

# Bayesian Inference in a Joint Model for Longitudinal and Time to Event Data with Gompertz Baseline Hazards

Josua Mwanyekange<sup>1</sup>, Samuel Mwalili<sup>1,2</sup> & Oscar Ngesa<sup>1,2,3</sup>

<sup>1</sup> Pan African University Institute for Basic Science, Technology and Innovation(PAUSTI), Department of Mathematics(Statistics), Nairobi, Kenya

<sup>2</sup> Jomo Kenyatta University of Agriculture and Technology(JKUAT), Department of Statistics and Actuarial Science, Nairobi, Kenya

<sup>3</sup> Taita Taveta University, Nairobi, Kenya

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

Received: June 19, 2018

Accepted: June 25, 2018

Online Published: August 24, 2018

doi:10.5539/mas.v12n9p159

URL: <https://doi.org/10.5539/mas.v12n9p159>

## Abstract

Longitudinal and time to event data are frequently encountered in many medical studies. Clinicians are more interested in how longitudinal outcomes influences the time to an event of interest. To study the association between longitudinal and time to event data, joint modeling approaches were found to be the most appropriate techniques for such data. The approaches involves the choice of the distribution of the survival times which in most cases authors prefer either exponential or Weibull distribution. However, these distributions have some shortcomings. In this paper, we propose an alternative joint model approach under Bayesian prospective. We assumed that survival times follow a Gompertz distribution. One of the advantages of Gompertz distribution is that its cumulative distribution function has a closed form solution and it accommodates time varying covariates. A Bayesian approach through Gibbs sampling procedure was developed for parameter estimation and inferences. We evaluate the finite samples performance of the joint model through an extensive simulation study and apply the model to a real dataset to determine the association between markers(tumor sizes) and time to death among cancer patients without recurrence. Our analysis suggested that the proposed joint modeling approach perform well in terms of parameter estimations when correlation between random intercepts and slopes is considered.

**Keywords:** joint modeling, Longitudinal data, time to event data, Bayesian inference, gompertz distribution, gibbs sampling, MCMC

## 1. Introduction

In many clinical studies, longitudinal outcomes are collected alongside with time to event data. For example, we collect repeated patients' information such as heights, weights, blood pressures and many other variables over time and at the same time we are also interested in the time to an event of interest (e.g., death). For such data, the focus is always on how longitudinal outcomes influence the time to an event. Approaches for analyzing these two processes separately have been extensively discussed in the literature. The most common methods used to analyze longitudinal outcomes are the random effect model proposed by Laird and Ware (1982), linear mixed effects model by Verbeke (1997) and Generalized linear models by Laing & Zeger (1986), Zeger *et al.* (1988). On the other hand, event time process is analyzed by making use of cox proportional hazard models (Hougaard 2012). However, analysis of longitudinal and time to event process separately have received many criticisms as they bear inefficient and bias results. Therefore, it is very important to take into account both processes in analysis.

Over the past two decades, extensive efforts have been made on statistical methods that simultaneously analyze longitudinal outcomes and time to event data. These methods offers great advantages over separate analyzing of each process. A comprehensive overview of joint modeling for longitudinal and time to event data was given in Tsiatis & Davidian (2004). More basic concepts and methods for joint models can be found in Ibrahim *et al.* (2010). In all these reviews, the two processes are jointly modeled by linking them together through a common latent structure. That is the longitudinal and survival process both share the same random effects. Extensions to earlier developments of joint modeling framework has been proposed. More recent developments with applications includes, Sweeting & Thompson (2011), Rizopoulos *et al.* (2012), McCrink *et al.* (2013) and Yang *et al.* (2016). Some authors have already proposed joint models for multiple longitudinal outcomes and repeated events of different outcome (Chi & Ibrahim.,2006; Musero *et al.*,2015; Huang &

Elashoff.,2010).

Up to date, there are three frequently used types of estimation approaches for joint models of longitudinal and time to event data, namely: maximum likelihood approach, two stage approach and Bayesian approach. According to Lang (2011), parameters of the joint model can be simultaneously estimated by incorporate both processes in a log joint likelihood function and use EM algorithm. The EM algorithm is commonly used for likelihood inferences of joint models due to its stability and generality Tsiatis & Davidian (2004), Tseng *et al.* (2005). However, the method can be computationally challenging when unobserved random effects have high dimensions. The two stage approach comprises two stages. The first stage fits linear mixed effects model to the longitudinal data and parameter estimate are obtained based on the fitted model without considering the survival model. In the second stage, survival model is fitted separately with longitudinal values predicted in stage one used as true observed values. Even though, two stage approach seem to more computationally friendly, the approach may lead to biased inferences. On the other hand, Bayesian approach has been studied by a number of authors such as Faucett & Thomas (1996), Ibrahim *et al.* (2004), Martins *et al.* (2010), Haung *et al.* (2011) and Rizopoulos & Ghosh (2011). Joint models may have too many unknown parameters which in some cases can be difficult to solve analytically. Bayesian approach uses prior information by incorporate them into current analysis called posterior distribution. Then Markov chain samples can be drawn from posterior distribution for inference. Conversely, the approach may not provide good inferences as the results may heavily affected by the prior information. Nevertheless, Bayesian is easy to implement which make it more useful for inference of joint models.

Previous work on joint model for longitudinal outcomes and time to event data with times varying covariates considered linear mixed effects models for longitudinal process and a proportion hazards model with specified baseline function for time to event process, Yu *et al.* (2006); Song *et al.* (2002). The most frequently used beeline functions for proportional hazards models assumes that survival data follow exponential or Weibull distribution. Some authors suggested that proportional hazard model with Weibull baseline function may be very flexible when time dependent covariates are included in the model (Casellas, 2007). However, generating survival data from such model with time varying covariates can be too complex. According to Austin (2013), one of the disadvantages of Weibull proportional hazards model with time dependent covariates is that its cumulative distribution function can not be derived in a closed form. Therefore, to compute the cumulative incidence and the hazard function, we need to make use of numerical integration method. On the other hand, proportional hazard model with exponential baseline function also has disadvantages as it assumes constant hazards. For these reasons, researchers willing to generate survival times that involve time dependent covariates are advised to consider using Gompertz distribution event times data (Austin, 2012).

Proportional hazards models with baseline function from Gompertz distribution seems to be more flexible as one can compute survival function without high programming needs. The Gompertz distribution was first introduced by Gompertz (1825). It is described as one of the fundamental mathematical models that accurately represents survival function based on the laws of mortality. That is the force of mortality or survival times tends to increase exponentially over time. It has been extensively used as growth model in many cancer assessments. It is for these reasons that Gompertz distribution plays an important role in modeling human mortality. Its earlier applications can be found in Ahuj & Nash (1967).

To the best of our knowledge, there has been no joint model for longitudinal outcome and time to event data with proportional hazard model assumed to have Gompertz distribution properties. Therefore, in this article, we propose such model with time varying covariates under the framework of Bayesian inferences. Our joint model is made of two submodels linked together by common random effects. The first submodel is a linear mixed effects model for modeling true and unobserved markers. The interrelationship between measurements and subject specific effects is accounted by random intercept and slope. The second submodel is the proportional hazards model with Gompertz baseline function. To estimate the parameters of the joint model, we use both **R** and *WinBUGS*, Ntzoufras (2011) softwares. The rest of this article is organized as follows; In section 2 we present the notations and formulations of the joint model. Section 3 presents estimations and inferences, which includes joint likelihood and Bayesian parameter estimation procedures such as prior and joint posteriors specifications and model selection. In section 4, we present some simulation studies in order to access the performance of the model. Section 5, presents the application of the model to real datasets. The last section presents a discussion.

