) d\mathbf{w}_i \\ &\propto \prod_{i=1}^n \int \left[\left(2\pi\sigma_\epsilon^2 \right)^{-n_i/2} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{n_i} [y_i - (x'_i(t_{ij})\beta + z'_i(t_{ij})\mathbf{w}_i)]^2 \right\} \right] \\ &\quad \times \alpha_{gh} \exp \left\{ \gamma_{gh} d_{i,r} + v'_{i,gh} \xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r}) \psi_{gh} \right\}^{\delta_{i,r+1}} \\ &\quad \times \exp \left\{ -\sum_{h \neq g} \int_0^{d_{i,r}} \alpha_{gh} \exp \left\{ \gamma_{gh} s_{i,r} + v'_{i,gh} \xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; s_{i,r}) \psi_{gh} \right\} ds_{i,r} \right\} \\ &\quad \times |\mathbf{A}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{w}'_i \mathbf{A}^{-1} \mathbf{w}_i) \right\} \times \exp \left\{ -\frac{1}{2} (\beta - \mu_\beta)' \tau_\beta^{-1} (\beta - \mu_\beta) \right\} \\ &\quad \times (\sigma_\epsilon^2)^{-a_1-1} \exp \left\{ -\frac{b_1}{\sigma_\epsilon^2} \right\} \times |\mathbf{A}|^{-(v_0+q+1)/2} \exp \left\{ \frac{1}{2} tr(\mathbf{A}_0 \mathbf{A}^{-1}) \right\} \\ &\quad \times \left[\prod_{h \neq g} \alpha_{gh}^{a_{gh}-1} \exp(-\alpha_{gh} b_{gh}) \times \exp \left\{ \frac{1}{2} (\psi_{gh} - \mu_{\psi_{gh}})' \tau_{\psi_{gh}}^{-1} (\psi_{gh} - \mu_{\psi_{gh}}) \right\} \right] \\ &\quad \times \exp \left\{ \frac{1}{2} (\gamma_{gh} - \mu_{\gamma_{gh}})' \tau_{\gamma_{gh}}^{-1} (\gamma_{gh} - \mu_{\gamma_{gh}}) \right\} \times \exp \left\{ \frac{1}{2} (\xi_{gh} - \mu_{\xi_{gh}})' \tau_{\xi_{gh}}^{-1} (\xi_{gh} - \mu_{\xi_{gh}}) \right\} \right] d\mathbf{w}_i. \tag{36} \end{aligned}$$

As stated earlier, at this point in time, the model has become too complex and analytical evaluation of posterior distribution is not possible. Therefore, Gibbs sampling can be used to obtain posterior estimates. In order to implement the Gibbs sampling technique, the full conditional distribution are needed (see Appendix)

3.2 Model Comparison and Selection

In this subsection, we consider selecting models via the Deviance Information Criterion (DIC) (spiegelhalter et al., 2002).

Consider the parameter vector ϕ and observed data vector \mathcal{D}_{obs} , then, the deviance $Dev(\phi)$ is defined as

$$Dev(\phi) = -2 \log \mathcal{L}(\phi|\mathcal{D}_{obs}) \tag{37}$$

where $\mathcal{L}(\phi|\mathcal{D}_{obs})$ is given in (32).

For our joint multi-state semi-Markov model, we let $Dev(\phi)$ to be

$$Dev(\phi) = -2 \sum_{i=1}^n \log(y_{ij}, d_{i,r}, \delta_{i,r+1}|\phi). \tag{38}$$

The posterior expected deviance is given by

$$\overline{Dev}(\phi) = E_{\phi|\mathcal{D}_{obs}} [Dev(\phi|\mathcal{D}_{obs})]. \tag{39}$$

and the deviance evaluated at the posterior means of ϕ denoted by $\bar{\phi}$ is defined by

$$Dev(\bar{\phi}) = Dev(E_{\phi|\mathcal{D}_{obs}}[\phi]). \tag{40}$$

The difference between equation (40) and (39) is an estimate of the effective number of parameters p_D written as

$$p_D = \overline{Dev}(\boldsymbol{\phi}) - Dev(\bar{\boldsymbol{\phi}}). \tag{41}$$

Then, the deviance information criterion, DIC is given by

$$DIC = 2\overline{Dev}(\boldsymbol{\phi}) - Dev(\bar{\boldsymbol{\phi}}). \tag{42}$$

From equation (40) and (41), DIC can be written as

$$\begin{aligned} DIC &= 2\overline{Dev}(\boldsymbol{\phi}) - (\overline{Dev}(\boldsymbol{\phi}) - p_D) \\ &= \overline{Dev}(\boldsymbol{\phi}) + p_D. \end{aligned} \tag{43}$$

Equivalently,

$$DIC = Dev(\bar{\boldsymbol{\phi}}) + 2p_D. \tag{44}$$

According to Spiegelhalter et al. (2002), the process of computing $Dev(\bar{\boldsymbol{\phi}})$ involves the computing of n integrals as shown in (28). Hence, this pose a major challenge on computing the overall likelihood of observed data in our joint model. One possible approach to approximate equation (44) is to use Adaptive Gaussian Quadrature (AGQ) proposed by Pinheiro and Bates, 1995. In this paper, we adapt Monte Carlo Markov Chain approach.

Let consider the s^{th} MCMC sample, $s = \{1, 2, \dots, S\}$. Also consider the i^{th} subject such that $i = \{1, 2, \dots, n\}$. Now the deviance for the i^{th} subject at the s^{th} MCMC sample is defined as

$$Dev_i^{(s)}(\boldsymbol{\phi}^{(s)}) = -2 \log \mathcal{L}_i(\boldsymbol{\phi}^{(s)} | \mathcal{D}_{obs}), \tag{45}$$

where

$$\mathcal{L}_i(\boldsymbol{\phi}^{(s)} | \mathcal{D}_{obs}) = \prod_{j=1}^{n_i} \prod_{r=0}^{m_i} \int_{\mathbf{w}_i} f_Y(y_{ij} | x_i, \mathbf{w}_i; \boldsymbol{\phi}_Y^{(s)}) f_T(d_{i,r} | \mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \boldsymbol{\phi}_T^{(s)}) f_w(\mathbf{w}_i | \boldsymbol{\phi}_w^{(s)}) d\mathbf{w}_i. \tag{46}$$

Then, the deviance over all individuals is given by

$$\begin{aligned} Dev(\boldsymbol{\phi}^{(s)} | \mathcal{D}_{obs}) &= -2 \sum_{i=1}^n \int_{\mathbf{w}_i} \left[\left\{ \sum_{j=1}^{n_i} \log f_Y(y_i | \mathbf{w}_i, x_i; \boldsymbol{\phi}_Y^{(s)}) \right\} + \left\{ \sum_{r=0}^{m_i} \log f_T(d_{i,r} | \mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \boldsymbol{\phi}_T^{(s)}) \right\} \right. \\ &\quad \left. + \log f_w(\mathbf{w}_i | \boldsymbol{\phi}_w^{(s)}) \right] d\mathbf{w}_i. \end{aligned} \tag{47}$$

