Computational Journal of Mathematical and Statistical Sciences
4(1), 72-95
DOI:10.21608/cjmss.2024.303748.1061
https://cjmss.journals.ekb.eg/

*Research article*

# Latent Gaussian Approach to Joint Modelling of Longitudinal and Mixture Cure Outcomes

**Ekong A. H.**[1*], **Olayiwola O. M.**[1], **Dawodu G. A.** [1], **and Osinuga I. A.**[2]

[1] Department of Statistics, Federal University of Agriculture Abeokuta, P.M.B 2240, Abeokuta, Ogun State, Nigeria

[2] Department of Mathematics, Federal University of Agriculture Abeokuta, P.M.B 2240, Abeokuta, Ogun State, Nigeria

* **Correspondence:** anieekong@outlook.com

**Abstract:** Joint modelling has become pervasive in analysing data from survival and longitudinal studies. There are several technqiues on joint analyses of datasets from both studies simultaneously. Our interest here is on approximate Bayesian inference using latent Gaussian models (LGMs) to analyse longitudinal and mixture cure outcomes with shared random effect. Longitudinal outcome was modelled using spline function to account for nonlinearity in longitudinal trajectories often seen in real life datasets, while survival outcome was modelled using Cox proportional hazards and logistic link function for latency and incidence components respectively, to account for the possibility of cure proportion. The LGMs require three levels of hierachy involving joint likelihoods of paramters and hyperparameters, defining a multivariate normal distribution for a latent field and finally priors for the hyperparamters. Posterior estimates were evaluated using Integrated Laplace approximation. Simulation study compared linear, quadratic and spline specifications for longitudinal trajectories and showed similar results for quadratic and spline function in small sample sizes and linear specification good only in large sample size. The approach was applied to renal transplantation data which comprised glomerular filtration rate of kidneys and survival event of time to graft failure. The contribution of this paper is that it adds to the literature on approximate Bayesian alternative to jointly modelling nonlinear trajectories of longitudinal outcomes and survival outcome with possibility of cure proportions and showed its computational merit over sample-based Bayesian approach.

**Keywords:** Survival analysis, joint modelling, latent Gaussian models, shared random effect, nonlinear trajectory
**Mathematics Subject Classification:** 97K60; 91G70
**Received:** 25 July 2024; **Revised:** 27 August 2024; **Accepted:** 10 September 2024; **Online:** 12 October 2024.

## 1. Introduction

Joint modelling has become a pervasive approach in analysing clinical data involving simultaneous observation of time to event outcome and longitudinal outcomes. This has also been applied in other applied areas such as for example finance (see Hu and Zhou [8]; Medina-Olivares et al. [12]). There are load of works in the literature on joint modelling with different types of models for the longitudinal submodels but mostly the mixed effect model is used. The survival submodels often have been applied to time to event data, and extended to recurrent events, competing risk and cure data.

Joint modelling of these types of data has been noted to have some merits over separate modelling of the longitudinal and survival data (See Tsiatis and Davidian [21]). Hickey et al. [6] gave a comprehensive review of literatures for implementation of joint models involving more than a single event time per subject. They considered the distributional and modelling assumptions, including the association structure, estimation approaches, software implementations, and clinical applications. Alsefri et al.[1] gave a review of developments in Bayesian joint models covering articles published up to July 2019. And from the literature, joint modelling with spatial components is rare, even with frequentist approach. Martins et al. [10], Martins et al. [11] and Rappl et al. [16] have been identified in the literature to have dealt with Bayesian joint modelling with spatial component, while Martins et al., [11] included a survival model with spatial long-term survivors, where Markov chain Monte Carlo (MCMC) was used for the posterior distribution evaluation.

Lázaro et al., [9] presented implementation of integrated Laplace approximation (INLA) in general mixture cure survival model with covariate information for the latency and the incidence model within a general scenario with censored and non-censored information. van Niekerk et al. [22] showed that a joint model with a linear bivariate Gaussian association structure is still a latent Gaussian model (LGM) and thus can be implemented using most existing packages for LGMs especially R-INLA and van Niekerk et al. [23] proposed a fully non-parametric spline component to competing risk joint model with nonlinear longitudinal trajectories to capture non-linear behaviour over time in the form of a random walk order two model.

There have been many literatures reporting the computational constraint of Markov chain Monte Carlo (MCMC) technique in joint modelling, and they have been shown to be limited to relatively small samples and model specifications, as well as have slow convergence properties (Rustand et al. [20]). The approximate Bayesian approach, INLA, introduced by Rue et al. [18] is slowly gaining usage for joint modelling as an alternative to MCMC, see van Niekerk et al. [23], Medina-Olivares et al., [13], Rustand et al. [19] and Rustand et al. [20].

The motivation for this study is firstly the fact that in real life situations, the trajectories of longitudinal biomarkers are not usually linear and hence linear mixed effects model will not adequately describe these trajectories. Secondly, in real life it is also noted that in many clinical trials there are patients who may not experience the survival event even after a long time after the end of the follow-up period. Hence, the possibility of cure proportion should be considered. Thirdly, with the pioneering work of Rue et al. [18], we can have an alternative to MCMC that have been shown to be efficient and time saving. Therefore, the contribution of this study is the modelling of longitudinal outcome with nonlinear trajectories and mixture cure survival event using LGM and implemented using INLA, for the case where the individual deviation is shared from the mean at time $t$ as defined by the random effects.

## 2. Methodology

Given sample observation $y_{im}$, on the $i$-th patient at the $m$-th time point, let $T_{im}$ be the event time for the $i$-th patient at the $m$-th time point, which may be right censored. The event indicator is given as $\delta_{im}$ = 1 if event is observed and $\delta_{im}$ = 0 if censored and $z_{im}$ then is the latent variable classifying the patient as cured or not at the end of the follow-up. We observe that any patient with survival time observation at a particular point in time is classified to the population of uncured patients.

The observed data for the $i$-th patient without any covariate is $\boldsymbol{D}_i$ = { $y_{im}$ , $T_{im}$ , $\delta_{im}$ , $z_{im}$ }. The $\boldsymbol{D}_i$ 's are assumed to be independent across patients, reflecting the belief that the disease process evolves independently for each patient. We also assume that $T_{im}$ and $\boldsymbol{y}_{im}$ are conditionally independent given some covariates of interest and a set of unobserved subject-level random effects.

### *2.1. Longitudinal model component*

Given longitudinal observation $y_{im}$ and assuming that for a marginal generalized linear model, the population is from some probability model with density $f(Y|X;\beta;U)$. We also assume that the longitudinal outcomes $y_{im}$, are conditionally independent and follow a well-defined distribution, $G$, with some density function $g$, linear predictor $\eta^L$ and hyperparameters $\theta^L$ , hence a structured additive model for the longitudinal component is given as follows:

$$g^{-1}\{E(y_{im} \mid X,\beta,U)\} = \eta^L = \beta_0 + \beta X + \sum_{i=1}^{n}\sum_{m=1}^{M} f_m(u_{im}) + b_{im}u_{im} + \epsilon \tag{2.1}$$

where $f_m(u_{im})$ is the $m$-th latent random effect of covariate $u_{im}$ for $i$-th patient, which could be spatial effects, temporal effects, patient or group-specific intercepts. $\beta$ represent the linear fixed effects of the covariates $X$, $b_{im}$ is the vector of random effects of intercept and slope, where $\beta_0$ plus $b_{im}$ gives the combined effect of the intercept and random intercepts terms specifying that the event depends on the patient-specific level of the longitudinal profile at time $t = 0$. We also have the structured random effect $f_m(u_{im})$ which we take as cubic splines with internal knots at 1 and 4 years to account for nonlinear trajectories of the longitudinal outcome, $\epsilon$ is the unstructured random effects.

### *2.2. Cure survival model component*

Given the event time $T_{im}$, let Z be a cure random variable defined as Z = 0 if that patient is susceptible for experiencing the event of interest, and Z = 1 if the patient is cured. Cure and uncured probabilities are P(Z = 1) = $\pi$ and P(Z = 0) = 1 - $\pi$, respectively. The survival functions for patients in the cured and uncured population, $S_c$(t) and $S_u$(t), t > 0, respectively, are

$$S_u(t) = P(T_{im} > t \mid Z = 0)$$

$$S_c(t) = P(T_{im} > t \mid Z = 1) = 1$$

The general survival function for $T_{im}$ can be expressed in terms of a mixture of both cured and uncured populations in the form

$$S(t) = P(T_{im} > t) = +(1-)S_u(t) \tag{2.2}$$

Cure fraction $\pi$ is also known as the incidence model and event time $T_{im}$ in the uncured population is also referred to as the latency model (Peng and Taylor, [15]).

*Covariates in the incidence model*

The effect of a baseline covariate vector $\boldsymbol{x}_1$ on the cure proportion is typically modelled by means of a logistic link function expressed as

$$logit\left[\ (\beta_1)\right] = \beta_1' x_1 \equiv (\beta_1) = \frac{exp\left\{\beta_1' x_1\right\}}{1 + exp\left\{\beta_1' x_1\right\}} \tag{2.3}$$

where $\beta_1$ is the vector of regression coefficients associated to $\boldsymbol{x}_1$ and $\pi$ is the cure proportion.

*Covariates in the latency*

The Cox proportional hazards model is usually formulated in terms of the hazard function for the event time as

$$h_u\left(t \mid h_{u0},\ \beta_2\right) = \lim_{\Delta t \to \infty} \frac{P\left(t \le T_{im} < t + \Delta t \mid T \ge t\right)}{\Delta t} = h_{u0}(t) exp\left\{\beta_2' x_2\right\} \tag{2.4}$$

where $h_{u0}(t)$ is the baseline hazard function that determines the shape of the hazard function. Model (2.4) can also be presented in terms of the survival function of $T_{im}$ as

$$S_u\left(t|S_{u0}, \beta_2\right) = \left[S_{u0}(t)\right]^{exp\left\{\beta_2' x_2\right\}} \tag{2.5}$$

where $S_{u0}(t) = exp\left\{-\int_0^t h_{u0}\left(s\right) ds\right\}$ represents the survival baseline function and some hyperparameter $\theta^S$.