## 2. Method

### 2.1 Longitudinal Sub-Model

Suppose we have observations for  $n$  subjects under study. 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 analyze the longitudinal process, we define the distribution of  $y_{ij}$  by a Linear Mixed Effect(LME) model,

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

where,

$$\mu_i^*(t_{ij}) = \mathbf{x}'_i(t_{ij})\boldsymbol{\beta}_L + \eta_i(t_{ij})$$

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;  $\boldsymbol{\beta}_L$  is a  $(p \times 1)$  corresponding column vector of the fixed effect coefficients ( $\boldsymbol{\beta}_{L,s}$ );  $\eta_i(t_{ij}) = \mathbf{z}'_i(t_{ij})\mathbf{w}_i$ , where  $\mathbf{z}_i$  denotes  $(q \times 1)$  design matrix for the random effects;  $\mathbf{w}_i \sim \mathcal{MVN}(0, \mathbf{A})$  and  $\epsilon_i$  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:  $\mathbf{x}'_i(t_{ij})\boldsymbol{\beta}_L + \eta_i(t_{ij})$  such that  $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2 I_{n_i})$ . It should be noted here that  $\mathbf{A}$  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 others. The variance covariance matrix is defined as

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

or

$$\mathbf{A} = \begin{pmatrix} \sigma_{w_{i0}}^2 & 0 \\ 0 & \sigma_{w_{i1}}^2 \end{pmatrix}$$

with no correlation.

### 2.1 Time to Event Sub-Model

We assumed a proportional hazards model with Gompertz baseline function. Let

$$\lambda_i(t) = \lambda_0(t) \exp(\boldsymbol{\xi}' v_i + \psi \mu_i^*(t)) \tag{2}$$

denote the hazards function for subject  $i$  at time  $t$ , where  $\lambda_0(t) = \alpha \exp(\gamma t)$  is a baseline function with  $\alpha > 0$  and  $\gamma$  takes any value. When  $\gamma > 0$  or  $\gamma < 0$ , the hazards function in monotone(increasing or decreasing). On the other hand, when  $\gamma = 0$ , the hazards function is equivalent to hazards function from an exponential distribution.  $v_i$  represents the vector of prognostic factor associated with vector of coefficient  $\boldsymbol{\xi}$  and  $\psi$  quantifies the strength between the two processes. Furthermore, the cumulative hazards/incidences function is defined as follows:

$$\Lambda_i(t) = \int_0^t \alpha \exp(\gamma s) \exp(\boldsymbol{\xi}' v_i + \psi \mu_i^*(s)) \partial s \tag{3}$$

Equation (3) can be simplified as follows:

$$\Lambda_i(t) = \alpha \int_0^t \exp(\gamma s + \boldsymbol{\xi}' v_i + \psi \mu_i^*(s)) \partial s \tag{4}$$

From equation (4), we deduce

$$\Lambda_i(t) = \alpha \exp(\boldsymbol{\xi}' v_i) \left[ \frac{\exp((\gamma + \psi \mu_i^*)s)}{\gamma + \psi \mu_i^*} \right]_0^t = \alpha \exp(\boldsymbol{\xi}' v_i) \left[ \frac{\exp((\gamma + \psi \mu_i^*)t) - 1}{\gamma + \psi \mu_i^*} \right] \tag{5}$$

So, the inverse cumulative hazards function is given as

$$\Lambda_i^{-1}(t) = \frac{1}{\gamma + \psi \mu_i^*} \log \left( \frac{(\gamma + \psi \mu_i^*)t}{\alpha \exp(\boldsymbol{\xi}' v_i)} + 1 \right) \tag{6}$$

and the individual survival function is expressed as

$$S_i(t) = \exp \left( - \left( \alpha \exp(\boldsymbol{\xi}' v_i) \left[ \frac{\exp((\gamma + \psi \mu_i^*)t) - 1}{\gamma + \psi \mu_i^*} \right] \right) \right). \tag{7}$$

Consequently, the individual event times can be generated as

$$T_i = \frac{1}{\gamma + \psi\mu_i^*} \log \left( \frac{(\gamma + \psi\mu_i^*)(-\log(U))}{\alpha \exp(\xi' v_i)} (-\log(U)) + 1 \right), \text{ if } -\log(U) < \alpha \exp(\xi' v_i) \left[ \frac{\exp((\gamma + \psi\mu^*)t) - 1}{\gamma + \psi\mu_i^*} \right], \quad (8)$$

where  $T_i$  stands for the survival for subject  $i$  and  $U$  is a random variable uniformly distributed with  $U \sim Uniform(0, 1)$ .

The distribution function of the time to event based on the proportional hazards function is

$$F_i(t) = 1 - \exp \left( - \left( \alpha \exp(\xi' v_i) \left[ \frac{\exp((\gamma + \psi\mu^*)t) - 1}{\gamma + \psi\mu_i^*} \right] \right) \right) \quad (9)$$

and the probability density function of time to event is

$$f_i(t) = \alpha \exp(\gamma t + \xi' v_i + \psi\mu_i^*(t)) \exp \left( - \left( \alpha \exp(\xi' v_i) \left[ \frac{\exp((\gamma + \psi\mu^*)t) - 1}{\gamma + \psi\mu_i^*} \right] \right) \right). \quad (10)$$

### 3. Estimation and Inferences

#### 3.1 The Joint Likelihood Function

Let  $C_i$  be the censoring time for subject  $i$ . Then, we have  $T_i = \min(T_i^*, C_i)$  denotes the  $i^{th}$  subject's observation time, where  $T_i^*$  is the true event time for that subject. Furthermore, we denote  $\delta_i = I(T_i = T_i^*)$ , where  $I$  is the indicator function. It follows that,  $\delta_i$  is equal to 1 if  $T_i$  is an event time and 0 if  $T_i$  is right censored for subject  $i$ . Letting  $\phi = \{\beta_L, \alpha, \psi, \gamma, A, \sigma_\epsilon^2, \xi\}$  denote the vector of all parameters in equation to be estimated. Note that  $y_{ij}$  is the longitudinal outcomes for subject  $i$  measured at time  $t_{ij}(j = 1, 2, \dots, n_i)$ . Then, given subject-specific random effects, the two processes are said to be independent of each other. Hence, the joint likelihood function of the longitudinal and time to event sub-models based on all observed data  $\mathcal{D}_{obs} = \{y_{ij}, T_i^*, \delta_i, x_i, v_i\}$  is defined by

$$\mathcal{L}(\phi | \mathbf{w}_i, \mathcal{D}_{obs}) = \int \prod_{i=1}^n \left[ \left\{ \prod_{j=1}^{n_i} f_y(y_{ij} | \mathbf{w}_i, \phi) \right\} \times \{f_T(T_i^*, \delta_i | \mathbf{w}_i, \phi)\} \times f_w(\mathbf{w}_i | \phi) \right] \partial \mathbf{w}_i, \quad (11)$$