The posterior mean deviance across the MCMC samples and the deviance at the posterior means of the parameters can then be estimated as follow respectively

$$\begin{aligned} \overline{Dev}(\boldsymbol{\phi}^{(s)}) &= \frac{1}{S} \sum_{s=1}^S \left[-2 \sum_{i=1}^n \int_{\mathbf{w}_i} \left[\left\{ \sum_{j=1}^{n_i} \log f_Y(y_i | \mathbf{w}_i, x_i; \boldsymbol{\phi}_Y^{(s)}) \right\} + \left\{ \sum_{r=0}^{m_i} \log f_T(d_{i,r} | \mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \boldsymbol{\phi}_T^{(s)}) \right\} \right. \right. \\ &\quad \left. \left. + \log f_w(\mathbf{w}_i | \boldsymbol{\phi}_w^{(s)}) \right] d\mathbf{w}_i \middle| \mathcal{D}_{obs} \right]. \end{aligned} \tag{48}$$

$$\begin{aligned} Dev(\bar{\boldsymbol{\phi}}) &= -2 \sum_{i=1}^n \int_{\mathbf{w}_i} \left[\left\{ \sum_{j=1}^{n_i} \log f_Y(y_i | \mathbf{w}_i, x_i; \bar{\boldsymbol{\phi}}_Y) \right\} + \left\{ \sum_{r=0}^{m_i} \log f_T(d_{i,r} | \mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \bar{\boldsymbol{\phi}}_T) \right\} \right. \\ &\quad \left. + \log f_w(\mathbf{w}_i | \bar{\boldsymbol{\phi}}_w) \right] d\mathbf{w}_i. \end{aligned} \tag{49}$$

Then, effective number of model parameters can be deduced from equation (48) and (49) as

$$\begin{aligned}
 p_D = & \frac{1}{S} \sum_{s=1}^S \left[-2 \sum_{i=1}^n \int_{\mathbf{w}_i} \left[\left\{ \sum_{j=1}^{n_i} \log f_Y(y_i | \mathbf{w}_i, x_i; \phi_Y^{(s)}) \right\} + \left\{ \sum_{r=0}^{m_i} \log f_T(d_{i,r} | \mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \phi_T^{(s)}) \right\} \right. \right. \\
 & \left. \left. + \log f_w(\mathbf{w}_i | \phi_w^{(s)}) \right] d\mathbf{w}_i \Big| \mathcal{D}_{obs} \right] + 2 \sum_{i=1}^n \int_{\mathbf{w}_i} \left[\left\{ \sum_{j=1}^{n_i} \log f_Y(y_i | \mathbf{w}_i, x_i; \bar{\phi}_Y) \right\} + \left\{ \sum_{r=0}^{m_i} \log f_T(d_{i,r} | \mathbf{w}_i, \delta_{i,r+1}, v_{i,gh}; \bar{\phi}_T) \right\} \right. \\
 & \left. \left. + \log f_w(\mathbf{w}_i | \bar{\phi}_w) \right] d\mathbf{w}_i. \tag{50}
 \end{aligned}$$

The DIC in equation (44) can be easily calculated from equation (49) and (50).

Base on the DIC, the model with the smallest DIC value is considered to be the model that would best predict a replicated dataset which has the same structure as the current observed datasets. However, according to Geedipally at el., 2014, the model is penalized by $\overline{Dev}(\phi)$ which will decrease as the number of parameters in the model increases and p_D , which compensates for this effect by favouring the model with a smaller number of parameters. Therefore, it is very important to note that the way the model is parametrized will have an impact on the outcomes of the deviance information criterion values (Mwanyekange at el., 2018).

4. Simulation Study

In this section, we perform simulation under Markov Chain Monte Carlo (MCMC) to assess the performance of the proposed methodology. The data were generated from a joint multi-state survival model with a univariate longitudinal variable and a multi-state process (transition times process). We illustrate the proposed methodology by considering the following Linear Mixed Effects model

$$\begin{aligned}
 y_i(t_{ij}) &= x'_i(t_{ij})\beta + z'_i(t_{ij})\mathbf{w}_i + \epsilon_i(t_{ij}) \\
 &= \mu_i^*(t_{ij}) + \epsilon_i \\
 &= \beta_0 + \beta_1 x_i + \beta_2 t_{ij} + w_{i0} + w_{i1} t_{ij} + \epsilon_i(t_{ij}),
 \end{aligned}$$

where $\epsilon_i(t_{ij}) \sim \mathcal{N}(0, \sigma_\epsilon^2) = \mathcal{N}(0, 0.540)$, and the vector of random effects $\mathbf{w}_i = (w_{i0}, w_{i1})$ is assumed to follow multivariate normal distribution $\mathcal{MVN}(0, \mathbf{A})$, where $\mathbf{A} = \begin{pmatrix} \sigma_{w_{i0}}^2 & \rho\sigma_{w_{i0}}\sigma_{w_{i1}} \\ \rho\sigma_{w_{i0}}\sigma_{w_{i1}} & \sigma_{w_{i1}}^2 \end{pmatrix}$ with $\sigma_{w_{i0}} = 0.750$, $\sigma_{w_{i1}} = 0.150$ and $\rho = 0.400$.

We generated the data of $n = 200$ sample sizes and for each simulation, 1000 replicated datasets were considered. Each subject is expected to have at-most $n_i = 8$ number of repeated measurements taken at baseline and thereafter 7 visits, that is $t_{ij} = \{0.00, \dots, 2.50\}$. After this period, subjects are said to be censored non-informatively

For the multi-state semi-Markov sub-model, we have

$$\lambda_{i,gh}(d_{i,r} | \mathbf{w}_i) = \alpha_{gh} \exp \{ \gamma_{gh} d_{i,r} \} \exp \{ v'_{i,gh} \xi_{gh} + \mu_i^*(d_{i,r}) \psi_{gh} \}, \tag{Model I} \tag{51}$$

$$\lambda_{i,gh}(d_{i,r} | \mathbf{w}_i) = \alpha_{gh} \exp \{ \gamma_{gh} d_{i,r} \} \exp \{ v'_{i,gh} \xi_{gh} + \psi_{gh,int} w_{0i} + \psi_{gh,slope} w_{1i} \}, \tag{Model II} \tag{52}$$

We set the vectors of true parameters of Model I and II as follows:

$$\xi_{gh} = (\xi_{01}, \xi_{03}, \xi_{12}, \xi_{13}, \xi_{23})' = (-0.150, -0.250, -0.391, -0.183, -1.514)',$$

$$\alpha_{gh} = (\alpha_{01}, \alpha_{03}, \alpha_{12}, \alpha_{13}, \alpha_{23}) = (0.512, 0.145, 0.723, 0.601, 0.254)',$$

$$\gamma_{gh} = (\gamma_{01}, \gamma_{03}, \gamma_{12}, \gamma_{13}, \gamma_{23})' = (0.230, 0.254, 0.149, 1.543, 2.114)',$$

$$\psi_{gh} = (\psi_{01}, \psi_{03}, \psi_{12}, \psi_{13}, \psi_{23})' = (0.024, 0.216, 0.058, 0.045, 0.142)',$$

$$\psi_{gh,int} = (\psi_{01,int}, \psi_{03,int}, \psi_{12,int}, \psi_{13,int}, \psi_{23,int})' = (0.024, 0.484, 0.126, -0.273, -0.0713)',$$

$$\psi_{gh,slope} = (\psi_{01,slope}, \psi_{03,slope}, \psi_{12,slope}, \psi_{13,slope}, \psi_{23,slope}) = (-0.015, -0.238, -0.045, 0.214, 0.241)'.$$