### *2.3. Joint Model of Longitudinal and Cure Survival Outcomes as LGMs*

The modelling approach assumes a logistic distribution for the probability of cure in the incidence model in (2.3) and the Cox proportional hazard (2.4) for the survival time with a Weibull baseline hazard function $h_{u0}\left(t \mid \lambda,\ \alpha\right) = \lambda\alpha t^{\alpha-1}$ with $\lambda$ and $\alpha$ as the scale and shape parameters respectively. $\gamma$ is the association parameter estimating the strength of association between the survival and longitudinal components, thus we define

$$h_i\left(s\right) = h_{u0}\left(s\right) \eta_i^S\left(s\right) \left( exp\left\{ - \int_0^t h_i\left(u\right) du \right\} + logit\left[\ \right] \right)$$

The linear predictors of the joint model becomes

$$\eta_i^{L,J}(t) = \eta_i^L(t)$$

$$\eta_i^{S,J}\left(s\right) = \eta_i^S\left(s\right) + \gamma\left(\eta_i^L\left(s\right)\right)$$

where $\gamma$ is a smooth function facilitates the joint estimation of the models and can define as in our case, a nonlinear longitudinal trajectory and event process with cure proportion by using the entire longitudinal predictors as shared random effect.

*2.4. Likelihood function of Joint Model*

The likelihood of the longitudinal outcome is

$$\mathcal{L}^{L}\left(y \mid \eta^{L}\right)=\prod_{i=1}^{N_{L}} g\left(y_{i} \mid \eta_{i}^{L}(t)\right) \tag{2.6}$$

Given survival observations $\boldsymbol{d}=\{T_{im}\ ,\ \delta_{im}\ ,\ z_{im}\}$ and parameter vector $R=(\beta_1,\beta_2,\alpha,\lambda)$**,** the likelihood for the mixture cure survival becomes

$$\mathcal{L}^{S}\ (d \mid R)=\prod_{i=1}^{N} \mathcal{L}_{i}\ (R \mid d)=\prod_{i=1}^{N} \eta_{i}^{S} R^{z_i}\left(1-\eta_{i}^{S} R\right)^{1-z_i} h_{iu}(t_i \mid R)^{\delta_i(1-z_i)} S_{iu}(t_i \mid R)^{1-z_i} \tag{2.7}$$

Our task here will be to present equations (2.6) and (2.7) as LGMs by showing its specific hierarchical structure. The first level of the hierarchy involves the likelihood function given a latent field $=(\beta,\beta_1,\beta_2,\eta^{S},\eta^{L},f_m\ (\bullet)\ ,b_{im},\lambda,\ \ \epsilon)$ and a vector of hyperparameters $\theta=\left(\theta^{L},\theta^{S},\alpha,\ \tau^{-1},\gamma\right)$, hence the likelihood function is

$$p\ (D_i \mid \chi_i,\theta)=\prod_{i=1}^{n}\prod_{m=1}^{M} p\ (y_{im} \mid \chi_{im},\theta)$$

where the parameters are constant and $D$ and $\chi$ have the same dimension.

The next level of the hierarchy involves the conditional distribution of the latent field $\chi$ which is assumed to have a multivariate Gaussian prior with zero mean, such that it forms a Gaussian Markov random field with sparse precision matrix matrix $Q(\theta_2)$, i.e. $\chi\sim N(0,\ Q^{-1}(\theta_2))$, this is given as

$$p\ (\chi \mid \theta)=(2\pi)^{nm}|Q(\theta_2)|^{\frac{1}{2}} exp\left(-\frac{1}{2}\chi' Q(\theta_2)\chi\right)$$

Then at the final level of the hierarchy, a prior on the hyperparameter vector $p(\theta)$ can then be formulated for the set of hyperparameters $\theta=(\theta_1,\theta_2)$, which could be non-normal.

From this hierarchical Bayesian formulation the joint posterior distribution is then given as:

$$\begin{aligned} p\ (\mathcal{X},\theta|D) &\propto p\ (\theta)\ p\ (\mathcal{X} \mid \theta)\prod_{i} p\ (D_i \mid \mathcal{X},\theta) \\ &\propto p\ (\theta)\ |Q\ (\theta_2)|^{\frac{1}{2}} exp\left(-\frac{1}{2}\chi^{'} Q\ (\theta_2)\chi+\sum_{i=1}^{n} log\ (D_i \mid \chi_i,\theta)\right) \end{aligned} \tag{2.8}$$

Within this framework the joint posterior density (2.8) and subsequently the marginal posterior densities, $p\ (\chi_i \mid D);\ i=1\ ,\ldots\ ,n$ and $p\ (\theta|D)$ can be efficiently and accurately calculated using the integrated Laplace approximation methodology developed by Rue et al. [18]. The marginal posterior densities becomes

$$p\ (\chi_i \mid D)=\int p\ (\chi_i,\theta \mid D)d\theta=\int p\ (\chi_i,\theta \mid D)\ p\ (\theta|D)d\theta$$

and

$$p\ (\theta_i \mid D)=\int p\ (\theta \mid D)d\theta_{-j}$$

*2.5. Posterior Estimation using Integrated Laplace Approximation*

To obtain the posterior distribution of the model parameters under Bayesian framework, by Bayes' theorem, the conditional posterior distribution

$$p(\theta_i,\chi_i \mid D_i) = \frac{p(D_i \mid \theta_i,\chi_i)\, p(\theta_i,\chi_i)}{p(D_i)} \propto p(D_i \mid \theta_i,\chi_i)\, p(\chi_i \mid \theta_i)\, p(\theta_i)$$

where $p(\chi_i \mid \theta_i)$ and $p(\theta_i)$ are prior distributions and the focus is on approximating the multidimensional integral from the marginal likelihood $p(D_i \mid \theta_i,\chi_i)$ and approximation technique of INLA has been shown to provide exact approximations to the posterior estimates at faster rates than sampling-based methods such as Markov Chain Monte Carlo (MCMC) especially for complex and hierarchical models (see Rustand et al. [20]).

We consider the Laplace transformation using a second-order Taylor series expansion for the integral of the density function $p(\chi)$ by taking the form of (Blangiardo and Cameletti, [2]).

$$\int_{-\infty}^{\infty} p(\chi)d\chi = \int_{-\infty}^{\infty} exp(logp(\chi))\, d\chi = \int_{-\infty}^{\infty} exp(g(\chi))\, d\chi \tag{2.9}$$

Since for unimodal functions the integral value is mainly determined by the behaviour around the mode of $g(\chi)$, a second-order Taylor approximation of $g(\chi)$ can be substituted for $g(\chi)$ to calculate an approximate value of the integral.

Let $\chi^*$ be the global maximum of $\chi$ which is defined as

$$\chi^* = argmax_{\chi} g(\chi),$$

then

$$\left.\frac{\partial g(\chi)}{\partial \chi}\right|_{\chi=\chi^*} = 0$$

for $g(\chi)$ to be approximated as

$$g(\chi) \approx g(\chi^*) + 0.5(\chi-\chi^*)' Hg(\chi^*)(\chi-\chi^*)$$

where $Hg(\chi^*)$ is the Hessian of $g(\chi^*)$, and equation (2.9) can be written as

$$\int_{-\infty}^{\infty} p(\chi)d\chi = \int_{-\infty}^{\infty} exp(g(\chi^*) + 0.5(\chi-\chi^*)' Hg(\chi^*)(\chi-\chi^*))\, d\chi$$

$$= exp(g(\chi^*)) \int_{-\infty}^{\infty} exp(0.5(\chi-\chi^*)' Hg(\chi^*)(\chi-\chi^*))\, d\chi$$

$$= exp(g(\chi^*)) \int_{-\infty}^{\infty} exp(-0.5(\chi-\chi^*)' \{-Hg(\chi^*)\}(\chi-\chi^*))\, d\chi$$

$$= exp(g(\chi^*))\, (2\pi)^{\frac{nm}{2}} |Hg(\chi^*)|^{-\frac{1}{2}} \times$$

$$\int_{-\infty}^{\infty} (2\pi)^{-\frac{nm}{2}} |Hg(\chi^*)|^{-\frac{1}{2}} exp(-0.5(\chi-\chi^*)' \{-Hg(\chi^*)\}(\chi-\chi^*))\, d\chi$$

The integral is associated with the density of a multivariate Gaussian distribution and putting $-Hg(\chi^*) = Q(\chi^*)$, the precision matrix for the random vector $\chi^*$ yields

$$\int_{-\infty}^{\infty} p(\chi)d\chi \approx exp(g(\chi^*))(2\pi)^{\frac{nm}{2}}|Hg(\chi^*)|^{-\frac{1}{2}} \times$$

$$\int_{-\infty}^{\infty} (2\pi)^{-\frac{nm}{2}}|Q(\chi^*)|^{-\frac{1}{2}} exp(-0.5(\chi-\chi^*)' Q(\chi^*)(\chi-\chi^*))\, d\chi$$

$$\approx (2\pi)^{\frac{nm}{2}}|Q(\chi^*)|^{-\frac{1}{2}} exp(g(\chi^*)).$$

The conditional posterior distribution of $p(\chi, \theta|D)$ is defined from the joint posterior distribution in Equation (2.8) as

$$p(\chi, \theta|D) \propto p(\theta)|Q(\theta)|^{\frac{1}{2}} exp\left(-\frac{1}{2}\chi' Q(\theta)\chi + \sum_{i=1}^{n} logp(D_i \mid \chi_i, \theta)\right)$$

which can be rewritten as, ignoring elements with $\chi$.

$$p(\chi|\theta, D) \propto exp\left(-\frac{1}{2}\chi' Q(\theta)\chi + \sum_{i=1}^{n} g_i(\chi_i)\right) \tag{2.10}$$