where

$$f_y(y_{ij} | \mathbf{w}_i, \phi) = (2\pi\sigma_\epsilon^2)^{-n_i/2} \exp \left\{ (-1/2\sigma_\epsilon^2) \left[ y_{ij} - (x_i'(t_{ij})\beta_L + \eta_i(t_{ij})) \right]^2 \right\} \quad (12)$$

is the probability density function of the longitudinal outcomes conditionally on the random effects  $\mathbf{w}_i$  and

$$f_w(\mathbf{w}_i | \phi) = (2\pi)^{q/2} |A|^{-1/2} \exp \left( (-1/2) (\mathbf{w}_i' A^{-1} \mathbf{w}_i) \right) \quad (13)$$

is defined as the probability density function of the random effects with  $q$  as dimension of the covariance matrix  $A$ . The likelihood of the survival sub-part  $f_T(T_i^*, \delta_i | \mathbf{w}_i, \phi) = \lambda_i(T_i^* | \mathbf{w}_i, \phi)^{\delta_i} \times S_i(T_i^* | \mathbf{w}_i, \phi)$  which is a version of equation (7) in section 2.

#### 3.2 Bayesian Inferences

The solution to the joint likelihood function (11) can not be achieved with the normal standard maximum likelihood procedure due to the fact that it involves integration of the longitudinal and survival components over the subject-specific random effects  $\mathbf{w}_i$ . This requires high programming needs. The best way to overcome such difficulty is to make use of Bayesian approach. Faucett and Thomas (1996), proposed a Bayesian Markov Chain Monte Carlo (MCMC) approach that simulates samples from posterior distribution and estimate all the unknown parameters by using non-informative prior  $p(\phi)$ .

Taking equation (11) and  $p(\phi)$ , we can define joint posterior of  $\phi$  as product of the likelihood of the observed data and some priors as

$$p(\phi | \mathcal{D}_{obs}) = \frac{\mathcal{L}(\phi | \mathbf{w}_i, \mathcal{D}_{obs}) p(\phi)}{\int \prod_{i=1}^n f_{y,T}(y_{ij}, T_i^* | \phi, \delta_i, x_i, v_i) p(\phi) \partial \phi}. \quad (14)$$

If the observed data is fixed, then we have

$$p(\phi | \mathcal{D}_{obs}) \propto \mathcal{L}(\phi | \mathbf{w}_i, \mathcal{D}_{obs}) p(\phi). \quad (15)$$

Therefore, this translate our joint posterior distribution of the parameter  $\phi$  into

$$p(\phi|\mathcal{D}_{obs}) \propto \prod_{i=1}^n \left[ \left\{ \prod_{j=1}^{n_i} f_y(\mathbf{y}_{ij}|\mathbf{w}_i, \phi) \right\} \times \{f_T(T_i^*, \delta_i|\mathbf{w}_i, \phi)\} \times f_w(\mathbf{w}_i; \phi) \right] \times p(\phi). \tag{16}$$

Due to computational complexity, log joint posterior distribution is more preferable. Hence,

$$\log p(\phi|\mathcal{D}_{obs}) \propto \sum_{i=1}^n \left[ \left\{ \sum_{j=1}^{n_i} \log f_y(\mathbf{y}_{ij}|\mathbf{w}_i, \phi) \right\} + \{\log f_T(T_i^*, \delta_i|\mathbf{w}_i, \phi)\} + \log f_w(\mathbf{w}_i; \phi) \right] + p(\phi). \tag{17}$$

### 3.2.1 Full Conditional Distributions

In order to implement the Bayesian procedure, we need the full conditional distribution of each of unknown parameters of the model. Then Gibbs sampler can be used to generate MCMC samples from the joint posterior density  $p(\phi|\mathcal{D}_{obs})$ . In Bayesian framework, this procedure involves iteratively sampling from its full conditional distributions with the remaining components(others) fixed to their current value. Hence, the full conditional distribution of the coefficient  $\beta_L$  of the linear mixed effect model is

$$p(\beta_L|others) \propto \prod_{i=1}^n \exp \left( (-1/2\sigma_\epsilon^2) \sum_{j=1}^{n_i} [\mathbf{y}_{ij} - \mu_i^*(t_{ij})]^2 \right) \times (\alpha \exp(\gamma t + \xi v_i + \psi \mu_i^*(t)))^{\delta_i} \times \exp(-\Lambda_i(t)) \times \exp \left( (-1/2)(\beta_L - \mu_{\beta_L})' \tau_{\beta_L}^{-1} (\beta_L - \mu_{\beta_L}) \right), \tag{18}$$

where  $\mu_{\beta_L}$  and  $\tau_{\beta_L}$  are the parameters of the prior of  $\beta_L$ .

The full conditional distribution of  $\xi$  is

$$p(\xi|others) \propto \prod_{i=1}^n (\alpha \exp(\gamma t + \xi v_i + \psi \mu_i^*(t)))^{\delta_i} \times \exp(-\Lambda_i(t)) \times \exp \left( (-1/2)(\xi - \mu_\xi)' \tau_\xi^{-1} (\xi - \mu_\xi) \right), \tag{19}$$

where  $\mu_\xi$  and  $\tau_\xi$  are the parameters of the independent normal prior of  $\xi$ .

The full conditional distribution of  $\gamma$  takes the form

$$p(\gamma|others) \propto \prod_{i=1}^n (\alpha \exp(\gamma t + \xi v_i + \psi \mu_i^*(t)))^{\delta_i} \times \exp(-\Lambda_i(t)) \times \exp \left( (-1/2)(\gamma - \mu_\gamma)' \tau_\gamma^{-1} (\gamma - \mu_\gamma) \right), \tag{20}$$

where  $\mu_\gamma$  and  $\tau_\gamma$  are the parameters of the independent normal prior of  $\gamma$ .

The full conditional distribution of shape parameter,  $\alpha$  in the survival sub-model is given by

$$p(\alpha|others) \propto \prod_{i=1}^n (\alpha \exp(\gamma t + \xi v_i + \psi \mu_i^*(t)))^{\delta_i} \times \exp(-\Lambda_i(t)) \times \alpha^{a_0-1} \exp(-\alpha b_0), \tag{21}$$

where  $a_0$  and  $b_0$  are the parameters of the independent gamma prior of  $\alpha$ .

The full conditional distribution of the association between the two processes  $\psi$  is

$$p(\psi|others) \propto \prod_{i=1}^n (\alpha \exp(\gamma t + \xi v_i + \psi \mu_i^*(t)))^{\delta_i} \times \exp(-\Lambda_i(t)) \times \exp \left( (-1/2)(\psi - \mu_\psi)' \tau_\psi^{-1} (\psi - \mu_\psi) \right) \tag{22},$$

where  $\mu_\psi$  and  $\tau_\psi$  are the specified parameters of the independent normal prior of  $\psi$ .