The values of baseline covariates were simulated as follows:

$$x_i \sim \text{Bernoulli}(0.50) \text{ and } v_{i,gh} \sim \mathcal{N}(15, 45)$$

We generate the vector of true times $t_i^* = (t_{i,01}^*, t_{i,03}^*, t_{i,12}^*, t_{i,13}^*, t_{i,23}^*)'$ by first simulate the survival probability, $U_{i,gh}$, from Uniform(0,1) for each subject and solve for $t_{i,gh}^*$ from the following equation:

$$\exp \left\{ - \int_r^{t_{i,gh}^*} \alpha_{gh} \exp\{\gamma_{gh} t_{gh}\} \exp\{\xi_{gh} v_{i,gh} + \chi_{i,gh}(t_{gh}) \psi_{gh}\} dt_{gh} \right\} = U_{i,gh}, \quad r \geq 0, \quad h \in \{1, 2, 3\}.$$

In particular, we generate $t_{i,01}^*$ and $t_{i,03}^*$ from

$$\exp \left\{ - \int_0^{t_{i,0h}^*} \alpha_{0h} \exp\{\gamma_{0h} t_{0h}\} \exp\{\xi_{0h} v_{i,0h} + \chi_{i,0h}(t_{0h}) \psi_{0h}\} dt_{0h} \right\} = U_{i,0h}, \quad h \in \{1, 3\},$$

and compute the time spent in state 0 for subject i as $d_{i,0}^* = t_{i,0h}^* - 0$.

For the true transition times from state 1, $t_{i,12}^*$ and $t_{i,13}^*$ are generated from

$$\exp \left\{ - \int_{t_{i,0h}^*}^{t_{i,1h}^*} \alpha_{1h} \exp\{\gamma_{1h} t_{1h}\} \exp\{\xi_{1h} v_{i,1h} + \chi_{i,1h}(t_{1h}) \psi_{1h}\} dt_{1h} \right\} = U_{i,1h}, \quad h \in \{2, 3\},$$

and $d_{i,1}^* = t_{i,1h}^* - t_{i,0h}^*$.

Finally, $t_{i,23}^*$ is generated from

$$\exp \left\{ - \int_{t_{i,1h}^*}^{t_{i,23}^*} \alpha_{23} \exp\{\gamma_{23} t_{23}\} \exp\{\xi_{23} v_{i,23} + \chi_{i,23}(t_{23}) \psi_{23}\} dt_{23} \right\} = U_{i,23},$$

with $d_{i,2}^* = t_{i,23}^* - t_{i,12}^*$.

The censoring times were drawn from an uniform distribution, $C_i = U(0.05, 2.50)$. Then, we compute $d_{i,r} = \min(d_{i,r}^*, C_i)$, $\delta_{i,r+1} = 1$ if $d_{i,r}^* \leq C_i$ and 0 otherwise.

Note that, the sojourn times from Model II are generated in a similar manner.

The summary statistics of the estimated regression coefficients are presented in Table 1 including the true parameter values, bias (Bias), estimates, standard error (S.E), root mean squared error (RMSE) and coverage probabilities (CP) of 95% credible intervals (Cr.I). We compared joint modeling approach that includes the whole biomarker trajectory in multi-state survival sub-model to the joint modeling approach in which only random effects are shared between two sub-models. It is evident from Table 1 that estimators of the regression parameters of the longitudinal sub-model are relatively unbiased in both modeling approaches. Our focus is mainly on the multi-state semi-Markov process. From the same table, we clearly see that coverage probability dwells around 95% nominal level in both approaches. Overall, the joint modeling approach with shared random effects shows better performance than the modeling approach with current level of biomarker value.

Table 1. Results from simulation study examining the performance of the proposed joint model

Parameter	True	Model I					Model II				
		Estimate	Bias	S.E	RMSE	CP(%)	Estimate	Bias	S.E	RMSE	CP(%)
<i>n</i> = 200											
Longitudinal sub-model											
β_0	5.155	5.126	-0.029	0.078	0.118	0.935	5.139	-0.016	0.073	0.113	0.945
β_1	3.015	3.131	0.116	0.073	0.170	0.965	3.147	0.132	0.062	0.148	0.945
β_2	0.145	0.118	-0.027	0.048	0.210	0.924	0.194	0.049	0.041	0.197	0.972
σ_{w_0}	0.750	0.749	-0.001	0.075	0.145	0.968	0.762	0.012	0.073	0.134	0.988
σ_{w_1}	0.150	0.100	-0.050	0.008	0.161	0.916	0.105	-0.045	0.008	0.159	0.918
σ_ϵ	0.540	0.527	-0.013	0.014	0.043	0.950	0.545	0.005	0.008	0.154	0.994
ρ	0.400	0.411	0.011	0.163	0.084	0.974	0.432	0.032	0.013	0.164	0.952
Transition Intensity Sub-model											
ξ_{01}	-0.150	-0.163	-0.013	0.013	0.324	0.924	-0.153	-0.003	0.009	0.075	0.981
ξ_{03}	-0.250	-0.241	0.009	0.035	0.125	0.953	-0.243	0.007	0.027	0.119	0.974
ξ_{12}	-0.391	-0.401	-0.010	0.184	0.495	0.965	-0.398	-0.007	0.142	0.362	0.954
ξ_{13}	-0.183	-0.167	0.016	0.054	0.148	0.978	-0.175	0.008	0.007	0.035	0.968
ξ_{23}	-1.514	-1.483	0.031	0.028	0.187	0.952	-1.531	0.017	0.017	0.142	0.948
α_{01}	0.512	0.463	0.100	0.194	0.493	0.898	0.518	0.006	0.002	0.012	0.956
α_{03}	0.145	0.156	0.014	0.021	0.352	0.960	0.152	0.010	0.008	0.053	0.968
α_{12}	0.723	0.728	0.005	0.013	0.156	0.924	0.715	-0.008	0.009	0.206	0.918
α_{13}	0.601	0.634	0.033	0.034	0.192	0.896	0.622	0.021	0.010	0.342	0.914
α_{23}	2.522	2.494	-0.026	0.053	0.575	0.918	2.419	-0.109	0.026	0.153	0.908
γ_{01}	0.230	0.239	0.009	0.034	0.063	0.952	0.218	-0.012	0.012	0.094	0.964
γ_{03}	0.254	0.248	-0.006	0.043	0.228	0.946	0.258	-0.004	0.048	0.420	0.953
γ_{12}	0.149	0.145	-0.004	0.021	0.032	0.938	0.154	0.005	0.008	0.096	0.942
γ_{13}	1.543	1.571	0.028	0.043	0.045	0.960	1.548	0.005	0.013	0.190	0.968
γ_{23}	2.114	2.124	0.010	0.195	0.289	0.956	2.183	0.069	0.040	0.321	0.948
ψ_{01}	0.024	0.018	-0.006	0.038	0.134	0.986	-	-	-	-	-
ψ_{03}	0.216	0.219	0.003	0.025	0.503	0.948	-	-	-	-	-
ψ_{12}	0.058	0.045	0.013	0.011	0.184	0.968	-	-	-	-	-
ψ_{13}	0.045	0.048	-0.003	0.037	0.253	0.956	-	-	-	-	-
ψ_{23}	0.142	0.143	0.001	0.022	0.055	0.974	-	-	-	-	-
$\psi_{01,int}$	0.024	-	-	-	-	-	0.029	0.003	0.007	0.034	0.940
$\psi_{01,slope}$	-0.015	-	-	-	-	-	-0.023	0.008	0.026	0.055	0.912
$\psi_{03,int}$	0.484	-	-	-	-	-	0.479	-0.005	0.043	0.251	0.934
$\psi_{03,slope}$	-0.238	-	-	-	-	-	-0.232	-0.006	0.034	0.329	0.924
$\psi_{12,int}$	0.126	-	-	-	-	-	0.118	-0.008	0.009	0.085	0.958
$\psi_{12,slope}$	-0.045	-	-	-	-	-	-0.049	0.004	0.142	0.365	0.894
$\psi_{13,int}$	-0.273	-	-	-	-	-	-0.271	-0.002	0.008	0.074	0.914
$\psi_{13,slope}$	0.214	-	-	-	-	-	0.308	0.094	0.169	0.512	0.793
$\psi_{23,int}$	-0.713	-	-	-	-	-	-0.722	0.009	0.013	0.048	0.935
$\psi_{23,slope}$	0.241	-	-	-	-	-	0.236	-0.005	0.007	0.127	0.940