*Gaussian approximation*

The Gaussian approximation of Equation (2.10), $p_G(\chi \mid \theta, D)$ is reached by matching the mode and the curvature at the mode of $p(\chi|\theta, D)$. The mode is computed iteratively by using a Newton Raphson method. Let $\mu^{(0)}$ be the initial value of the mode, and expand $g_i(\chi_i)$ around $\mu_i^{(0)} = \left(\mu_{i1}^{(0)}, \ldots, \mu_{iN}^{(0)}\right)$ to the second order Taylor expansion,

$$g_i(\chi_i) \approx g_i\left(\mu_i^{(0)}\right) + b_i'\chi_i - \frac{1}{2}c_i'\chi_i'\chi_i \tag{2.11}$$

where $b_i$ and $c_i$ depend on $\mu^{(0)}$. Substituting equation (2.11) into equation (2.10) yields

$$p_G(\chi \mid \theta, D) \approx g_i\left(\mu_i^{(0)}\right) exp\left(-\frac{1}{2}\chi'(Q+c)\chi + b'\chi\right)$$

$$\propto exp\left(-\frac{1}{2}\chi'(Q+c)\chi + b'\chi\right).$$

A Gaussian approximation of $p_G(\chi \mid \theta, D)$ is obtained, with the precision matrix $(Q+diag(c))$ and mode $\mu^{(2.1)}$, which is the solution of $(Q+diag(c))\mu^{(1)} = b$. The process can then be iterated, with $\mu^{(2.1)}$as the new starting value, until it converges to a Gaussian distribution with, say, mean $\mu^{(j)} \to \mu^{(*)} = \chi^*$ and precision matrix $Q^{(j)} \to Q^{(*)} = Q+diag(c^*)$, $j = 1, 2, \ldots$, where an appropriate convergence criterion must be used.

The resulting approximation will then be (Opitz, [14]):

$$p_G(\chi \mid \theta, D) \propto exp\left(-\frac{1}{2}(\chi-\chi^*(\theta))'(Q(\theta)+diag(c))(\chi-\chi^*(\theta))\right),$$

where $\boldsymbol{c}$ is the second-order term in the Taylor expansion of $\sum_{i=1}^{n} logp\left(D_i \mid \chi_i, \theta\right)$ at modal value $\chi^*(\theta)$.

For the marginal posterior conditional distribution $p(\chi_i \mid \theta, D)$ included in the computation of the marginal posterior $p(\chi_i \mid D)$, Rue et al. [18] discussed three approximations $\tilde{p}(\chi_i \mid \theta_k, D)$ where $\theta_k$ are weighted points to be used in the integration, Gaussian, full Laplace and simplified Laplace approximation. The Gaussian approximation is generally not best if the true density of $p(\chi_i \mid \theta, D)$ is not symmetric, the full Laplace approximation is a correction of Gaussian approximation and accurate but at a very expensive computational cost, the simplified Laplace approximation which is based on the Taylor series expansion of the full Laplace approximation is sufficiently accurate for most applications (Blangiardo and Cameletti, [2]).

If the mean of $\chi$ is $\mu$, the density of $\chi$ is

$$p(\chi) = (2\pi)^{-n/2}|Q|^{1/2} exp\left[-\frac{1}{2}(\chi-\mu)^T Q(\chi-\mu)\right] \tag{2.12}$$

The sparse matrix $Q$ is factorised as Cholesky triangle product $LL^T$, and only non-zero terms are computed due to the Markov property and $L_{ji} = 0$. Let $L^T\chi = r$ where $r \sim N(0,1)$**,** then we have that $L_{ii}\chi_i = r_i - \sum_{k=i+1}^{n} L_{ki}\chi_k$ for $i = n, \ldots, 1$. Multiplying each side with $\chi_j$, $j \geq i$ and taking the expectation yields the recursion

$$\Sigma_{ij} = \frac{\partial_{ij}^2}{L_{ii}} - \frac{1}{L_{ii}} \sum_{k=i+1}^{n} L_{ki}\Sigma_{kj} \qquad j \geq i, \quad i = n, \ldots, 1$$

where $\Sigma = Q^{-1}$ is the covariance matrix and $\partial_{ij} = 1$ if $i = j$ and $\partial_{ij} = 0$ otherwise. These recursion results in Gaussian approximations $\tilde{p}_G(\chi \mid \theta, D)$ with mean $\mu_i(\theta)$ and marginal variance $\sigma_i^2(\theta)$.

*Laplace approximation*

Laplace approximation is obtained from Gaussian approximation to $\chi_{-i}|\chi_i, \theta, D$, $\tilde{p}_{GG}$. Laplace approximation is given as

$$\tilde{p}_{LA}(\chi_i \mid \theta, D) \propto \left.\frac{p(\chi, \theta, D)}{\tilde{p}_{GG}(\chi_{-i}|\chi_i, \theta, D)}\right|_{\chi_{-i}=\chi_{-i}^*(\chi_i,\theta)} \tag{2.13}$$

Rue et al. [18] noted that $\tilde{p}_{GG}$ is recomputed for each value of $\chi_i$*and* $\theta$ since its precision matrix depends on $\chi_i$*and* $\theta$**,** and they proposed two modifications. The first avoids the optimisation step in computing $\tilde{p}_{GG}(\chi_{-i}|\chi_i, \theta, D)$ by approximating the modal configuration,

$$\chi_{-i}^*(\chi_i, \theta) \approx E_{\tilde{p}_G}(\chi_{-i}|\chi_i)$$

The second modification uses, for some $a_{ij}(\theta)$ when $j \neq i$,

$$\frac{E_{\tilde{p}_G}\left(\chi_j|\chi_i\right) - \mu_j(\theta)}{\sigma_j(\theta)} = a_{ij}(\theta)\frac{\chi_i - \mu_j(\theta)}{\sigma_i(\theta)}$$

The idea is to compute a set around $i$, $R_i(\theta)$, where only $\chi_j$s for which the dependence between $\chi_j$ and $\chi_i$ decay as the distance between nodes $i$ and $j$ increases determine the marginal of $\chi_i$, hence $R_i(\theta) = \left\{j : \left|a_{ij}(\theta)\right| > 0.001\right\}$.

The density $\tilde{p}_{LA}(\chi_i \mid \theta, D)$ is then represented by

$$\tilde{p}_{LA}(\chi_i \mid \theta, D) \propto N\left\{\chi_i; \mu_i(\theta), \sigma_i^2(\theta)\right\} exp\left\{cubic\ spline\left(\chi_i\right)\right\} \tag{2.14}$$

The cubic spline is fitted to the difference of the log-density of $\tilde{p}_{LA}(\chi_i \mid \theta, D)$ and $\tilde{p}_G(\chi_i \mid \theta, D)$at the selected abscissa points, and then the density is normalized by using quadrature integration (Rue et al. [18]).

*Simplified Laplace approximation*

The simplified Laplace approximation $\tilde{p}_{SLA}(\chi_i \mid \theta, D)$ is obtained via a series expansion of $\tilde{p}_{LA}(\chi_i \mid \theta, D)$ around $\chi_i = \mu_i(\theta)$ to correct the Gaussian approximation $\tilde{p}_G(\chi_i \mid \theta, D)$ for location and skewness. Assume

$$d_j^{(3)}(\chi_i, \theta) = \frac{\partial^3}{\partial \chi_j^3} log\{p(D_i \mid \chi_i, \theta)\}\Big|_{\chi_j = E_{\tilde{p}_G}(\chi_j \mid \chi_i)} \tag{2.15}$$

exists, then

$$log\left\{\tilde{p}_{SLA}\left(\chi_i^s \mid \theta, D\right)\right\} = constant - \frac{1}{2}\left(\chi_i^{(s)}\right)^2 + \gamma_i^{(1)}(\theta)\chi_i^{(s)} + \frac{1}{6}\left(\chi_i^{(s)}\right)^3 \gamma_i^{(3)}(\theta) + \cdots$$

where

$$\gamma_i^{(1)}(\theta) = \frac{1}{2}\sum_{j \in \mathfrak{T}} \sigma_i^2(\theta)\left\{1 - corr_{\tilde{p}_G}(\chi_i, \chi_j)^2\right\} d_j^{(3)}\{\mu_i(\theta), \theta\}\sigma_j(\theta) a_{ij}(\theta)$$

$$\gamma_i^{(3)}(\theta) = \sum_{j \in \mathfrak{T}} d_j^{(3)}\{\mu_i(\theta), \theta\}\left\{\sigma_j(\theta) a_{ij}(\theta)\right\}^3. \tag{2.16}$$

We refer to Rue et al. [18] for more details on the procedures as well as further discussions on the approximations error and its asymptotics with practical issues.

### *2.6. Joint Model diagnostics in INLA*

Conditional Predictive Ordinates (CPO) are a cross-validation criterion for model assessment computed for each observation as

$$CPO_i = p\left(\hat{D}_i = D_i \mid D_{-i}\right) = \int p\left(\hat{D}_i = D_i \mid \chi\right) p(\chi \mid D_{-i})\, d\chi$$

with $\hat{D}_i$ denoting the predicted longitudinal outcome and time to event for the $i$-th patient, $\chi$ the all-parameter vector, $D_{-i}$ the data vector without the $i$-th observation, and $p(\chi \mid D_{-i})$ the posterior distribution of $\chi$ predicted without $D_i$.

The CPO for each observation is the posterior probability of observing that observation when the model is fit using all the data except the one in question. Large values of CPO indicate a better fit of the model to the data, while small values implies not a good fit for that observation.

A measure that summarises CPO is

$$-\sum_{i=1}^{n} log(CPO_i)$$

with smaller values meaning better model fit.