The full conditional distribution of inverse variance,  $\frac{1}{\sigma_\epsilon^2}$  takes the form

$$p\left(\frac{1}{\sigma_\epsilon^2}|others\right) \propto \prod_{i=1}^n (2\pi\sigma_\epsilon^2)^{-n_i/2} \exp \left( (-1/2\sigma_\epsilon^2) \sum_{j=1}^{n_i} [\mathbf{y}_{ij} - \mu_i^*(t_{ij})]^2 \right) \times (\sigma_\epsilon^2)^{-a_1-1} \exp(-b_1/\sigma_\epsilon^2), \tag{23}$$

where  $a_0$  and  $b_0$  are the parameters of the inverse gamma of  $\sigma_\epsilon^2$ .

The full conditional distribution of the inverse covariance variance,  $A^{-1}$  takes the form

$$p(A^{-1}|others) \propto \prod_{i=1}^n |A|^{-n_i/2} \exp\left(-\frac{1}{2}(\mathbf{w}'_i A^{-1} \mathbf{w}_i)\right) \times |A|^{-(\nu_0+q+1)/2} \exp\left(-\frac{1}{2}tr(A_0 A^{-1})\right), \tag{24}$$

where  $\nu_0$  and  $A_0$  are the parameters of the inverse wishart prior of  $A$ .

More simplified full conditional distribution can be derived the way as in Faucett & Thomas (1996) and Zhang *et al.* (2017).

### 3.3 Model Selection

In this paper, we propose two forms of the random effects terms in the joint model:

- a) Correlated random intercepts and slope - Case I
- b) Uncorrelated random intercepts and slopes - Case II

For the model selection, we consider the Deviance information criteria(DIC)(Spiegelhalter *et al.*, 2002) which defined as a version of well known Akaike’s information criteria(AIC) and Bayesian criteria(BIC). The idea of using of DIC is that, DIC uses the posterior distribution which allows it to take into account the prior information of both longitudinal and time to event sub-models. Consider  $\phi$ , the collection of parameters in the joint model and  $\mathcal{D}_{obs} = \{y_{ij}, T_i^*, \delta_i, \mathbf{x}_i, \nu_i\}$ , the observed data. Now, let specify the deviance function as

$$D(\phi) = -2 \sum_{i=1}^n \log f(\mathcal{D}_{obs}|\phi) = -2 \log(\phi|\mathbf{w}_i, \mathcal{D}_{obs}), \tag{25}$$

where  $\log(\phi|\mathbf{w}_i, \mathcal{D}_{obs})$  is the likelihood in equation(11). Define  $\overline{D(\phi)} = E\left[-2 \sum_{i=1}^n \log f(y_{ij}, T_i^*, \delta_i, \mathbf{x}_i, \nu_i; \phi)\right]$  as the expectation of the deviance under posterior and  $\overline{\phi} = E[\phi]$  posterior means of the parameters. According to Spiegelhalter *et al.*, 2002, the difference between two measures denoted by  $p_D = \overline{D(\phi)} - D(\overline{\phi})$  can be interpreted as the posterior estimate of the effective number of parameters and it measures the complexity of the model. Hence adding it to the posterior mean Deviance, gives a measure of fit that penalized for complexity . Therefore,

$$DIC = D(\overline{\phi}) + 2p_D = \overline{D(\phi)} + p_D. \tag{26}$$

Based 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 dataset. However, as stated by Geedipally *et al.*, 2014, the model is penalized by the  $\overline{D(\phi)}$  which will decrease as the number of the parameters in the model increases and  $p_D$ , which compensates for this effect by favoring model with a smaller number of parameters. Therefore, it is very important to note that the way the model is parameterized will influence the outcomes of the Deviance Information Criterion (DIC) values.

## 4. Simulation Studies

In this section, we performed two sets of simulations studies under Markov Chain Monte Carlo (MCMC) in order to assess the performance of the proposed methodology.

### 4.1 Simulation Study 1

The data are generated from a joint model with a single longitudinal variable and a time to event variable. Each subject is expected to have a record of  $n_i = 5$  biomarker values recorded at baseline and thereafter 4 visits are scheduled at equally spaced time interval  $t_{ij} = \{0.00, 0.15, 0.30, 0.45, 0.60\}$ . After this period, subjects are said to be censored non-informatively. Specifically, we first generate data using a Linear Mixed Effects model for the longitudinal sub-model;

$$y_i(t_{ij}) = \mu_i^*(t_{ij}) + \mathbf{z}'_i(t_{ij})\mathbf{w}_i + \epsilon_{ij}, \tag{27}$$

where  $\mu_i^*(t_{ij})$  is the true unobserved biomarker value and  $\epsilon_i(t_{ij}) \sim N(0, \sigma_\epsilon^2)$ , where  $\sigma_\epsilon^2 = 0.5$  is the measurement error variance. We then simulate longitudinal data from a linear curve,  $\mu_i^*(t_{ij}) = \beta_0 + \beta_1 x_i + \beta_2 t_{ij} + w_{i0} + w_{i1} t_{ij}$ , where the vector of random effects  $\mathbf{w}_i = (w_{i0}, w_{i1})$  is simulated from a multivariate normal distribution  $\mathcal{MVN}(0, \mathbf{A})$ , with variance covariance matrix  $\mathbf{A} = \begin{pmatrix} \sigma_{w_{i0}}^2 & \rho\sigma_{w_{i0}}\sigma_{w_{i1}} \\ \rho\sigma_{w_{i0}}\sigma_{w_{i1}} & \sigma_{w_{i1}}^2 \end{pmatrix} = \begin{pmatrix} 0.75^2 & -0.02 \\ -0.02 & 0.15^2 \end{pmatrix}$ . Note that  $\rho = -0.2$ . We set coefficients  $\beta = (\beta_0, \beta_1, \beta_2)' = (3.00, 0.50, 5.00)'$ .

For the survival sub-model, we have

$$\lambda_i(t|v_i, \mu_i^*(t)) = \lambda_0(t) \exp(\xi v_i + \psi \mu_i^*(t)) \tag{28}$$

where  $\lambda_0(t) = \alpha \exp(\gamma t)$ , with  $\alpha = 0.02$  and  $\gamma = 1.2$ ,  $v_i$  is a vector of baseline covariates with associated coefficients  $\xi = (\xi_0, \xi_1)'$ , with  $\xi_0 = -0.74$  and  $\xi_1 = -0.015$ . The values of baseline covariates were simulated as follows:

$$x_i \sim N(12, 4) \text{ and } v_i \sim Bin(1, 0.5)$$

The parameter  $\psi$  was set at 0.2. Since our interest is in the time to event, we generate the vector of true event times  $T_i^*$  by first simulate the survival probability,  $U_i$ , from *Uniform*(0, 1) for each subject and solve for  $T_i^*$  from the following equation:

$$U_i = \exp \left\{ - \int_0^{T_i^*} \alpha \exp(\gamma s) \exp(\xi v_i + \psi \mu_i^*(s)) \partial s \right\} \tag{29}$$