5. An Application to Real Data

5.1 Data

In this section, we consider the data from the Federation Francophone de Cancerologie Digestive (FFCD) 2000-2005 Multi-center phase III clinical trial of patients diagnosed with metastatic colorectal cancer. The study was conducted between February 2002 and January 2007 in France by FFCD. The main aim of the study was to examine the efficacy of two treatment effects: Sequential arm (S) and Combination arm (C). For the purpose of the present study, we consider datasets presented in Król et al. (2016) and Krol et al. (2017) in which 150 patients were randomly selected from the same clinical trial. The data contains individual follow-up of tumor size measure (sum of the longest diameters of target lesions) and time of new lesions (recurrent events). Moreover, some baseline covariates (age, WHO performance status and previous resection), treatment arm (Combination vs Sequential), time of death or the last observed time for right censored. A total of 906 tumor size measurements were recorded at subject specific follow-up. During the study period, 289 cases

of recurrence and 121 death were also recorded. In this study, we chose to apply our joint multi-state semi-Markov model to this datasets including only patients up to their second recurrence.

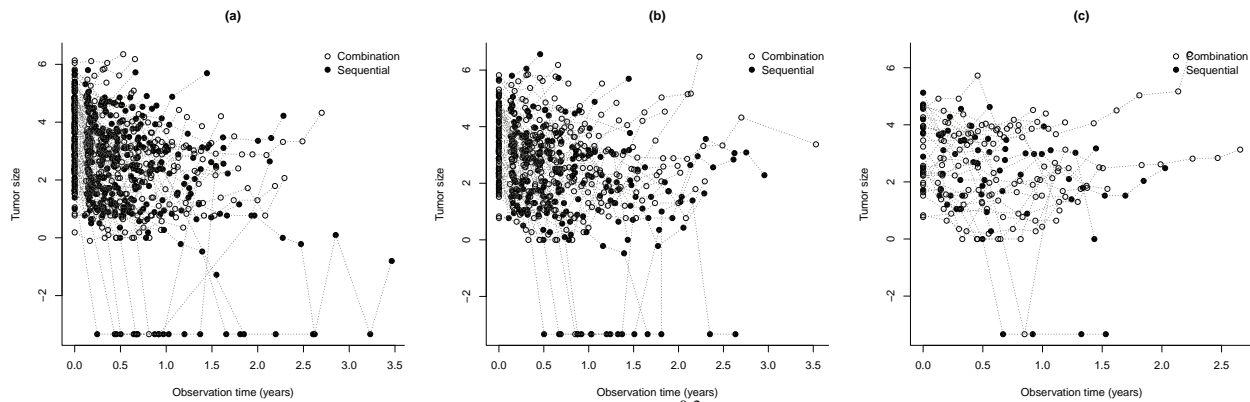


Figure 3. Individual profiles of tumor size ($SLD^* = (SLD^{0.3} - 1)/0.3$) measured during a visit. (a) Profiles for individuals expecting to directly transit from state 0 to either state 1 or state 3. (b) Profiles for individuals expecting to directly transit from state 1 to either state 2 or state 3. (c) Profiles for individuals expecting to transit from state 2 to either state 3 or censored