The probability integral transform (PIT) is a predictive measure alternative to the CPO which is for validating and comparing models. The predictive density is examined for a subset of the observed data based on all the observations and it is given as

$$p(\chi_i \mid D_{-i}, \theta) \propto \frac{p(\chi_i \mid D, \theta)}{p(D_i \mid \chi_i, \theta)}$$

$$p(\chi_i \mid D_{-i}) \propto \frac{p(\theta \mid D)}{p(D_i \mid D_{-i}, \theta)}$$

where

$$p(D_i \mid D_{-i}, \theta) = \int p(D_i \mid \chi_i, \theta)\, p(\chi_i \mid D_{-i}, \theta) d\chi_i$$

$$PIT_i = Prob(D_i^{new} \leq (D_i \mid D_{-i}))$$

An unusually small or large $PIT_i$ (assuming continuous observations) indicates a possibly surprising observation which may require further attention (Rue et al. [18]). Furthermore, if the histogram of the $PIT_i$s is too far from a uniform distribution, the model can be questioned (Hicketier, [5]).

Deviance information criteria is noted to be a popular information criterion for hierarchical models and is well defined for improper priors in most cases (Rue et al. [18]). It is defined as two times the mean of the deviance minus the deviance of the mean. It is given as

$$DIC(\chi, \theta) = -2 \sum_{i \in \mathcal{I}} log\{p(D_i \mid \chi_i, \theta)\} + constant$$

The mean of the deviance is evaluated by first computing the conditional mean conditioned on $\theta$ by using univariate numerical integration for each $i \in \mathcal{I}$, then by integrating out $\theta$ with respect to $p(\theta \mid D)$. The deviance of the mean requires the posterior mean of each $\chi_i$, $i \in \mathcal{I}$, which is computed from the posterior marginals of $\chi_i$s.

The Watanabe-Akaike information criterion (WAIC) is also called the widely applicable Bayesian information criterion (WABIC) is similar to DIC only that the effective number of parameters is computed differently. See Watanabe [24] and Gelman et al. [3] for details.

$$WAIC = -2(D_{WAIC} - p_{WAIC}),$$

where $D_{WAIC}$ measures the fit of the model defined as $D_{WAIC} = \sum_{i=1}^{n} \sum_{m=1}^{M} logE_{\chi|D}[p(D_i \mid \chi)]$ and $p_{WAIC}$ denotes the effective number of parameters defined as $p_{WAIC} = \sum_{i=1}^{n} \sum_{m=1}^{M} Var_{posterior}[logp(D_i \mid \chi)]$. The lower the WAIC, the better the fit. This is preferable to the DIC because it computes the effective number of parameters for the variance for each data point separately, and then takes the sum (Gelman et al. [3]).

The marginal likelihood (MLIK) obtained in INLA can be used to compute posterior probabilities of models fitted and in turn the Bayes factor which can be used to compare, for example model $m_1$ and model $m_2$ as

$$\frac{p(m_1 \mid D)}{p(m_2 \mid D)} = \frac{p(D \mid m_1)}{p(D \mid m_2)} \times \frac{p(m_1)}{p(m_2)}$$

### *2.7. Description of the Renal transplantation Application Data*

The approach discussed in this paper will be applied to Renal transplantation data collected on 407 patients suffering from chronic kidney disease who underwent, between 21 January 1983 and 16 August 2000, a primary renal transplantation with a graft from a deceased or living donor in the University Hospital of the Catholic University of Leuven (Belgium). Chronic renal disease is a progressive loss of renal function over a period of months or years through five stages. Each stage is a progression through an abnormally low and progressively worse glomerular filtration rate (GFR) (Hickey et al. [7]). The

dataset records 3 repeated measures (2 continuous and 1 binary), and an event time. The clinical interest has been noted to lie in the long-term performance of the new graft, and especially in the time to graft failure survival. During the follow-up period, patients were periodically tested for the condition and performance of their kidneys. Our interest in this study is the longitudinal response variable GFR which measures the filtration rate of the kidneys. Rizopoulos and Ghosh [17] showed that the renal graft failure data is unusual in the subject-specific longitudinal evolutions, in that there exist highly nonlinear longitudinal profiles for the longitudinal outcomes, for which linear or quadratic models are not suitable. They proposed a natural cubic spline-based approach in order to flexibly capture the possibly nonlinear shapes of the subject-specific evolutions and explicitly tuned the degree of smoothness of the nonlinear evolutions by estimating the knots' position for the natural spline basis.

## 3. Simulation studies evaluation of the models

We evaluate the performances of the model with spline, quadratic and linear specifications for the longitudinal trajectories by working on simulation studies, the survival component was fitted with a mixture cure model, while varying the sample size of the simulated data as 50, 80, 150, 300, 500 and 1000 and the number of integration points was set to 500. We used the metrics of mean square error (MSE) and bias to see how each model estimated the true model parameters. The model parameters true values were chosen for the longitudinal fixed effect components as intercept was 4; covariates x1Y and x2Y were both 1; time variable timej was -0.5; residual error (variance) was 0.97. For the random effect, we chose intercept as 0.75; timej as 0.75; intercept and timej interaction was 0.3. For the survival component we chose Weibull shape parameter value as 1.3; Weibull scale value as 1.12; for incidence part, the covariates x1 and x2 both were 0.6; for latency part, the cure binary covariates z1 and z2 were both 0.3; and the association parameter as 0.25. The parameters values were used to simulate the datasets from using the models in equation (2.1), equation (2.3) and equation (2.5) and fitted using linear, quadratic and spline longitudinal trajectories. The results of the simulation are presented in Table 1, Table 2 and Table 3.

From the MSE and Bias values in the tables, we deduced from the results that the estimation of fixed effect of time and the random effect of time in the longitudinal component showed similar results with quadratic and spline specifications for the sample sizes considered except for larger sample size of 1000 where linear specification was better. For the mixture cure component, we saw that in small sample sizes linear specification resulted in closer estimates for latency and association parameters (as seen for sample sizes 50 and 80). The lower MSE and bias values for the mixture cure components and the association parameters for the spline specification showed that the spline functions resulted in better estimates for the incidence and latency parts as sample size increased.

Hence we conclude from the simulation results that sample size plays a crucial role in the estimation output no matter the longitudinal trajectories specification as the MSE and Bias varied across sample sizes. Overall, the spline specification was better in estimating the incidence and latency parameters of the mixture cure model. The nonlinear trajectory of the longitudinal outcome accounted for the shared random effect of the baseline covariate vectors x1 and x2 on the cure proportion.

**Table 1.** Simulation Results for N = 50 and 80

| | Linear | | Quadratic | | Spline | | Linear | | Quadratic | | Spline | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LONG. | N = 50 | | | | | | N = 80 | | | | | |
| *Fixed effect* | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS |
| intercept | 0.02650 | 0.16280 | 0.00002 | 0.00470 | 0.00312 | 0.05590 | 0.19097 | -0.43700 | 0.06729 | -0.25940 | 0.06770 | -0.26020 |
| x1Y | 0.21059 | -0.45890 | 0.53876 | -0.73400 | 0.33304 | -0.57710 | 0.48094 | 0.69350 | 0.02579 | -0.16060 | 0.13148 | 0.36260 |
| x2Y | 0.10621 | 0.32590 | 0.08283 | 0.28780 | 0.10543 | 0.32470 | 0.16728 | 0.40900 | 0.19927 | 0.44640 | 0.03415 | -0.18480 |
| timej | 0.31821 | -0.56410 | 0.00171 | 0.04140 | 0.92141 | -0.95990 | 0.53363 | -0.73050 | 1.10965 | -1.05340 | 0.07568 | -0.27510 |
| Res. err. (var.) | 0.02289 | -0.15130 | 0.02493 | -0.15790 | 0.02883 | -0.16980 | 0.20286 | 0.45040 | 0.20385 | 0.45150 | 0.19731 | 0.44420 |
| *Random effect* | | | | | | | | | | | | |
| intercept | 0.19237 | -0.43860 | 0.19572 | -0.44240 | 0.20061 | -0.44790 | 0.00560 | -0.07480 | 0.01259 | -0.11220 | 0.01687 | -0.12990 |
| timej | 0.12674 | -0.35600 | 0.12243 | -0.34990 | 0.10214 | -0.31960 | 0.15777 | -0.39720 | 0.12517 | -0.35380 | 0.12773 | -0.35740 |
| Inter* timej | 0.00582 | -0.07630 | 0.00604 | -0.07770 | 0.00654 | -0.08090 | 0.00021 | 0.01450 | 0.00104 | 0.03220 | 0.00031 | 0.01750 |
| SURVIVAL | | | | | | | | | | | | |
| shape | 0.00740 | -0.08600 | 0.00104 | 0.03220 | 0.00129 | -0.03590 | 0.00487 | -0.06980 | 0.00196 | -0.04430 | 0.00355 | -0.05960 |
| scale | 0.21734 | 0.46620 | 0.21902 | 0.46800 | 0.21800 | 0.46690 | 0.03629 | -0.19050 | 0.03549 | -0.18840 | 0.03679 | -0.19180 |
| *Incidence* | | | | | | | | | | | | |
| x1 | 0.00260 | -0.05100 | 0.00225 | -0.04740 | 0.00208 | -0.04560 | 0.21428 | 0.46290 | 0.24187 | 0.49180 | 0.23002 | 0.47960 |
| x2 | 0.04137 | -0.20340 | 0.04060 | -0.20150 | 0.04084 | -0.20210 | 0.03572 | 0.18900 | 0.05153 | 0.22700 | 0.04928 | 0.22200 |
| *Latency* | | | | | | | | | | | | |
| z1 | 0.59167 | -0.76920 | 0.59275 | -0.76990 | 0.59151 | -0.76910 | 0.17189 | -0.41460 | 0.16826 | -0.41020 | 0.15642 | -0.39550 |
| z2 | 0.00005 | -0.00720 | 0.00004 | -0.00610 | 0.00002 | -0.00460 | 0.00007 | -0.00830 | 0.00012 | -0.01080 | 0.00018 | -0.01360 |
| Association | 0.08638 | -0.29390 | 0.08638 | -0.29390 | 0.04490 | -0.21190 | 0.01318 | 0.11480 | 0.02226 | 0.14920 | 0.01880 | 0.13710 |