To obtain the cumulative function, we let

$$\begin{aligned} \Lambda_i(t|v_i, \mu_i^*(t)) &= \int_0^{T_i^*} \alpha \exp(\gamma s) \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + \beta_2 s + w_{i0} + w_{i1} s)) \partial s \\ &= \int_0^{T_i^*} \alpha \exp(\gamma s + \xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + \beta_2 s + w_{i0} + w_{i1} s)) \partial s \\ &= \alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0})) \times \int_0^{T_i^*} \exp((\gamma + \psi \beta_2 + \psi w_{i1}) s) \partial s \\ &= \alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0})) \times \left[ \frac{\exp((\gamma + \psi \beta_2 + \psi w_{i1}) s)}{\gamma + \psi \beta_2 + \psi w_{i1}} \right]_0^{T_i^*} \\ &= \alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0})) \times \left[ \frac{\exp((\gamma + \psi \beta_2 + \psi w_{i1}) T_i^*)}{\gamma + \psi \beta_2 + \psi w_{i1}} - \frac{1}{\gamma + \psi \beta_2 + \psi w_{i1}} \right] \\ &= \frac{\alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0}))}{\gamma + \psi \beta_2 + \psi w_{i1}} \times [\exp((\gamma + \psi \beta_2 + \psi w_{i1}) T_i^*) - 1] \end{aligned} \tag{30}$$

Hence, we have

$$U_i = \exp \left( - \frac{\alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0}))}{\gamma + \psi \beta_2 + \psi w_{i1}} \times [\exp((\gamma + \psi \beta_2 + \psi w_{i1}) T_i^*) - 1] \right) \tag{31}$$

Now, solving for  $T_i^*$ , from equation (29), we have

$$\begin{aligned} \log(U_i) &= - \frac{\alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0}))}{\gamma + \psi \beta_2 + \psi w_{i1}} \times [\exp((\gamma + \psi \beta_2 + \psi w_{i1}) T_i^*) - 1] \\ T_i^* &= \frac{1}{\gamma + \psi \beta_2 + \psi w_{i1}} \log \left( - \frac{\gamma + \psi \beta_2 + \psi w_{i1}}{\alpha \exp(\xi_0 + \xi_1 v_i + \psi(\beta_0 + \beta_1 x_i + w_{i0}))} \log(U_i) + 1 \right) \end{aligned} \tag{32}$$

We draw censoring time  $C_i$  from an uniform distribution *Uniform*(0.05, 8). Then, we compute  $T_i = \min(T_i^*, C_i)$  and event indicator  $\delta_i = 1$  if  $T_i^* \leq C_i$  and 0 otherwise. The censoring was recored between 40% – 45%.

#### 4.2 Simulation Study 2

In this simulation study, subjects are expected to have 10 repeated measurements taken at each visit between  $0 \leq t \leq 12.5$ . We generated longitudinal data from the following model:

$$y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 t_{ij} + \beta_3 * x_{i2} + \beta_4 x_{i3} + w_{i0} + w_{i1} t_{ij} + \epsilon_{ij} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n_i \tag{33}$$

where  $0 < t_{ij} \leq 10$ ,  $\beta = \{5.00, 2.00, 1.00, -2.50\}$ ,  $x_{ik}, k = 1, 2, 3$  were generated from a Bernoulli distribution with different success probabilities and  $\epsilon_i(t_{ij}) \sim N(0, \sigma_\epsilon^2)$ , with  $\sigma_\epsilon^2 = 0.2$ . For the vector of random effects  $w_i = (w_{i0}, w_{i1})$  with no correlation between random intercept and slope was generated from a multivariate normal distribution  $MVN(0, A^2)$ , where  $A = \text{diag}(\sigma_{w_{i0}}, \sigma_{w_{i1}})^2$ , with  $\sigma_{w_{i0}} = 0.75$  and  $\sigma_{w_{i1}} = 0.15$ .

Also, a proportional hazards model with Gompertz baseline was considered:

$$\lambda_i(t) = \alpha \exp(\gamma t) \exp(\xi_0 + \xi_1 v_{i1} + \xi_2 v_{i2} + \psi(\beta_0 + \beta_1 x_{i1} + \beta_2 t + \beta_3 * x_{i2} + \beta_4 x_{i3} + w_{i0} + w_{i1} t)) \tag{34}$$

where  $\gamma = 1.5$  and  $\xi = \{-4.00, -1.00, -2.10\}$ ,  $\alpha = 0.15$ ,  $\psi = 0.02$ , the covariate  $v_{i1}$  and  $v_{i2}$  were generated from normal distribution (i.e.,  $\mathcal{N}(25, 6)$ ) and uniform distribution (i.e.,  $Uniform(0, 3)$ ), respectively. The true event times was then generate from

$$U_i = \exp\left(-\int_0^{T_i^*} \alpha \exp(\gamma s) \exp(\xi_0 + \xi_1 v_{i1} + \xi_2 v_{i2} + \psi(\beta_0 + \beta_1 x_{i1} + \beta_2 s + \beta_3 * x_{i2} + \beta_4 x_{i3} + w_{i0} + w_{i1} s)) \partial s\right) \tag{35}$$

$$T_i^* = \frac{1}{\gamma + \psi(\beta_2 + w_{i1})} \log\left(-\frac{(\gamma + \psi(\beta_2 + w_{i1}))}{\alpha \exp(\xi_0 + \xi_1 v_{i1} + \xi_2 v_{i2} + \psi(\beta_0 + \beta_1 x_{i1} + \beta_3 x_{i2} + \beta_4 x_{i3} + w_{i0}))} \times \log(U_i) + 1\right) \tag{36}$$

We generate censoring times,  $C_i$  from an exponential distribution with mean 10.

### 4.3 Simulation Study Results

The aim of the simulation study was to investigate the performance of the joint model. The results are summarized in Table 1 and 2. We have considered several quantities to determined the behavior of the estimators  $\hat{\phi}$  by comparing to the true  $\phi$  as follows;

- 1) The estimated bias:  $\mathcal{B}(\phi) = \frac{1}{N} \sum \hat{\phi}_i - \phi = \bar{\phi} - \phi$ , where a negative bias indicates an underestimation while a positive bias indicates an overestimation.
- 2) The Root Mean Square Error:  $RMES(\hat{\phi}) = \sqrt{\frac{1}{N} \sum_i (\hat{\phi}_i - \phi)^2}$ , this measures the accuracy of the estimates. The lower the RMSE, the more accurate effects estimates.
- 3) The standard error:  $S.E(\hat{\phi}) = \frac{\sigma}{\sqrt{n}}$ , defined as equal to standard deviation divided by the square root of the sample size. This implies that the larger the sample size, the smaller the standard error.
- 4) The 95% coverage probability(CP):  $\hat{\phi} \pm 1.96 \times se(\hat{\phi})$ , the proportion of 1000 simulated data sets for which 95% confidence intervals included the true estimates. The closer the outcomes to 95%CP(0.95), the more accurate the estimates.

Table 1. Simulation study 1 results with correlation random effects from 1000 replications of 200 and 500 subjects.