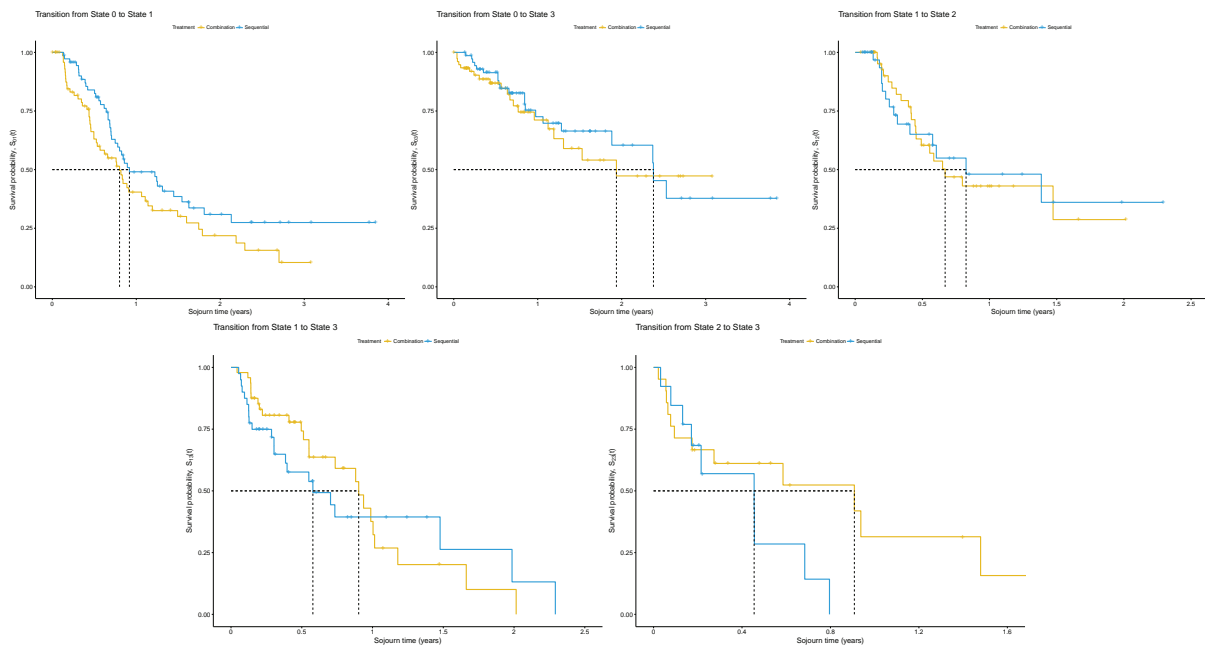


Figure 4. Kaplan-Meier curves for each transition

The multi-state semi-Markov process are presented through transitions between 4 states (see figure 1). At the first follow-up in state 0 (0 recurrence), a patient can either experience a transition to state 1 (1st recurrence) or to an absorbing state 3 (death). After the first recurrence, a patient may experience a transition to state 2 (2nd recurrence) or move to an absorbing state 3. After the second recurrence, we only considered the transition to an absorbing state 3, that is, patients with 3rd recurrence are ignored in this study. In total, 90 patients experienced transition to state 1, 35 patients had experienced second recurrence after the first recurrence. 41 patients die without experiencing any recurrence, 41 and 22 patients die after first and second recurrence respectively.

5.1.1 Longitudinal Sub-model

The following covariates are included in the longitudinal process (Linear mixed effect model); Treatment (1: Sequential arm, 2: Combination arm), Age (1: < 60 years, 2: 60 – 69 years, 3: > 69 years), who.PS (1: status 0, 2: status 1 and 3: status 2) and year the observation time. That is

$$\begin{aligned}
 y_i(t_{ij}) &= \mu_i^*(t_{ij}) + \epsilon_{ij} \\
 &= \beta_0 + \beta_1 age_i + \beta_2 who.PS_i + \beta_3 year_{ij} \times treatment_i + w_{i0} + w_{i1} \times year_{ij} + \epsilon_{ij},
 \end{aligned} \tag{53}$$

where y_i denotes tumor size, which is the transformed sum of the longest diameter ($SLD^* = (SLD^{0.3} - 1)/0.3$) measured during the visit, the random effect $w_i = (w_{i0}, w_{i1})'$ is assumed to be normal distributed with mean zero and variance-covariance matrix A , $w_i \sim \mathcal{MVN}(0, A)$ and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

5.1.2 Multi-state Semi-Markov Sub-model

The following proportional transition intensities model includes previous resection (prev.resection (0: No, 1: Yes)) covariate, gap-time (duration) and the function of current values of tumor size-(Model I)

$$\lambda_{i,gh}(d_{i,r}) = \lambda_{0,gh}(d_{i,r}) \exp \left\{ \xi_{0,gh} + \xi_{1,gh} Prev.res_{i,gh} + \psi_{gh} \mu_i^*(d_{i,r}) \right\}, \quad \forall r \geq 0, \tag{54}$$

where $\lambda_{0,gh}(d_{i,r}) = \alpha_{gh} \exp \left\{ \gamma_{gh} d_{i,r} \right\}$ is the Gompertz baseline function and $d_{i,r}$ is the sojourn time in each transient state.

Model II, use the shared random effects in the transition intensities function

$$\lambda_{i,gh}(d_{i,r}) = \lambda_{0,gh}(d_{i,r}) \exp \left\{ \xi_{0,gh} + \xi_{1,gh} Prev.res_{i,gh} + \psi_{gh,int} w_{i,0} + \psi_{gh,slope} w_{i,1} \right\}, \quad \forall r \geq 0. \tag{55}$$

In our Gibbs algorithm, we ran three parallel MCMC chains with different random generated initial values for 205 000 iterations. In order to avoid the influence of pre-convergence in our final posterior inference, 55 000 iterations were discarded as burn-in. Hence, our posterior inferences are based on the last 150 000 iterations. For the prior distributions, we used standard non-informative prior distributions for the parameters. That is, β , γ_{gh} , ξ_{gh} and ψ_{gh} are assumed to have normal distributions with mean zero and large variances ($\beta \sim \mathcal{N}(0, 100)$, $\gamma_{gh} \sim \mathcal{N}(0, 100)$, $\xi_{gh} \sim \mathcal{N}(0, 100)$, $\psi_{gh} \sim \mathcal{N}(0, 100)$). Similarly, we take standard normal prior for the association parameters $\psi_{gh,int}$ and $\psi_{gh,slope}$ in Model II. For the shape parameter in the baseline transition intensity function α_{gh} , an independent gamma prior. As stated in previous section, an inverse wishart prior distribution was taken ($A \sim Inv - Wishart(A_0, \nu_0)$), where A_0 is the prior distribution for covariance matrix with ν_0 degree of freedom. Finally, inverse gamma was chosen for the error variance $\sigma_\epsilon^2 = 1/\tau$, where $\tau \sim \Gamma(0.01, 0.01)$.