**Table 2.** Simulation Results for N = 150 and 300

| | Linear | | Quadratic | | Spline | | Linear | | Quadratic | | Spline | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LONG. | N = 150 | | | | | | N = 300 | | | | | |
| *Fixed effect* | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS |
| intercept | 0.04105 | -0.20260 | 0.00998 | -0.09990 | 0.03014 | -0.17360 | 0.04635 | -0.21530 | 0.05162 | -0.22720 | 0.04951 | -0.22250 |
| x1Y | 0.04584 | 0.21410 | 0.16695 | -0.40860 | 0.00000 | -0.00060 | 0.01891 | 0.13750 | 0.18879 | -0.43450 | 0.00714 | -0.08450 |
| x2Y | 0.01809 | 0.13450 | 0.01166 | 0.10800 | 0.06677 | 0.25840 | 0.00006 | -0.00800 | 0.00001 | -0.00380 | 0.00245 | 0.04950 |
| timej | 0.08352 | -0.28900 | 0.36918 | -0.60760 | 0.03580 | 0.18920 | 0.12222 | -0.34960 | 0.20630 | -0.45420 | 0.02602 | -0.16130 |
| Res. err. (var.) | 0.00575 | -0.07580 | 0.00716 | -0.08460 | 0.00808 | -0.08990 | 0.00248 | -0.04980 | 0.00245 | -0.04950 | 0.00224 | -0.04730 |
| *Random effect* | | | | | | | | | | | | |
| intercept | 0.00496 | -0.07040 | 0.00697 | -0.08350 | 0.00654 | -0.08090 | 0.01094 | 0.10460 | 0.01034 | 0.10170 | 0.00943 | 0.09710 |
| timej | 0.03165 | -0.17790 | 0.00874 | -0.09350 | 0.00796 | -0.08920 | 0.16088 | 0.40110 | 0.17539 | 0.41880 | 0.13119 | 0.36220 |
| inter*timej | 0.04198 | -0.20490 | 0.03486 | -0.18670 | 0.04048 | -0.20120 | 0.09248 | -0.30410 | 0.09753 | -0.31230 | 0.09102 | -0.30170 |
| SURVIVAL | | | | | | | | | | | | |
| shape | 0.00336 | 0.05800 | 0.00317 | 0.05630 | 0.00282 | 0.05310 | 0.00207 | -0.04550 | 0.00194 | -0.04410 | 0.00154 | -0.03920 |
| scale | 0.00050 | -0.02230 | 0.00003 | 0.00560 | 0.00000 | 0.00110 | 0.00198 | 0.04450 | 0.00213 | 0.04620 | 0.00186 | 0.04310 |
| *Incidence* | | | | | | | | | | | | |
| x1 | 0.04260 | 0.20640 | 0.03779 | 0.19440 | 0.03437 | 0.18540 | 0.00248 | 0.04980 | 0.00228 | 0.04770 | 0.00104 | 0.03220 |
| x2 | 0.03222 | 0.17950 | 0.02969 | 0.17230 | 0.02873 | 0.16950 | 0.00098 | 0.03130 | 0.00076 | 0.02760 | 0.00011 | 0.01030 |
| *Latency* | | | | | | | | | | | | |
| z1 | 0.03996 | -0.19990 | 0.04372 | -0.20910 | 0.04524 | -0.21270 | 0.00015 | 0.01240 | 0.00013 | 0.01160 | 0.00001 | 0.00310 |
| z2 | 0.00008 | 0.00920 | 0.00002 | 0.00420 | 0.00001 | 0.00300 | 0.00006 | -0.00780 | 0.00008 | -0.00910 | 0.00025 | -0.01590 |
| Association | 0.03501 | 0.18710 | 0.02676 | 0.16360 | 0.02602 | 0.16130 | 0.00091 | -0.03020 | 0.00094 | -0.03060 | 0.00156 | -0.03950 |

**Table 3.** Simulation Results for N = 500 and 1000

| | Linear | | Quadratic | | Spline | | Linear | | Quadratic | | Spline | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LONG. | N = 500 | | | | | | N = 1000 | | | | | |
| *Fixed effect* | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS | MSE | BIAS |
| intercept | 0.00546 | -0.07390 | 0.00599 | -0.07740 | 0.00371 | -0.06090 | 0.00263 | -0.05130 | 0.00311 | -0.05580 | 0.00319 | -0.05650 |
| x1Y | 0.01501 | 0.12250 | 0.19202 | -0.43820 | 0.00212 | -0.04600 | 0.00990 | 0.09950 | 0.20621 | -0.45410 | 0.00552 | -0.07430 |
| x2Y | 0.00812 | 0.09010 | 0.00834 | 0.09130 | 0.00182 | -0.04270 | 0.01518 | 0.12320 | 0.01510 | 0.12290 | 0.12938 | 0.35970 |
| timej | 0.11169 | -0.33420 | 0.06817 | -0.26110 | 0.05890 | -0.24270 | 0.18361 | -0.42850 | 0.43573 | -0.66010 | 0.01820 | 0.13490 |
| Res. err. (var.) | 0.00951 | -0.09750 | 0.00931 | -0.09650 | 0.00949 | -0.09740 | 0.00190 | -0.04360 | 0.00247 | -0.04970 | 0.00292 | -0.05400 |
| *Random effect* | | | | | | | | | | | | |
| intercept | 0.04516 | 0.21250 | 0.04783 | 0.21870 | 0.04744 | 0.21780 | 0.00129 | -0.03590 | 0.00174 | -0.04170 | 0.00175 | -0.04180 |
| timej | 0.41977 | 0.64790 | 0.38465 | 0.62020 | 0.43600 | 0.66030 | 0.02002 | 0.14150 | 0.05217 | 0.22840 | 0.06770 | 0.26020 |
| Inter*timej | 0.15817 | -0.39770 | 0.15445 | -0.39300 | 0.16281 | -0.40350 | 0.08066 | -0.28400 | 0.09114 | -0.30190 | 0.09591 | -0.30970 |
| SURVIVAL | | | | | | | | | | | | |
| shape | 0.00003 | -0.00510 | 0.00003 | -0.00530 | 0.00001 | -0.00350 | 0.00392 | -0.06260 | 0.00320 | -0.05660 | 0.00312 | -0.05590 |
| scale | 0.01212 | -0.11010 | 0.01237 | -0.11120 | 0.01184 | -0.10880 | 0.00339 | -0.05820 | 0.00265 | -0.05150 | 0.00264 | -0.05140 |
| *Incidence* | | | | | | | | | | | | |
| x1 | 0.03226 | 0.17960 | 0.03172 | 0.17810 | 0.03151 | 0.17750 | 0.02641 | -0.16250 | 0.02699 | -0.16430 | 0.02650 | -0.16280 |
| x2 | 0.00249 | -0.04990 | 0.00258 | -0.05080 | 0.00263 | -0.05130 | 0.00052 | 0.02280 | 0.00036 | 0.01890 | 0.00028 | 0.01660 |
| *Latency* | | | | | | | | | | | | |
| z1 | 0.00245 | -0.04950 | 0.00250 | -0.05000 | 0.00253 | -0.05030 | 0.00197 | -0.04440 | 0.00228 | -0.04770 | 0.00243 | -0.04930 |
| z2 | 0.00046 | 0.02150 | 0.00046 | 0.02140 | 0.00042 | 0.02040 | 0.00075 | -0.02730 | 0.00097 | -0.03110 | 0.00102 | -0.03200 |
| Association | 0.00233 | -0.04830 | 0.00229 | -0.04790 | 0.00231 | -0.04810 | 0.00000 | -0.00170 | 0.00003 | -0.00560 | 0.00005 | -0.00740 |

## 4. Application to Renal Transplantation Data

### *4.1. Exploratory analysis on renal transplantation data*

The longitudinal trajectories of the glomerular filtration rate (GFR) of the kidneys of the 407 patients show that it is neither linear nor could be expressed as quadratic trajectory (see Figure 1 on the longitudinal plots of GR) as noted by Rizopoulos and Ghosh [17].

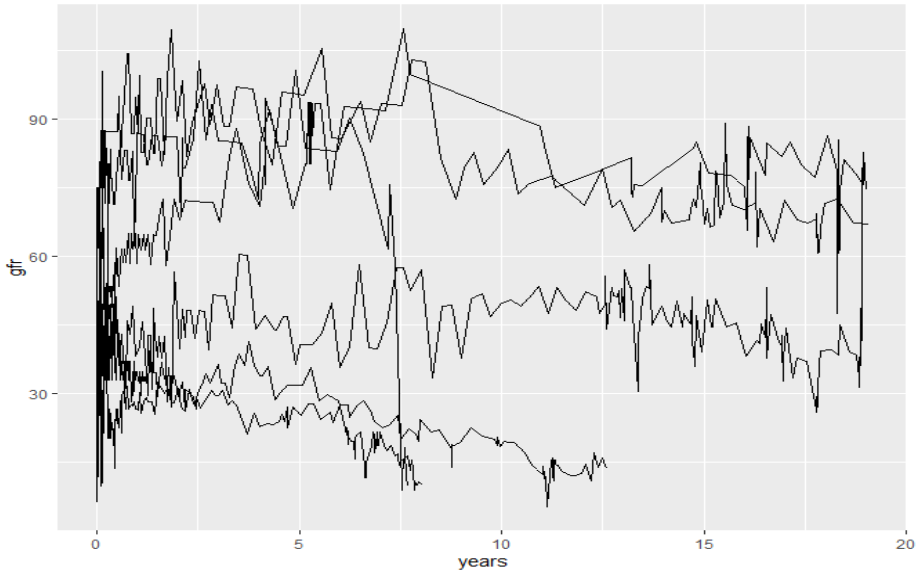


**Figure 1.** Spaghetti Plot of GFR of Patient's kidney (ml/min/1.73m)

We decided to see how these nonlinear trajectories are distributed between the female and male gender (Figure 2) and we observed quite a remarkable difference in the longitudinal trajectories of the glomerular filtration rate of the kidneys of female patients from the male patients. We see that the GFR of female patients were remarkably lower than the GFR of male patients, though there were also some male patients who had lower GFR also, within the range of those of observed female patients with equally lower GFR. This can give us clue to the dichotomous nature of the survival event as to the possibilities of some patients no experiencing the graft failure during the follow-up period.