Parameter	n = 200						n = 500				
	True( $\phi$ )	Mean( $\hat{\phi}$ )	Bias	S.E	RMSE	95%CP	Mean( $\hat{\phi}$ )	Bias	S.E	RMSE	95%CP
$\beta_0$	3.000	2.996	-0.004	0.020	0.501	0.920	3.027	0.027	0.010	0.324	0.935
$\beta_1$	0.500	0.500	0.000	0.000	0.002	0.895	0.500	0.000	0.000	0.001	0.920
$\beta_2$	5.000	4.999	-0.001	0.004	0.064	0.930	5.008	0.008	0.003	0.034	0.935
$\sigma_{w_{i0}}$	0.750	0.737	-0.013	0.004	0.043	0.945	0.802	0.052	0.003	0.015	0.960
$\sigma_{w_{i1}}$	0.150	0.342	-0.058	0.009	0.125	0.900	0.164	0.064	0.005	0.065	0.954
$\rho$	-0.200	-0.256	0.056	0.014	0.045	0.935	-0.120	-0.080	0.008	0.013	0.945
$\sigma_\epsilon$	0.500	0.496	-0.040	0.005	0.012	0.915	0.496	-0.04	0.003	0.008	0.930
$\xi_1$	-0.740	-0.619	-0.019	0.038	0.432	0.895	-0.619	-0.019	0.024	0.352	0.900
$\xi_2$	-0.015	-1.508	-0.008	0.626	0.054	0.945	-1.508	-0.008	0.626	0.026	0.945
$\psi$	0.200	0.009	-0.011	0.008	0.324	0.964	0.009	-0.011	0.003	0.131	0.972
$\alpha$	0.020	0.018	-0.002	0.012	0.075	0.940	0.025	0.005	0.008	0.043	0.945
$\gamma$	1.200	1.268	0.068	0.035	0.642	0.910	1.351	0.151	0.012	0.136	0.915

Table 2. Simulation study 2 results with uncorrelated random effects



Parameter	$n = 200$						$n = 500$				
	True( $\phi$ )	Mean( $\hat{\phi}$ )	Bias	S.E	RMSE	95%CP	Mean( $\hat{\phi}$ )	Bias	S.E	RMSE	95%CP
$\beta_0$	5.000	5.010	0.010	0.008	0.319	0.898	5.015	0.015	0.006	0.124	0.904
$\beta_1$	2.000	2.003	0.003	0.006	0.011	0.964	1.998	-0.002	0.004	0.044	0.995
$\beta_2$	1.000	1.001	0.001	0.002	0.023	0.945	1.003	0.003	0.001	0.052	0.950
$\beta_3$	2.000	2.002	0.002	0.042	0.058	0.974	2.004	0.004	0.016	0.015	0.978
$\beta_4$	1.500	1.505	0.005	0.011	0.128	0.892	1.499	-0.001	0.029	0.044	0.965
$\sigma_{i0}$	0.750	0.689	-0.061	0.005	0.246	0.975	0.739	-0.011	0.002	0.043	0.985
$\sigma_{i1}$	0.150	0.116	-0.034	0.029	0.054	0.945	0.140	-0.010	0.020	0.039	0.974
$\sigma_\epsilon$	0.200	0.201	0.001	0.001	0.115	0.994	0.201	0.001	0.000	0.023	0.999
$\xi_0$	-4.000	-3.985	-0.015	0.043	0.065	0.964	-4.005	0.005	0.021	0.048	0.960
$\xi_1$	-1.000	-0.943	-0.057	0.052	0.143	0.926	-1.021	0.021	0.015	0.042	0.972
$\xi_2$	2.100	2.085	-0.015	0.034	0.484	0.944	2.241	0.141	0.029	0.053	0.970
$\psi$	0.020	0.025	0.005	0.024	0.431	0.945	0.023	0.003	0.014	0.125	0.960
$\alpha$	0.150	0.143	-0.007	0.043	0.563	0.965	0.158	0.008	0.023	0.254	0.955
$\gamma$	1.500	1.546	0.046	0.034	0.654	0.991	1.495	-0.005	0.017	0.372	0.993

The summary statistics of the estimated regression coefficients are summarized in table 1 and 2, respectively for  $n = 200$  and  $n = 500$  subjects. In each simulation, 1000 replications were performed. We can clearly see that the proposed methodology performed well in terms of parameter estimations. The biases are relatively small and the probability of 95% credible intervals dwells around the 0.95 values. On the other hand, simulation with large sample size have smaller standard error, hence the root mean square error. Overall, better results were obtained with joint model with correlated random effects. That is, the correlation between random intercept and slope positively influences the estimates.

### 5. Real Data Application

#### 5.1 Data

In this section, we apply the joint model to the 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 Federation Francophone de Cancerologie Digestive(FFCD). The main aim of the study was to examine the efficacy of two treatment effects: Sequential arm(S) and combination arm(C). We consider datasets presented by Kro'1 *et al.* (2016 & 2017), in which 150 patients were randomly selected from the same clinical trial. The data contains individual progression of disease such as tumor size, time of new lesions(recurrent events), baseline covariates(Age, WHO performance status and previous resection: combination arm vs sequential arm), time to death or the last observed time for right censored. A total of 906 tumor size measurements were recored at subject specific follow-up time. During the study, 289 recurrences and 121 deaths were also recorded. We chose to model the longitudinal outcomes together with time to death without recurrence. In our model, we included a total of 716 tumor size measurements of 41 deaths and 109 right censored. The aim of this application is to examine the effects of longitudinal dynamics and baseline covariates on subjects who died without experiencing any recurrence.

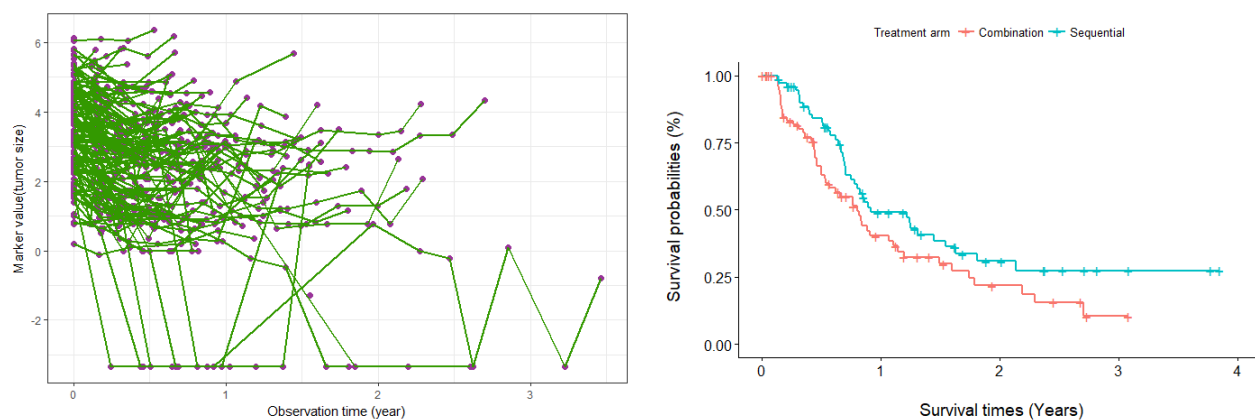


Figure 1. Individual profiles for Longitudinal measurements(right) and the Kaplan-Meier estimates of the survival function among patients with no recurrence(left)

#### 5.1.1 Longitudinal Sub-Model

For the longitudinal process, we includes the following covariates:

Treatment(1: Sequential, 2: Combination), Age(1: < 60 years, 2: 60 – 69 years, 3: > 69 years), who.PS(1: status 0. 2: status 1 and 3: status 2).

$$Y_{ij} = \beta_0 + \beta_1 Age_i + \beta_2 who.PS_i + \beta_3 Year * Treatment_i + w_{i0} + w_{i1} * Year + \epsilon_{ij}, \tag{37}$$

where  $y_{ij}$  represents tumor size, which is the transformed sum of the longest diameters ( $SLD^* = (SLD^{0.3} - 1)/0.3$ ) measured during the visit,  $w_{i0}$  and  $w_{i1}$  are the subject specific random effects for intercept and slope respectively.