Table 2. Posterior estimates of the longitudinal sub-model for the two models

Transition	Parameter	Model I			Model II		
		Estimate	S.E	95%Cr.I	Estimate	S.E	95%Cr.I
0 → 1	β_0 (Intercept)	2.923	0.010	(2.047,3.846)	2.936	0.010	(2.057,3.859)
	β_1 (Age)	-0.092	0.004	(-0.394,0.227)	-0.071	0.003	(-0.396,0.243)
	β_2 (Who.PS)	0.287	0.004	(-0.084,0.640)	0.255	0.003	(-0.118,0.610)
	β_3 (Year vs Trt)	-0.590	0.003	(-0.838,-0.333)	-0.607	0.002	(-0.884,-0.328)
	ρ	0.036	0.001	(-0.171,0.242)	0.042	0.000	(-0.165,0.248)
	$\sigma_{w_{i0}}$	2.522	0.001	(1.955,3.232)	2.514	0.001	(1.953, 3.222)
	$\sigma_{w_{i1}}$	3.180	0.002	(2.235,4.442)	3.160	0.003	(2.222,4.409)
	σ_{ϵ}^2	0.820	0.001	(0.771,0.873)	0.821	0.000	(0.772,0.874)
0 → 3	β_0 (Intercept)	2.869	0.010	(1.969,3.767)	3.127	0.010	(2.254,3.980)
	β_1 (Age)	-0.095	0.004	(-0.425,0.235)	-0.086	0.003	(-0.381,0.213)
	β_2 (Who.PS)	0.317	0.004	(-0.031,0.660)	0.173	0.004	(-0.173,0.527)
	β_3 (Year vs Trt)	-0.528	0.003	(-0.769,-0.271)	-0.684	0.003	(-0.968,-0.403)
	ρ	0.051	0.001	(-0.158,0.258)	0.021	0.001	(-0.185,0.230)
	$\sigma_{w_{i0}}$	2.526	0.000	(1.960,3.237)	2.534	0.000	(1.965,3.249)
	$\sigma_{w_{i1}}$	3.13	0.003	(2.204,4.369)	3.308	0.003	(2.309,4.652)
	σ_{ϵ}^2	0.822	0.000	(0.772,0.876)	0.821	0.001	(0.772,0.876)
1 → 2	β_0 (Intercept)	3.003	0.019	(1.540,4.545)	2.991	0.018	(1.449,4.453)
	β_1 (Age)	-0.479	0.002	(-0.655,-0.299)	-0.480	0.001	(-0.655,-0.305)
	β_2 (Who.PS)	0.332	0.002	(0.149,0.522)	0.334	0.001	(0.151,0.519)
	β_3 (Year vs Trt)	-1.045	0.004	(-1.367,-0.715)	-1.041	0.002	(-1.368,-0.718)
	ρ	0.401	0.003	(-0.098,0.760)	0.412	0.002	(-0.076,0.764)
	$\sigma_{w_{i0}}$	2.849	0.001	(1.995,4.203)	2.854	0.006	(2.000,4.180)
	$\sigma_{w_{i1}}$	3.110	0.006	(2.154,4.607)	2.977	0.005	(2.101,4.331)
	σ_{ϵ}^2	1.482	0.000	(1.398,1.573)	1.483	0.000	(1.398,1.574)
1 → 3	β_0 (Intercept)	2.996	0.020	(1.433,4.633)	2.990	0.018	(1.523,4.426)
	β_1 (Age)	-0.478	0.002	(-0.651,-0.299)	-0.490	0.001	(-0.666,-0.314)
	β_2 (Who.PS)	0.329	0.002	(0.142,0.520)	0.338	0.001	(0.155,0.522)
	β_3 (Year vs Trt)	-1.051	0.004	(-1.391,-0.725)	1.089	0.001	(0.751,1.423)
	ρ	0.376	0.003	(-0.133,0.749)	0.442	0.003	(-0.045,0.779)
	$\sigma_{w_{i0}}$	8.746	0.044	(4.107,18.23)	8.536	0.046	(3.991,17.96)
	$\sigma_{w_{i1}}$	9.225	0.037	(4.424,18.72)	9.180	0.038	(4.399,18.810)
	σ_{ϵ}^2	1.483	0.000	(1.398,1.574)	1.483	0.000	(1.398,1.573)
2 → 3	β_0 (Intercept)	2.402	0.053	(-1.443,5.984)	3.129	0.051	(-1.011,3.973)
	β_1 (Age)	-0.166	0.004	(-0.437,0.113)	-0.161	0.001	(-0.441,0.120)
	β_2 (Who.PS)	0.367	0.004	(0.081,0.653)	0.372	0.001	(0.086,0.657)
	β_3 (Year vs Trt)	0.780	0.007	(0.211,1.304)	0.754	0.003	(0.203,1.298)
	ρ	0.368	0.004	(-0.324,0.845)	0.368	0.004	(-0.388,0.841)
	$\sigma_{w_{i0}}$	21.23	0.231	(6.896,58.610)	0.089	0.001	(0.023,0.203)
	$\sigma_{w_{i1}}$	21.1	0.173	(7.138,58.420)	0.089	0.000	(0.025,0.198)
	σ_{ϵ}^2	1.417	0.000	(1.287,1.561)	1.417	0.000	(1.288,1.562)

Table 3. Posterior estimates of the Multi-state survival sub-model for the two modeling approaches

Transition	Parameter	Model I			Model II		
		Estimate	S.E	95%Cr.I	estimate	S.E	95%Cr.I
0 → 1	ξ_{01} (Prev.res)	-0.200	0.005	(-0.622,0.225)	-0.224	0.002	(-0.661, 0.218)
	γ_{01} (Duration)	-0.010	0.004	(-0.372,0.344)	-0.058	0.001	(-0.378,0.243)
	ψ_{01}	0.092	0.002	(-0.051,0.241)	-	-	-
	$\psi_{12,int}$	-	-	-	0.125	0.001	(-0.041,0.298)
	$\psi_{12,slope}$	-	-	-	-0.035	0.002	(-0.276,0.241)
	α_{01}	0.619	0.005	(0.303,1.131)	0.797	0.001	(0.519,1.144)
0 → 3	ξ_{03} (Prev.res)	-0.499	0.008	(-1.141,0.146)	-0.764	0.003	(-1.519,-0.044)
	γ_{03} (Duration)	0.252	0.006	(-0.258,0.722)	0.245	0.002	(-0.254,0.748)
	ψ_{03}	0.290	0.003	(0.035, 0.574)	-	-	-
	$\psi_{12,int}$	-	-	-	0.495	0.002	(0.146,0.892)
	$\psi_{12,slope}$	-	-	-	-0.328	0.002	(-0.733,0.080)
	α_{03}	0.170	0.002	(0.042,0.414)	0.321	0.001	(0.146,0.563)
1 → 2	ξ_{12} (Prev.res)	-0.183	0.009	(-0.945,0.622)	-0.226	0.005	(-1.029,0.581)
	γ_{12} (duration)	0.163	0.015	(-0.950,1.496)	0.073	0.005	(-0.868,1.004)
	ψ_{12}	0.091	0.005	(-0.298,0.496)	-	-	-
	$\psi_{12,int}$	-	-	-	0.087	0.004	(-0.340,0.523)
	$\psi_{12,slope}$	-	-	-	-0.054	0.004	(-0.479,0.353)
	α_{12}	0.710	0.011	(0.089,1.817)	0.765	0.003	(0.319,1.413)
1 → 3	ξ_{13} (Prev.res)	-0.391	0.008	(-1.061,0.255)	-0.481	0.007	(-1.395,0.337)
	γ_{13} (Duration)	0.670	0.011	(-0.125,1.783)	1.029	0.009	(0.123,2.137)
	ψ_{13}	0.056	0.004	(-0.235,0.405)	-	-	-
	$\psi_{13,int}$	-	-	-	-0.236	0.007	(-0.863,0.394)
	$\psi_{13,slope}$	-	-	-	0.209	0.007	(-0.365,0.842)
	α_{13}	0.857	0.011	(0.172,1.914)	0.869	0.003	(0.423,1.482)
2 → 3	ξ_{23} (Prev.res)	-1.744	0.022	(-3.579,-0.175)	-1.725	0.020	(-3.894, -0.006)
	γ_{23} (Duration)	2.577	0.047	(-0.355, 6.507)	2.21	0.028	(-0.220,5.452)
	ψ_{23}	0.076	0.011	(-0.621, 0.856)	-	-	-
	$\psi_{13,int}$	-	-	-	-0.734	0.021	(-2.504,0.709)
	$\psi_{13,slope}$	-	-	-	0.298	0.014	(-0.597,1.638)
	α_{23}	3.408	0.100	(0.202,13.530)	2.612	0.054	(0.336,8.007)

Table 4. Deviance Information Criterion

Model	$\overline{Dev(\hat{\phi})}$	$Dev(\bar{\hat{\phi}})$	p_D	DIC
I	2197.0	2169.0	28.44	2226.0
II	2177.0	2153.0	24.17	2201.0

Next, we present results obtained from the samples posterior distribution. Table 2 presents the posterior estimates long with their standard errors (S.E) and 95% credible intervals (Cr.I) for longitudinal sub-model for both models. We clearly see that the joint modeling approach with the current value of biomarker and joint modeling approach with shared random effects are fairly close to each other in terms of parameter estimation. We observed that age of patients is only significantly associated with tumor size in both models for patients transiting from state 1 to state 3, since the 95% credible interval for β_1 does not include the value 0. However, this has negative relationship to tumor size, implying that, as patient’s age increases, the tumor size of the patient will decrease. Moreover, the time dependent treatment effect seems to a significant factor tumor size in all transitions for both both models. On the other hand, the performance status (Who.PS) indicate significant effect on the tumor size for 1 → 2, 1 → 3 and 2 → 3 transitions provided that the 95% credible interval for parameter β_2 does not include the value 0. In all the transitions observed, random effects have 95% credible intervals excluding value 0, implying significant heterogeneity for tumor size trajectory.