Figure 3 further shows the longitudinal trajectories of the GFR of the kidneys of 6 randomly selected patients and we also observed that the nonlinear trajectories of the GFR is clearly seen for each of the GFR in each patient. Interest decided to see how these nonlinear trajectories are distributed between the female and male sexes impact the time to graft failure or the long term survivor of the kidneys.

We further describe the GFR data using other covariate variables such as age and weights of patients including gender of patients to see how the event status and GFR longitudinal trajectories varies. Figure 4 shows the boxplots of age of patient at day of surgery, preoperative weight of patients and GFR of patients against the event status and gender of Patients. The first frame shows that the mean age of patients at day of surgery was lower in patients that had graft failure than those who did not (especially among female patients, as it can be seen from the plot that interestingly the mean age of male patients seem to be similar of both event groups) in the period of follow-up.

The preoperative weight of the patients was remarkably different between the patients who experience graft failure and those who did not. The same pattern of the mean weight of female patients lower

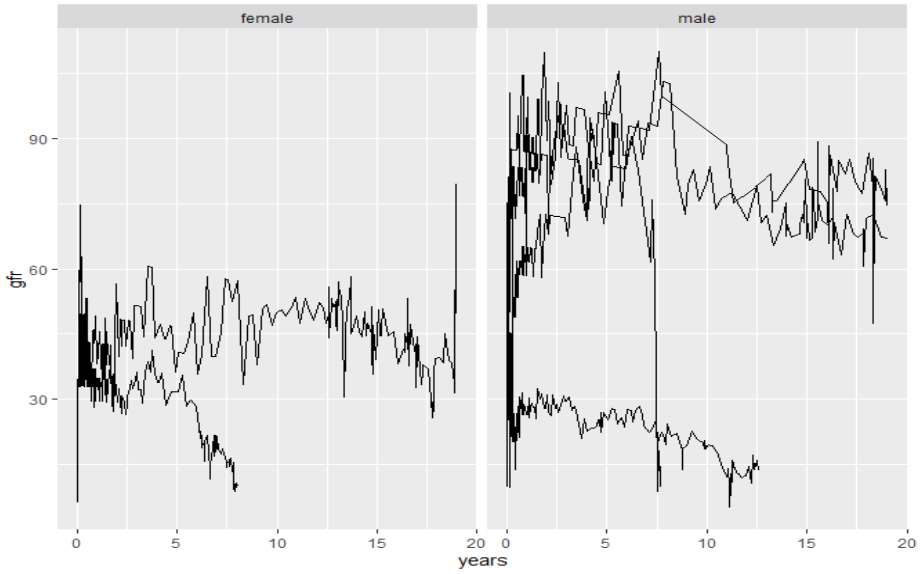


**Figure 2.** Spaghetti Plot of GFR of kidneys of Female and Male Patients

than those of the male patients is observed in both the graft failure group and the censored group. For the GFR of the kidneys of the patients, the mean GFR was around 40 to 50 ml/min/1.73m, for which the female patients had lower GFR compared to the male patients in both the graft failure and the censored groups. As noted in the previous sections, chronic kidney disease is a progressive loss of renal function over a period of months or years through different stages consisting of a progression through an abnormally low and progressively worse glomerular filtration rate, there seemed to be some abnormally in the data as there were unusual data-points with high GFR values and yet appeared under the graft failure group. This observation is seen with the tails in the density plot of GFR shown in Figure 5.

From Figure 5, we observe that the mean value of the GFR is higher in the censored group than the graft failure group, even though their tails at higher values waned together. Similar pattern is observed for the gender factor as the mean value of GFR is higher in male patients than female patients, yet there are both female and male patients with higher values of GFR.

From Figure 6 which shows the different survival curves estimated by splitting the data according to the different types of genders, we observe that survival seems to be worst for male patients than for female patients, which shows that the female patients group has higher survival than the male patient group.

The result from the Kaplan-Meier estimates of the probability of survival for the female and male genders shows that there is high survival among the female patients, but the data descriptions earlier seen showed that female patients had lower GFR than male patients, hence, looking at the longitudinal observations and the survival data separately can lead to conflicting conclusions (see Hickey et al. [6]). Also, in survival modelling, censorship and time to event are assumed to be independent, however, there may be association between these two variables for which other covariates observed may be able to explain. Clinical state can be modelled on several covariates using longitudinal data and censoring also depends on these covariates (Gómez-Rubio, [4]), hence the need for joint modelling, which puts into consideration the association between the covariates and the survival variables.

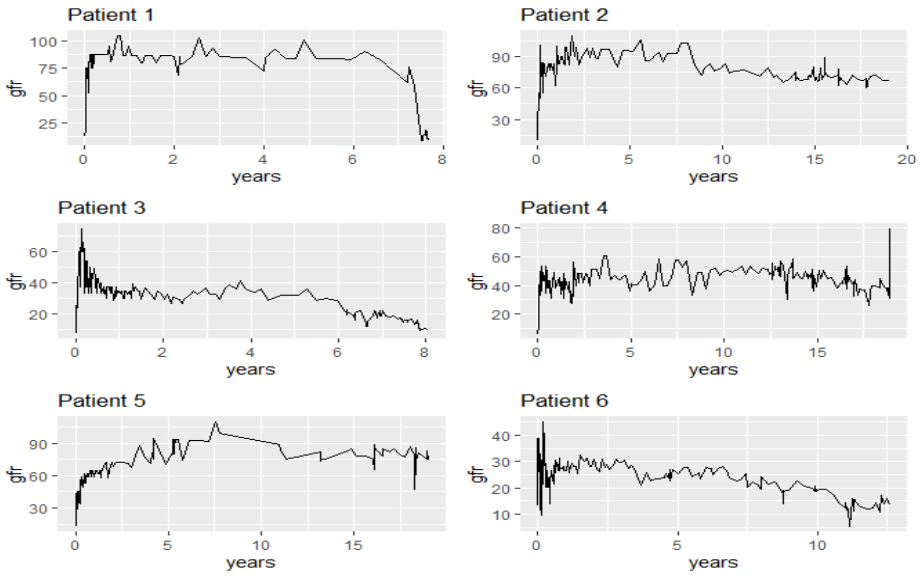


**Figure 3.** Spaghetti Plot of GFR of the kidneys of 6 randomly selected Patients

### *4.2. Joint modelling on application data and Discussion of Results*

Our approach here is to fit the joint model for the renal transplantation data using the LGM framework as implemented with INLA by considering the possibilities of cure proportion in the survival component of the modelling.

Longitudinal component:

$$gfr_{i,m}(t) = \eta_i^L(t) + \varepsilon_i(t)\,(\beta_{10}+b_{i10}) + \left(b_{i11}+\beta_{yr1}\right) years_i + \beta_{wt1} weight_i + \beta_{age1} age_i + \beta_{male1} gender_i + \beta_{yr2} years_i F_1(t) + \beta_{wt2} weight_i F_1(t) + \beta_{age2} age_i F_1(t) + \beta_{yr3} years_i F_2(t) + \beta_{wt3} weight_i F_2(t) + \beta_{age3} age_i F_2(t) + \varepsilon_i(t)$$

Incidence model component:

$$logit\left[\pi_i(\beta_1)\right] = \beta_{20} + \beta_{wt4} weight_i + \beta_{age4} age_i + \beta_{male2} gender_i$$

Latency model component:

$$h_{iu}(t \mid h_{u0},\ \beta_2) = \left(\lambda\alpha t^{\alpha-1}\right) exp\left\{\eta_i^L(t) + \gamma\left(\beta_{wt5} weight_i + \beta_{age5} age_i + \beta_{male3} gender_i\right)\right\}$$

We fit two numbers natural cubic spline basis functions $F_1(t)$ and $F_2(t)$ with internal knots at 2 years for longitudinal outcome GFR. For priors specification, as noted earlier, we assume a multivariate Gaussian prior for the latent field $\chi$ with precision matrix $Q(\theta)$ conditioned on $\theta$, which we also assume prior distributions $p(\theta)$ for which all the regression coefficients and the Weibull log(λ) scale parameters follow a vague normal distribution centred at zero ($\mathcal{N}(0, 1000)$) while the shape, $\alpha \sim$ Ga(0.01, 0.01), parameter assumes a log-gamma distribution.

The result of the modelling of the glomerular filtration rate (GFR) jointly with the risk of renal graft failure after kidney transplantation is given in the following tables showing the longitudinal component and the cure survival output comprising the incidence and latency models incorporated in the joint

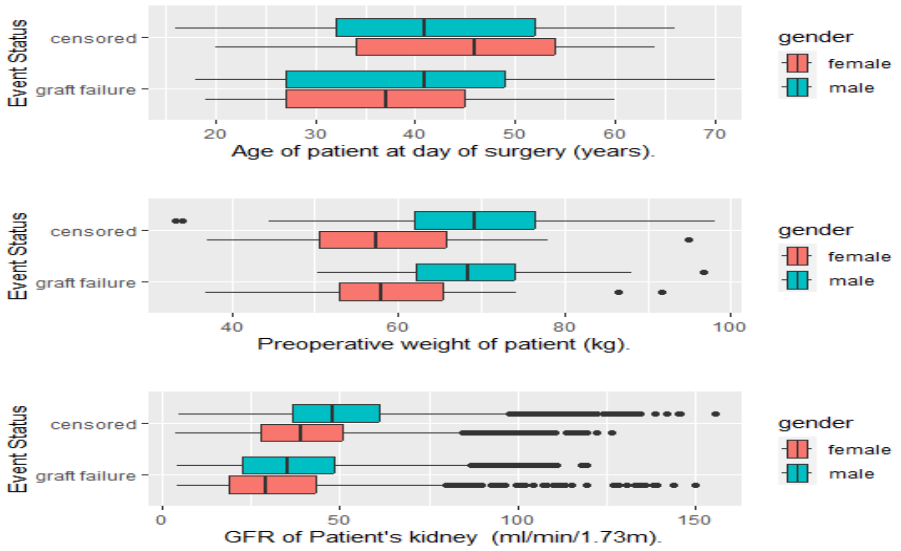


**Figure 4.** Boxplot of Age, Weight and GFR vs Event status and Gender of Patients

model coefficients outputs. Table 4 shows that the time posterior with the spline components are significant in the nonlinear trajectory of GFR except for the linear coefficient, and we see similar significance for the covariates of weight, age and gender. The random effects of intercept and slope are respectively significant 205.113 with 95% credible interval (181.523, 233.550) and 16.181 with 95% credible interval (14.284, 18.159).