5.1.2 Survival Submodel

The following proportional hazards model with Gompertz baseline function and Prev-resection(0: No, 1: Yes) covariate was used

$$\lambda_i(t) = \alpha \exp(\gamma time1_i + \xi_0 + \xi_1 Prev.resection + \psi(\beta_0 + \beta_1 Age_i + \beta_2 who.PS_i + \beta_3 time1_i * Treatment_i + w_{i0} + w_{i1} * time1_i)) \tag{38}$$

In the Gibbs sampler algorithm, we ran three MCMC chains with random generated initial values for 205000 iterations. In order to avoid influences of pre-convergence on our final posterior inferences, 55000 iterations were discarded as burn-in. Therefore, our posterior inferences were based on the last 150000 iterations. As presented in subsection 3.2.2, we consider informative priors for models’ coefficients, that is  $\beta_L, \xi$  and  $\gamma$  are assumed to have normal distribution with mean zero and a large variance ( $\beta_L \sim \mathcal{N}(0, 100), \xi \sim \mathcal{N}(0, 100), \gamma \sim \mathcal{N}(0, 100)$ ). We have also assume an independent gamma prior for  $\alpha, \alpha \sim \mathcal{G}(0.01, 0.01)$ . For the random intercept  $\sigma_{w_{i0}}$  and slope  $\sigma_{w_{i1}}$ , we use inverse-Wishart prior distribution  $A \sim Inv - Wishart(A_0, q)$ , where  $A_0$  is the prior distribution for covariance matrix with  $q = 2$  degrees of freedom.

Table 3. Posterior mean, standard error and 95% credible interval for the joint model

Parameters	Case I			Case II		
	Estimate	SE	95%CI	Estimate	SE	95%CI
<b>Longitudinal submodel</b>						
<i>Fixed Effect</i>						
Intercept( $\beta_0$ )	3.177	0.009	(2.461,3.908)	3.232	0.012	(2.520,3.925)
Age( $\beta_1$ )	-0.095	0.004	(-0.410,0.221)	-0.019	0.006	(-0.417,0.207)
Who.PS( $\beta_2$ )	0.308	0.004	(0.035,0.636)	0.295	0.005	(-0.010,0.597)
Year $\times$ Treatment( $\beta_3$ )	-0.514	0.003	(-0.766,-0.269)	-0.575	0.004	(-0.800,-0.338)
<i>Random Effect</i>						
$\rho$	0.052	0.001	(-0.157,0.258)	-	-	-
$\sigma_{w_{i0}}^2$	6.457	0.007	3.828,10.400)	6.848	0.010	(4.052,11.040)
$\sigma_{w_{i0}}\sigma_{w_{i1}}$	7.901	0.009	(5.040,11.110)	-	-	-
$\sigma_{w_{i1}}^2$	10.170	0.020	(0.653,1.976)	10.870	0.033	(5.202,20.490)
$\sigma_{\epsilon}^2$	0.822	0.000	(0.773,0.876)	0.821	0.000	(0.771,0.874)
<b>Time to death without recurrence</b>						
Intercept( $\xi_0$ )	1.149	0.074	(-2.595,6.704)	0.643	0.064	(-2.660,3.854)
Prev.resection( $\xi_1$ )	-0.488	0.008	(-1.115,0.087)	-0.458	0.011	(-1.105,0.148)
Time( $\gamma$ )	0.188	0.007	(-0.350,0.675)	0.092	0.012	(-0.560,0.665)
$\alpha$	0.298	0.014	(0.001,1.850)	0.3091	0.023	(0.005,2.108)
$\psi$	0.264	0.004	(0.003,0.560)	0.244	0.005	(-0.019,0.504)
<b>Model selection</b>						
$\bar{D}$	1926.0	0.204	(1878.0,1978.0)	1924.0	0.209.0	(1877.0, 1975.0)
$p_D$	79.1			80.6		
DIC	2001.0			2005.0		

In table 3, we present the posterior estimates, standard error and 95% credible interval of the estimates based on the proposed joint models. Although, there is little difference between parameter values in both cases, we can clearly see that all the parameters are statistically significant by looking at the value of standard error. For the longitudinal outcome, we see that there is a decrease in tumor size measurements contributed by age and time varying treatment. We also observed a positive effect of previous resection on the risk of death. On the other hand, the value of association parameter,  $\psi$  in both cases indicate that there is a positive association between the two processes. Based on Deviance information criteria, a joint model(Case I) with correlated random intercepts and slopes seems to be the model that would best fit the data as it has relatively small DIC value compared to case II with uncorrelated random intercepts and slopes.

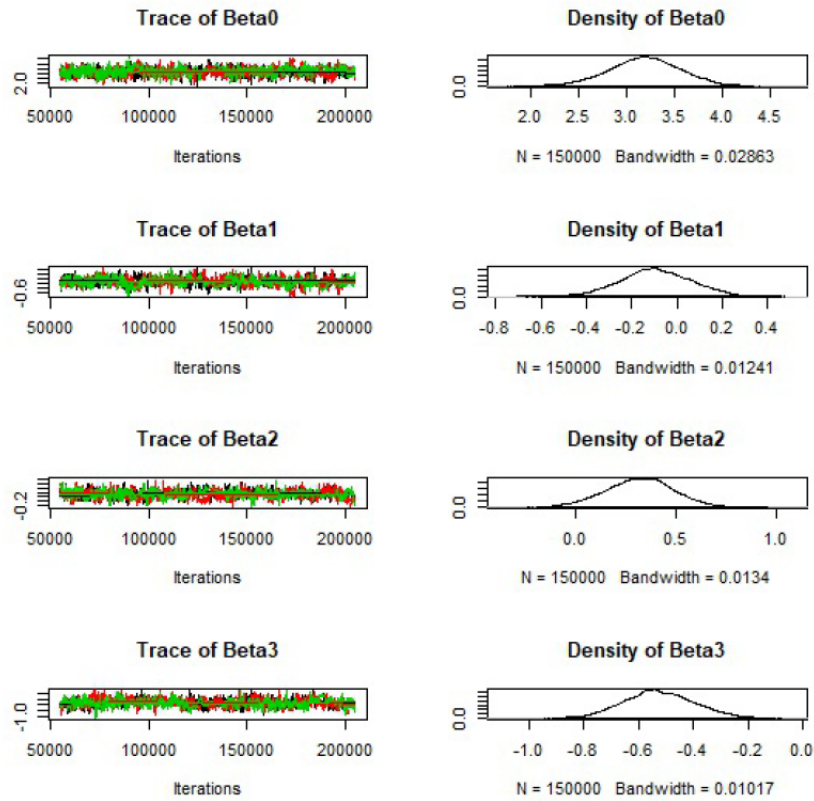


Figure 2. Posterior marginal distribution trace and density plots for  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$ .

Figure 2 show the the trace and density plots for posterior marginal distributions of selected parameters. We clearly see that, the MCMC of all parameters have converged to their target posterior distributions.