For the transition intensity process, the covariate of interest were duration and previous resection. From Table 3, we notice that patients previous resection had a negative effect on the transition intensity. However, 95% credible intervals for this

parameter ξ_{gh} in both models include value 0 for $0 \rightarrow 1$, $0 \rightarrow 3$, $1 \rightarrow 2$, and $1 \rightarrow 3$, implying that the negative effect is not significant. Time spent in each transient state is expected to have association in respect to different length of times moving from one state to the another. However, it is worth mentioning that sojourn time is negatively associated with transition intensity from transient states to the next possible transient state and positively associated for transitions from transient states to an absorbing state. The association between the longitudinal process and transition intensity process for $0 \rightarrow 1$, $0 \rightarrow 3$, $1 \rightarrow 2$ and $2 \rightarrow 3$ showed positive strength. Nevertheless, studies by Król et al. (2016) and Król et al. (2017) gave some important details on the study. It can be asserted from their studies that, performance status (Who.PS) show a strong effect on the progression to both absorbing state (death) and transient state(new lesions). Therefore, we can confidently say that this study provides some additional insights to the study of longitudinal outcomes and survival times in joint modeling framework.

Finally, we compared the two joint modeling approaches using the Deviance Information Criterion (DIC). We observed the DIC values at each transition. From Table 4, it is evident that joint model with shared random effects appears to be a better model with lower DIC values.

6. Discussion

Joint models of longitudinal and multi-state data are less explored in literature. Hence, in this article, we have discussed the Bayesian approach for joint modeling of longitudinal and multi-state data. We have used linear mixed effect model for the longitudinal process and semi-proportional transition intensity function for the multi-state process. In addition, we assume that multi-state data exhibit a semi-Markov process whereby transition intensity from a transient state depends on the sojourn/holding time at which subject stayed at the current state. Furthermore, the sojourn time in each transient state is assumed to follow Gompertz distribution. We have used different formulations that illustrates how the two processes are linked together. In the first formulation (Model I), the multivariate function which represents the true value of the biomarker is included in the transition intensity function. While in the second formulation (Model II), the multivariate function represents the random effects that are included in the transition intensity function.

We conducted a simulation study in order to assess the performance of the proposed joint modeling approach. Furthermore, we applied the proposed joint modeling approaches to a sample of patients from colorectal phase III clinical trial FFC02000-05 dataset with the aim of identifying prognostic factors that are significantly associated with either longitudinal or transition intensity process. Generally, small differences between the two models were observed for the parameter estimates. We have found out that the proposed methods behave well in terms of parameter estimation as its posterior estimates converges to their target distributions.

One limitation of our study may be the number of states. We have only looked at four states because it will be much complex with many states. Hence, this joint modeling approach can be extended to more complex cases, in which more than one longitudinal processes can be modelled together with multi-state survival process. However such extensions comes with computation and implementation challenges. Therefore, substantial work is needed in terms of software development for this growing field of research.

In short we have introduced the model that jointly analyse longitudinal and multi-state data with an assumption that multi-state data has a semi-Markov process. The simulation study in section 4 empirically demonstrated good performance in terms of parameter estimation. Moreover, we have demonstrated how this model can be applied to real life data. Even though this study cover only a small area in a field of joint modeling framework, we are confident to say that it had contributed to the existing literature.

Disclosure Statement

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Pan African University (PAUSTI) in collaboration with Jomo Kenyatta University of Agriculture and Technology (JKUAT). The authors are grateful to the anonymous reviewers for their valuable comments and suggestions.

References

- Ahuja, J. C., & Nash, S. W. (1967). The generalized Gompertz-Verhulst family of distributions. *The Indian Journal of Statistics, Series A, Sanhkya* 29 A, 141-156.
- Chi, Y. -Y., & Joseph, G. I. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2), 432-445. <https://doi.org/10.1111/j.1541-0420.2005.00448.x>
- de Castro, M., Chen, M. H., & Zhang, Y. (2015). Bayesian path specific frailty models for multi-state survival data with

- applications. *Biometrics*, 71(3), 760-771. <https://doi.org/10.1111/biom.12298>
- Dendale, P., De Keulenaer, G., Troisfontaines, P., Weytjens, C., Mullens, W., Elegeert, I., ... & Hansen, D. (2012). Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. *European journal of heart failure*, 14(3), 333-340. <https://doi.org/10.1093/eurjhf/hfr144>
- El-Gohary, A., Alshamrani, A., Al-Otaibi, A. N. (2013). The generalized Gompertz distribution. *Applied Mathematical Modelling*, 37(1-2), 13-24. <https://doi.org/10.1016/j.apm.2011.05.017>
- Faucett, C. L., & Duncan, C. T. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine*, 15(15), 1663-1685.
- Ferrer, L. et al. (2016). Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in medicine*, 35(22), 3933-3948. <https://doi.org/10.1002/sim.6972>
- Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand, J. F., & Daurès, J. P. (2005). A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(6), 825-833. <https://doi.org/10.1002/bimj.200410170>
- Foucher, Y., Giral, M., Soullillou, J.-P., & Daurès, J.-P. (2007). A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Statistics in Medicine*, 26(30), 5381-5393. <https://doi.org/10.1002/sim.3100>
- Geedipally, S. R., Lord, D., & Dhavala, S. S. (2014). A caution about using deviance information criterion while modeling traffic crashes. *Safety science*, 62, 495-498. <https://doi.org/10.1016/j.ssci.2013.10.007>
- Gompertz, B. (1824). On the nature of the function expressive of the law of human mortality and on the new mode of determining the value of life contingencies Phil. Trans. *The Royal Society*, 115, 513-580. <https://www.jstor.org/stable/107756>
- Guo, X., & Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1), 16-24. <https://doi.org/10.1198/0003130042854>
- Hougaard, P. (2012). *Analysis of multivariate survival data*. Springer Science & Business Media.
- Huang, X., Li, G., & Elashoff, R. M. (2010). A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and its interface*, 3(2), 185.
- Ibrahim, J. G., Chu, H., & Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16), 2796. <https://doi.org/10.1200/JCO.2009.25.0654>
- Król, A., & Loc, F. et al. (2016). 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. *Biometrics*, 72(3), pp. 907-916. <https://doi.org/10.1111/biom.12490>
- Król, A., Mauguen, A., Mazroui, Y., Laurent, A., Michiels, S., & Rondeau, V. (2017). Tutorial in joint modeling and prediction: A statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *arXiv preprint arXiv:1701.03675*. <https://doi.org/10.18637/jss.v081.i03>
- Lawrence, G. A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., & Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in medicine*, 34(14), 2181-2195. <https://doi.org/10.1002/sim.6141>
- McCrink, L. M., Marshall, A. H., & Cairns, K. J. (2013). Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, 81(2), 249-269. <https://doi.org/10.1111/insr.12018>
- Mwanyekange, J., Mwalili, S., & Ngesa, O. (2018). Bayesian Inference in a Joint Model for Longitudinal and Time to Event Data with Gompertz Baseline Hazards. *Modern Applied Sciences*, 12(9), 159-172. <https://doi.org/10.5539/mas.v12n9p159>
- Njagi, E. N., Rizopoulos, D., Molenberghs, G., Dendale, P., & Willekens, K. (2013). A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients. *Statistical Modelling*, 13(3), 179-198. <https://doi.org/10.1177/1471082X13478880>
- Rizopoulos, D., & Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12), 1366-1380. <https://doi.org/10.1002/sim.4205>

Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der, L. A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639. <https://doi.org/10.1111/1467-9868.00353>

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3), 439-454. <https://www.jstor.org/stable/2061224>

Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330-339.