From Table 5 regarding the association with clinical endpoints, we observed some negative association between the GFR and the risk of graft failure, -0.109 with 95% credible interval of -0.119 and -0.100. We see that smaller values of GFR as seen by the negative association parameter are associated with a higher risk for a graft failure. We can also observe from the baseline covariates that only gender seemed to impact to some extent the risk of graft failure as the risk for graft failure is 0.113 times more likely in males than in females, we saw this pattern from Kaplan-Meier estimates of the probability of survival in Figure 6.

Figure 7 shows the patient-specific intercepts and the respective unique patient identifications for each patient (random intercept) in the longitudinal and survival components. Estimates of the shared random effects from the longitudinal component copied to the survival component are used check for dependence between the longitudinal and survival components. From Figure 7, there is a linear dependence between them because the random effects in the survival component are the random effects in the longitudinal component multiplied by a constant scaling factor, represented by the red line whose slope is the posterior mean of the scaling factor.

### *4.3. Comparing different specification for longitudinal GFR trajectory*

Comparing the nonlinear trajectory specification for the longitudinal process with a linear and a quadratic model specifications using the model predictive measures of DIC, WAIC, CPO and marginal likelihood (MLIK), shows that the spline specification for the longitudinal outcome, GFR is a better fit than the linear and quadratic specifications from the table below. This is shown in Table 6 by the smaller values of DIC, WAIC, CPO and bigger marginal likelihood values for the spline longitudinal

**Table 4.** Longitudinal Component Model Coefficients of Joint model with Mean Posterior and 95% credible Intervals

| Longitudinal outcome: - | | | | |
|---|---|---|---|---|
| *Fixed effects*: | Mean | SD | Lower 95% CI | Upper 95% CI |
| $\beta_{10}$ | 3.088 | 9.181 | -14.906 | 21.083 |
| $\beta_{yr1}$ | -0.310 | 0.415 | -1.123 | 0.503 |
| $\beta_{yr2}$ | 13.546 | 7.205 | -0.575 | 27.668 |
| $\beta_{yr3}$ | -14.101 | 6.931 | -27.686 | -0.516 |
| $\beta_{wt1}$ | 1.092 | 0.924 | -0.719 | 2.902 |
| $\beta_{age1}$ | 0.210 | 0.925 | -1.602 | 2.022 |
| $\beta_{male1}$ | 2.162 | 0.684 | 0.821 | 3.503 |
| $\beta_{wt2}$ | 3.466 | 8.928 | -14.033 | 20.965 |
| $\beta_{age2}$ | 3.547 | 8.929 | -13.953 | 21.048 |
| $\beta_{wt3}$ | -3.538 | 8.807 | -20.799 | 13.722 |
| $\beta_{age3}$ | -3.675 | 8.807 | -20.937 | 13.586 |
| $\sigma_{\epsilon}$ | 128.153 | 0.899 | 126.178 | 129.650 |
| *Random effects*: | | | | |
| $\sigma_{b10}$ | 205.113 | 13.509 | 181.523 | 233.550 |
| $\sigma_{b11}$ | 16.181 | 1.004 | 14.284 | 18.159 |
| $cov_{b10,b11}$ | -10.585 | 3.606 | -16.657 | -2.800 |

**Table 5.** Cure Survival Components Model Coefficients of Joint model with Mean Posterior and 95% credible Intervals

| Survival outcome: - | | | | |
|---|---|---|---|---|
| *Incidence*: | Mean | SD | Lower 95% CI | Upper 95% CI |
| $\beta_{20}$ | -1.534 | 0.654 | -2.816 | -0.252 |
| $\beta_{wt4}$ | 0.001 | 0.011 | -0.020 | 0.022 |
| $\beta_{age4}$ | 0.017 | 0.009 | 0.000 | 0.034 |
| $\beta_{male2}$ | 0.065 | 0.238 | -0.401 | 0.532 |
| *Latency*: | | | | |
| $\alpha$ | 0.975 | 0.001 | 0.973 | 0.976 |
| $\lambda$ | 0.016 | 0.011 | 0.004 | 0.044 |
| $\beta_{wt5}$ | -0.001 | 0.011 | -0.023 | 0.020 |
| $\beta_{age5}$ | -0.007 | 0.009 | -0.024 | 0.010 |
| $\beta_{male3}$ | 0.113 | 0.225 | -0.329 | 0.555 |
| *Association parameter*: | | | | |
| $\gamma$ | -0.109 | 0.005 | -0.119 | -0.100 |

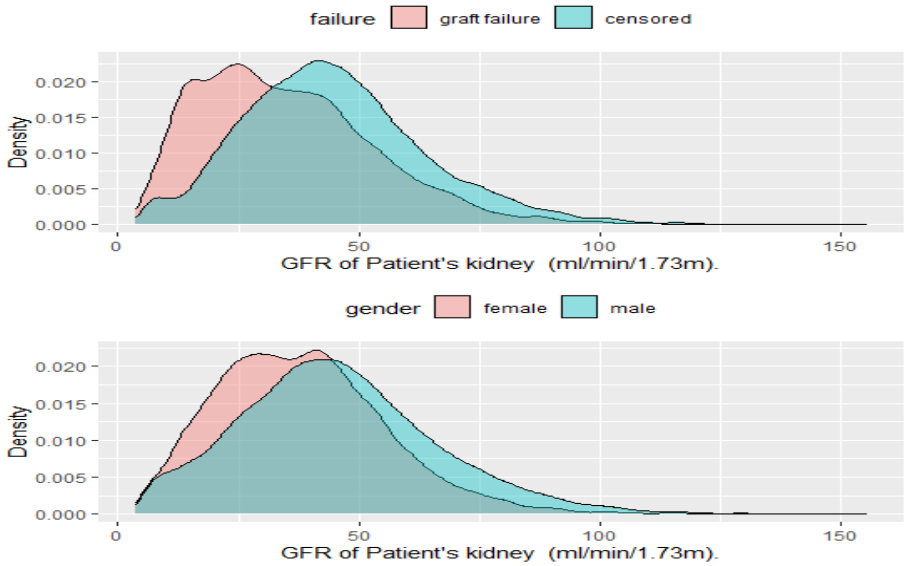


**Figure 5.** Density Plot of GFR vs Event status and Gender of Patients

**Table 6.** Summary of values of DIC, WAIC, CPO and marginal likelihood for Linear, Quadratic and Spline fits to the Renal GFR data.

| Model | DIC | WAIC | CPO | MLIK | Time (secs.) |
|---|---|---|---|---|---|
| Spline trajectory | 512648.70 | 511513.90 | 259609.95 | -289643.80 | 1330.73 |
| Linear trajectory | 528069.60 | 527405.70 | 266069.60 | -284868.00 | 208.52 |
| Quadratic trajectory | 579324.30 | 578161.80 | 292799.30 | -322290.20 | 2461.22 |

model compared to those values for the linear and quadratic longitudinal model specifications. Best model is one with smaller DIC, WAIC, CPO and bigger Marginal likelihood.

This better fit of spline function for GFR using INLA is in line with the result reported by Rizopoulos and Ghosh [17] who used natural cubic splines basis specification for the longitudinal outcomes using MCMC run with WinBUGS for which they noted that a linear fit revealed the potential bias in measuring the effect of the true underlying longitudinal outcomes to the time-to-graft failure.

## 5. Conclusion

This paper presented the modelling of longitudinal outcome and mixture cure survival event under shared random effect using latent Gaussian modelling approach, which involves the deterministic approximate Bayesian inference of Laplace approximation in evaluating the posterior distribution of the resulting Bayesian modelling. The paper looked at joint model with longitudinal and cure survival outcomes expressed as a latent Gaussian model and hence how the integrated Laplace approximation introduced by Rue et al. [18] can be used to evaluate its posteriors.

For the longitudinal outcome a spline specification was used to capture its complex evolution and the survival cure component is based on the specifications by Lázaro et al., [9], in which latent indicators in the inferential process for classifying individuals in the cured and uncured groups was introduced. The INLA was used to fit the marginal posterior distribution for the longitudinal component, survival

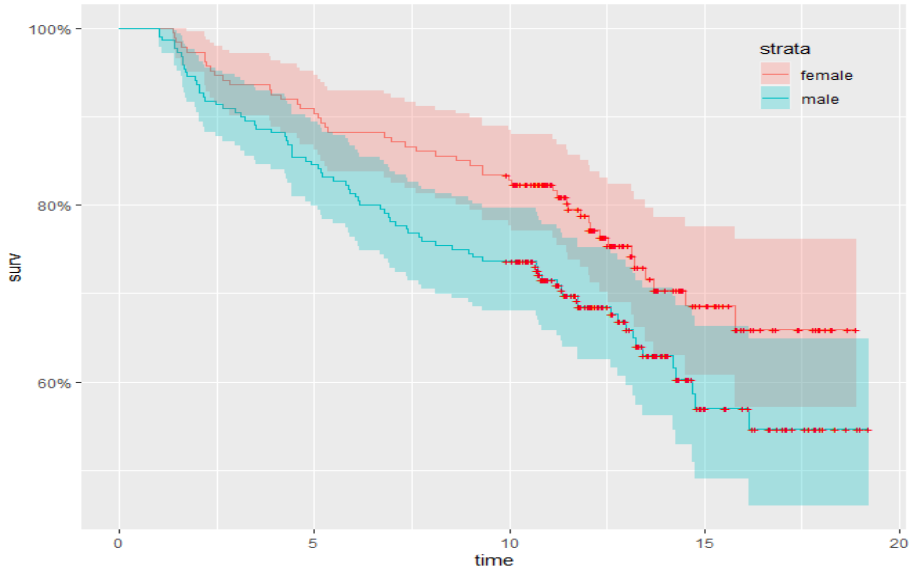


**Figure 6.** Survival estimates of patients according to gender of patients

component and the relevant element given the latent indicator variable classifying subjects as cured or uncured.