## 6. Discussion

Joint modeling for longitudinal outcome and time to event data has gained increasing popularity in literature. However, when it comes to the choice of baseline function on survival sub-part, many authors assumed that the survival times follow exponential or weibull distributions. In this paper, we developed a joint model under a Bayesian prospective assuming that the proportional hazards model for the survival times has a Gompertz baseline hazard function. We think that generating survival times from a Gompertz distribution have more advantages over Weibull distribution as its cumulative distribution has a closed form solution which make it more easier to simulate survival data in presence of time varying covariates. We started building separate models for each process and then link them together through a common latent variable. Our model incorporate both time invariant(fixed) and time varying covariates that forces the hazards of the outcomes to change over time. On the other hand, the inter-relationship between markers was accounted by subject-specific random effects.

Due to high programming needs for fitting the joint likelihood function, we proposed a Bayesian approach that estimates the parameters by simulating samples from posterior distribution. Specifically, Gibbs sampler was used for posterior inferences as it provides conveniently way to the fit of complex models. We have conducted an extensive examination of the model parameter estimation through simulation studies (simulation study 1- correlated random effects, and simulation study 2- uncorrelated random effects). The simulation results for both simulation studies are presented by looking at several quantities such as Bias-the difference between the average estimate over all simulations and the true parameter value, S.E- standard error of the estimates that measures the accuracy of predictions, RMSE-the square root of the mean error and CP- the coverage probability. The results from the two simulation studies, indicated adequate performance of the joint model. They highlighted however some weakness of the model when few sample sizes were used.

There is ample of additional work needed in joint modeling framework. This paper covered only a small area of this fast growing field of research. Therefore, this work can be extended further to accommodate multiple longitudinal outcomes and competing risks as done by Musoro (2014). In future, we plan to develop a joint model for longitudinal and time to event data assuming that survival times follow a Generalized Gompertz distribution with three parameters (Haile et al.. 2016).

In short, we have introduced a more flexible joint model for longitudinal and time to event data assuming that the survival time follow a Gompertz distribution. We further demonstrated that this model can easily be used in practice through a study, the FFCD 2000-2005 multi-center phase III clinical trial of patients diagnosed with metastatic colorectal cancer. In both simulation and application studies, two cases (case I with correlated random intercepts and slope and case II with uncorrelated random intercepts and slopes) were considered. It is clear that our work contributed to this fascinating research area by making use of more flexible methodology to develop a joint model.

## 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., & Nash, S. W. (1967). The generalized gompertz-verhulst family of distributions. *Sankhya: The Indian Journal of Statistics, Series A*, 141-156. Retrieved from <https://www.jstor.org/stable/25049460>
- Austin, P. C. (2012). Generating survival times to simulate cox proportional hazards models with time-varying covariates.
- Austin, P. C. (2013). Correction: generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in Medicine*, 32(6), 1078-1078. <https://doi.org/10.1002/sim.5723>
- Casellas, J. (2007). Bayesian inference in a piecewise weibull proportional hazards model with unknown change points.
- Chi, Y. Y., & Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2), 432-445. <https://www.jstor.org/stable/3695862>
- Faucett, C. L., & Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured co- variates: a gibbs sampling approach. *Statistics in medicine*, 15(15), 1663-1685. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960815\)15:15<1663::AID-SIM294>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1)
- 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.
- Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, 115, 513-583. Retrieved from <http://www.jstor.org/stable/107756>
- Haile, S., Jeong, J. H., Chen, X., & Cheng, Y. (2016). A 3-parameter gompertz distribution for survival data with competing risks, with an application to breast cancer data. *Journal of Applied Statistics*, 43(12), 2239-2253. <https://doi.org/10.1080/02664763.2015.1134450>
- Hougaard, P. (2012). Analysis of multivariate survival data. *Springer Science & Business Media*. Retrieved from <https://books.google.co.ke/books?id=VJfzBwAAQBAJ>
- 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), 185-195.
- Huang, Y., Dagne, G., and Wu, L. (2011). Bayesian inference on joint models of hiv dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine*, 30(24), 2930-2946. <https://doi.org/10.1002/sim.4321>
- 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), 2796-2801. <http://doi.org/10.1200/JCO.2009.25.0654>
- Ibrahim, J., Chen, M., & Sinha, D. (2004). BAYESIAN METHODS FOR JOINT MODELING OF LONGITUDINAL AND SURVIVAL DATA WITH APPLICATIONS TO CANCER VACCINE TRIALS. *Statistica Sinica*, 14(3), 863- 883. Retrieved from <http://www.jstor.org/stable/24307419>
- Journal of Animal Breeding and Genetics*, 124(4), 176-184. <https://doi.org/10.1111/j.1439-0388.2007.00658.x>
- Kro l, A., Ferrer, L., Pignon, J. P., Proust-Lima, C., Ducreux, M., Bouch e, O., Michiels, S., & Rondeau, V. (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), 907-916. <https://doi.org/10.1111/biom.12490>
- Kro l, A., Mauguén, 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://www.jstatsoft.org/v081/i03>
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963-974.
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13- 22. <https://doi.org/10.1093/biomet/73.1.13>
- Martins, R., Silva, G. L., and Andreozzi, V. (2017). Joint analysis of longitudinal and survival aids data with a spatial fraction of long-term survivors: A bayesian approach. *Biometrical Journal*, 59(6), 1166-1183. <https://doi.org/10.1002/bimj.201600159>
- 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>
- Musoro, J. Z., Geskus, R. B., & Zwinderman, A. H. (2015). A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant. *Biometrical Journal*, 57(2), 185-200. <https://doi.org/10.1002/bimj.201300167>
- Ntzoufras, I. (2011). Bayesian Modeling Using WinBUGS. Wiley Series in Computational Statistics. Wiley. <https://books.google.co.ke/books?id=1yZnwnOxXsEC>
- Rizopoulos, D. (2012). Joint models for longitudinal and time-to-event data: With applications in R. CRC Press.
- 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>
- Song X, Davidian M, Tsiatis A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, 4(3), 511-528. <https://doi.org/10.1093/biostatistics/3.4.511>

- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.
- Sweeting, M. J., & Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5), 750-763. <https://doi.org/10.1002/bimj.201100052>
- Tseng, Y. K., Hsieh, F., & Wang, J. L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92(3), 587-603. <https://doi.org/10.1093/biomet/92.3.587>
- Tsiatis, A., & Davidian, M. (2004). Joint Modeling of Longitudinal and Time-To-Event Data: An Overview. *Statistica Sinica*, 14(3), 809-834. Retrieved from <http://www.jstor.org/stable/24307417>
- Verbeke, G. (1997). Linear mixed models for longitudinal data. *Springer*, page 63-153. <https://doi.org/10.1007/978-1-4419-0300-6>
- Yang, L., Yu, M., and Gao, S. (2016). Joint models for multiple longitudinal processes and timeto- event outcome. *Journal of statistical computation and simulation*, 86(18), 3682-3700. <https://doi.org/10.1080/00949655.2016.1181760>
- 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. <https://doi.org/10.1198/016214507000000400>
- Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 1049-1060. <https://www.jstor.org/stable/2531734>
- 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>

### 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/3.0/>).