Yu, M., Taylor, J. M. G., & Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103(481), 178-187.

Zhang, D., Chen, M. H., Ibrahim, J. G., Boye, M. E., & 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(1), 121-133. <https://doi.org/10.1080/10618600.2015.1117472>

Appendix

Full Conditional Densities

This section presents details for the full conditional distributions of the parameters used in the Gibbs sampling algorithm, where $p[\cdot|\cdot]$ represent conditional densities.

i) The full conditional distribution of β is given by

$$p[\beta|rest] \propto \prod_{i=1}^n \left[\exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{n_i} [y_{ij} - (x'_i(t_{ij})\beta + z'_i(t_{ij})\mathbf{w}_i)]^2 \right\} \right. \\ \times \prod_{h \neq g} [\alpha_{gh} \exp \{ \gamma_{gh} d_{i,r} + v'_{i,gh} \xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r}) \psi_{gh} \}]^{\delta_{i,r+1}} \\ \left. \times \exp \left\{ -\sum_{h \neq g} \Lambda_{i,gh}(d_{i,r} | v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r})) \right\} \right] \times \exp \left\{ -\frac{1}{2} (\beta - \mu_\beta)' \tau_\beta^{-1} (\beta - \mu_\beta) \right\},$$

where

$$\Lambda_{i,gh}(d_{i,r} | v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r})) = \int_0^{d_{i,r}} \alpha_{gh} \exp(\gamma_{gh} s + v'_{i,gh} \xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; s) \psi_{gh}) ds$$

ii) The full conditional of $1/\sigma_\epsilon^2$ takes the form

$$p \left[\frac{1}{\sigma_\epsilon^2} | rest \right] \propto \prod_{i=1}^n (2\pi\sigma_\epsilon^2)^{-n_i/2} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{n_i} [y_{ij} - (x'_i(t_{ij})\beta + z'_i(t_{ij})\mathbf{w}_i)]^2 \right\} \\ \times (\sigma_\epsilon^2)^{-a_1-1} \exp \left\{ -\frac{b_1}{\sigma_\epsilon^2} \right\}$$

iii) The full conditional distribution of the inverse covariance variance \mathbf{A}^{-1} takes the form

$$p[\mathbf{A}^{-1}|rest] \propto \prod_{i=1}^n |\mathbf{A}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{w}'_i \mathbf{A}^{-1} \mathbf{w}_i) \right\} \times |\mathbf{A}|^{-(v_0+q+1)/2} \exp \left\{ -\frac{1}{2} tr(\mathbf{A}_0 \mathbf{A}^{-1}) \right\}$$

iv) The full conditional distribution of shape parameter α_{gh} in the survival sub-model is given by

$$p[\alpha_{gh}|rest] \propto \prod_{i=1}^n \prod_{r=1}^{m_i} [\alpha_{gh} \exp \{ \gamma_{gh} d_{i,r} + v'_{i,gh} \xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r}) \psi_{gh} \}]^{\delta_{i,r+1}} \\ \times \exp \left\{ -\sum_{h \neq g} \Lambda_{i,gh}(d_{i,r} | v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r})) \right\} \times \alpha_{gh}^{a_{gh}-1} \exp \{ -\alpha_{gh} b_{gh} \}$$

v) The full conditional distribution of ψ_{gh} that quantifies association between the two processes is given by

$$p[\psi_{gh}|rest] \propto \prod_{i=1}^n \prod_{r=1}^{m_i} [\alpha_{gh} \exp\{\gamma_{gh}d_{i,r} + v'_{i,gh}\xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r})\psi_{gh}\}]^{\delta_{i,r+1}} \\ \times \exp\left\{-\sum_{h \neq g} \Lambda_{i,gh}(d_{i,r}|v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r}))\right\} \times \exp\left\{-\frac{1}{2}(\psi_{gh} - \mu_{\psi_{gh}})' \tau_{\psi_{gh}}^{-1}(\psi_{gh} - \mu_{\psi_{gh}})\right\}$$

vi) The full conditional distribution of γ_{gh} takes the form

$$p[\gamma_{gh}|rest] \propto \prod_{i=1}^n \prod_{r=1}^{m_i} [\alpha_{gh} \exp\{\gamma_{gh}d_{i,r} + v'_{i,gh}\xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r})\psi_{gh}\}]^{\delta_{i,r+1}} \\ \times \exp\left\{-\sum_{h \neq g} \Lambda_{i,gh}(d_{i,r}|v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r}))\right\} \times \exp\left\{-\frac{1}{2}(\gamma_{gh} - \mu_{\gamma_{gh}})' \tau_{\gamma_{gh}}^{-1}(\gamma_{gh} - \mu_{\gamma_{gh}})\right\}$$

vii) The full conditional distribution of ξ_{gh} is

$$p[\xi_{gh}|rest] \propto \prod_{i=1}^n \prod_{r=1}^{m_i} [\alpha_{gh} \exp\{\gamma_{gh}d_{i,r} + v'_{i,gh}\xi_{gh} + \chi'_{i,gh}(\mathbf{w}_i; d_{i,r})\psi_{gh}\}]^{\delta_{i,r+1}} \\ \times \exp\left\{-\sum_{h \neq g} \Lambda_{i,gh}(d_{i,r}|v_{i,gh}, \chi_{i,gh}(\mathbf{w}_i; d_{i,r}))\right\} \times \exp\left\{-\frac{1}{2}(\xi_{gh} - \mu_{\xi_{gh}})' \tau_{\xi_{gh}}^{-1}(\xi_{gh} - \mu_{\xi_{gh}})\right\}$$

These densities are obtained through Gibbs sampling algorithm by retaining only those terms in equation (36) involve β , σ_{ϵ}^2 , A^{-1} , α_{gh} , ξ_{gh} , γ_{gh} and ψ_{gh} . Note that "rest" denotes the vector of all parameters and data excluding the parameter to be estimated, e.g., $[\beta|\sigma_{\epsilon}^2, A^{-1}, \alpha_{gh}, \xi_{gh}, \gamma_{gh}, \psi_{gh}; \mathcal{D}_{obs}]$

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).