The modelling approach was applied to the Renal transplantation data collected on 407 patients suffering from chronic kidney disease who underwent, between 21 January 1983 and 16 August 2000, a primary renal transplantation with a graft from a deceased or living donor in the University Hospital of the Catholic University of Leuven (Belgium), with focus on the longitudinal biomarker glomerular filtration rate (GFR) and time to graft failure as survival event. Two numbers natural cubic spline basis functions with internal knots at 2 years was fit for longitudinal outcome, GFR and for priors specification, a multivariate Gaussian prior for the latent field was assumed with precision matrix. All the regression coefficients and the Weibull log-scale parameters follow a vague normal distribution centred at zero, while the shape parameter followed a log-gamma distribution.

A simulation study on the performance of linear, quadratic and spline specifications for the longitudinal nonlinear trajectories and mixture cure was carried out, with varying sample sizes and comparisons were made using MSE and Bias. We deduced from the results that the estimation of fixed effect of time and the random effect of time in the longitudinal component showed similar results with quadratic and spline specifications for the sample sizes considered except for larger sample size of 1000 where linear specification was better. For the mixture cure component, we saw that in small sample sizes linear specification resulted in closer estimates for latency and association parameters as sample size increased, the MSE and bias values for the mixture cure components and the association parameters with spline specification showed lowest values indicating better estimates. Sample size plays a crucial role in the estimation output no matter the longitudinal trajectories specification as the MSE and Bias varied across sample sizes. Overall, the spline specification was best in estimating the incidence and latency parameters of the mixture cure model.

We saw from the application results from Renal transplantation data that the baseline covariates for only gender seemed to impact to some extent the risk of graft failure as the risk for graft failure is 0.113 times more likely in males than in females, we saw this pattern from Kaplan-Meier estimates of the

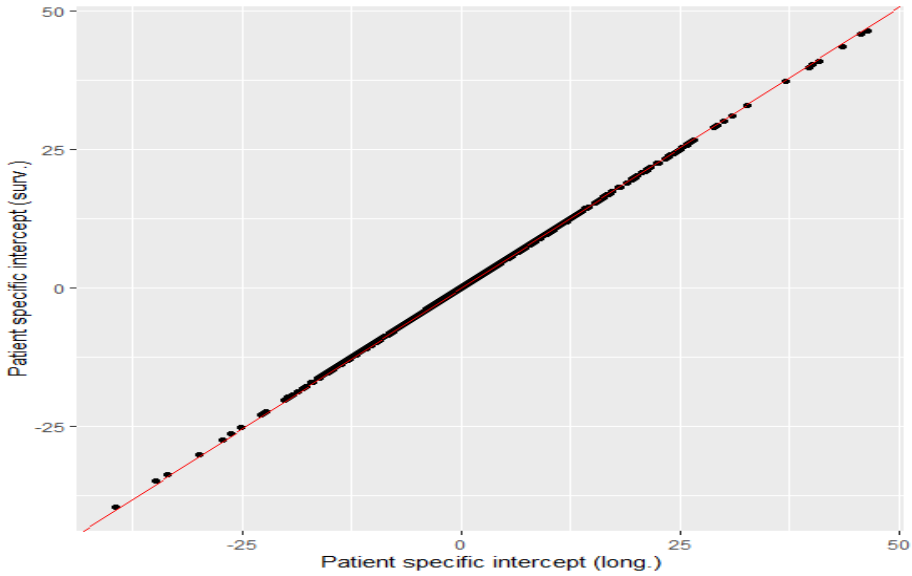


**Figure 7.** Patient-Specific Intercepts in Longitudinal model vs their copied values in survival sub-model

probability of survival. The odds of being cured was therefore higher in females than for males after renal transplantation.

The merit of this approach is that apart from capturing the nonlinear trajectory of the longitudinal outcome, the association between the longitudinal component and survival components which included cure fractions was also estimated. The proportion of cure patients as a function of the association between longitudinal and event-time process can be modelled under the latent Gaussian model framework. The suitability of the spline nonlinear trajectory specification for the longitudinal process was seen by comparing it with a linear and a quadratic model specifications using DIC, WAIC and CPO and computation time. The results showed that the spline specification was best for application data.

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Declaration of competing interest**

The authors declare that they have no known competing financial or personal interests that could have appeared to influence the work reported in this paper.

## References

1. Alsefri, M., Sudell, M., García-Fiñana, M., & Kolamunnage-Dona, R. (2020). Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology*, 20(1), 94. https://doi.org/10.1186/s12874-020-00976-2
2. Blangiardo M. and Cameletti M. (2015). Spatial and Spatio-temporal Bayesian Models with R –INLA. John Wiley & Sons, Chichester

3. Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding Predictive Information criteria for Bayesian Models. *Statistics and Computing* 24(4): 997 – 1016. https://doi.org/10.1007/s1122-013-9416-2.

4. Gómez-Rubio, V. (2020). Bayesian Inference with INLA. In Bayesian Inference with INLA. Chapman and Hall/CRC. https://doi.org/10.1201/9781315175584

5. Hicketier A. (2015). Spatio-temporal modeling of Hantavirus in Germany. Thesis, University of Stockholm, Sweden

6. Hickey, G. L., Philipson, P., Jorgensen, A., & Kolamunnage-Dona, R. (2018a). Joint Models of Longitudinal and Time-to-Event Data with More Than One Event Time Outcome: A Review. *The International Journal of Biostatistics*, 14(1). https://doi.org/10.1515/ijb-2017-0047

7. Hickey, G. L., Philipson, P., Jorgensen, A., & Kolamunnage-Dona, R. (2018b). JoineRML: A joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Medical Research Methodology*, 18(1), 1–14. https://doi.org/10.1186/s12874-018-0502-1

8. Hu, W., & Zhou, J. (2019). Joint modeling: an application in behavioural scoring. *Journal of the Operational Research Society*, 70(7), 1129–1139. https://doi.org/10.1080/01605682.2018.1487821

9. Lázaro, E., Armero, C., & Gómez-Rubio, V. (2020). Approximate Bayesian inference for mixture cure models. *Test*, 29(3), 750–767. https://doi.org/10.1007/s11749-019-00679-x

10. Martins, R., Silva, G. L., & Andreozzi, V. (2016). Bayesian joint modeling of longitudinal and spatial survival AIDS data. *Statistics in Medicine*, 35(19), 3368–3384. https://doi.org/10.1002/sim.6937

11. Martins, R., Silva, G. L., & 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

12. Medina-Olivares, V., Calabrese, R., Crook, J., & Lindgren, F. (2023a). Joint models for longitudinal and discrete survival data in credit scoring. *European Journal of Operational Research*, 307(3), 1457–1473. https://doi.org/10.1016/J.EJOR.2022.10.022

13. Medina-Olivares, V., Lindgren, F., Calabrese, R., & Crook, J. (2023b). Joint models of multivariate longitudinal outcomes and discrete survival data with INLA: An application to credit repayment behaviour. *European Journal of Operational Research*, 310(2), 860–873. https://doi.org/10.1016/j.ejor.2023.03.012

14. Opitz T (2017). Latent Gaussian modeling and INLA: A review with focus on space-time applications. *Journal of the French Statistical Society* 158(3):62-85

15. Peng, Y., & Taylor, J. M. G. (2014). Cure Models. In Handbook of Survival Analysis (pp. 113–134).

16. Rappl, A., Kneib, T., Lang, S., & Bergherr, E. (2023). Spatial Joint Models through Bayesian Structured Piece-wise Additive Joint Modelling for Longitudinal and Time-to-Event Data. http://arxiv.org/abs/2302.07020

17. 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

18. Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71(2), 319–392. https://doi.org/10.1111/J.1467-9868.2008.00700.X

19. Rustand, D., van Niekerk, J., Rue, H., Tournigand, C., Rondeau, V., & Briollais, L. (2023). Bayesian estimation of two-part joint models for a longitudinal semicontinuous biomarker and a terminal event with INLA: Interests for cancer clinical trial evaluation. *Biometrical Journal*, 65(4). https://doi.org/10.1002/bimj.202100322

20. Rustand, D., van Niekerk, J., Krainski, E. T., Rue, H., & Proust-Lima, C. (2024). Fast and flexible inference for joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations. *Biostatistics*, 2024, 25, 2, pp. 429–448. https://doi.org/10.1093/biostatistics/kxad019

21. Tsiatis, A. A., & Davidian, M. (2004). Joint modelling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14(3), 809–834.

22. van Niekerk, J., Bakka, H., & Rue, H. (2019). Joint models as latent Gaussian models - not reinventing the wheel. http://arxiv.org/abs/1901.09365

23. van Niekerk, J., Bakka, H., & Rue, H. (2021). Competing risks joint models using R-INLA. *Statistical Modelling*, 21(1–2), 56–71. https://doi.org/10.1177/1471082X20913654

24. Watanabe, S. (2013). A widely Applicable Bayesian Information Criterion. *Journal of Machine Learning Research* 14: 867 – 897.