Sensitivity analysis on joint modelling of longitudinal and mixture cure outcomes with data missingness and outliers using INLA: application to aortic valve replacement surgery data

Ekong A.H. [a, *], Olayiwola O.M. [a], Dawodu G.A. [a] , Wale-Orojo, O.A.[a], Akintunde, A.A.[a] and Osinuga I.A. [b]

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

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

**Corresponding author email address*: anieekong@outlook.com.

Abstract

Joint modelling under Bayesian paradigm has gained a lot of traction especially with sampling-based estimation, however approximate Bayesian estimation of integrated Laplace approximation (INLA) is slowly gaining grounds. Prior specification has also been a recurring discuss in Bayesian analysis with prior sentivity becoming part of the data analysis process. This work presents joint modelling of longitudinal and cure proportion using latent Gaussian model with INLA and prior sensitivity analysis for the model in the presence of data value missingness and outliers. The approach assumed inverse-Wishart prior distribution for the covariance matrix of the random effects and Gaussian priors for the joint model fixed effects, while the penalised complexity prior was assumed for the Weibull shape parameters of the baseline hazard function. Four different prior specification settings were studied for fixed and random effects and the association parameter. The study was applied to aortic valve replacement surgery data to assess the effects of covariates on a biomarker and risk of event, with spline trajectories. The best prior setting was arrived at via the lowest values of DIC, WAIC and log marginal-likelihood and was Gaussian prior for fixed effects and association parameter each with (mean, precision) values as (0, 0.001), (0, 0.001), (0, 0.001), and parameters from Wishart distribution on the precision matrix for random effects as (100, 1) and it gave robust results with missing values and outliers. The posterior estimates from the best prior settings showed significant covariates on the biomarker and on the conditional failure time latency model. The study contributes to the literature on approximate Bayesian alternative to jointly modelling of longitudinal and mixture cure outcomes in the area of prior specification and data value missingness and outiers.

Keywords: prior specification, association structure, Laplace approximation, shared random effect, nonlinear trajectory

1. Introduction

Many experiments and trials give rise to opportunity for the collection of different types of dataset at the same time, for example longitudinal and survival datasets. These types of datasets are usually analysed separately. However, since the two datasets are collected form the same individuals simultaneously, analysing them separately can overlook latent association of the two components that could shed better light to the subject of inquiry. Consequently, conclusions may be biased and insufficient as a result of measurement error and missing data. Joint modelling has become a pervasive approach in analysing these two datasets as a way of remedying the separate analysis

(See Tsiatis and Davidian 2004). Joint modelling has been used in medical and health studies, engineering, finances etc. There are load of works in the literature on joint modelling with different types of models for the longitudinal part and survival part. Alsefri et al. (2020) gave a review of developments in Bayesian joint models covering articles published up to July 2019.

Classical survival analysis models such as the Cox proportional hazard (PH) model (Cox, 1972) and the accelerated failure time (AFT) (Kalbfleisch and Prentice, 2002) model are based on the assumption that given enough follow up time, every subject will eventually experience the event of interest or censored. In some real situations, there are cases in which some individuals will never experience the event of interest, even if the follow-up is indefinite. In the case of possibility of cured subjects, the population is assumed to be made up of two groups of subjects: the susceptible, who will one day experience the event of interest, and the cured, who may not experience the event of interest during the follow-up period. Cancer trials are also cases where there is a strong rationale for the existence of cured subjects because if the treatment is successful, the original cancer is removed and the subject will not experience recurrence of the disease. This is particularly true for patients in early cancer stages (Peng and Taylor, 2014). Cure models are particularly appropriate in cancer trials where there is scientific interest in factors associated with the probability of cure and factors associated with the time to recurrence for non-cured individuals.

Bayesian approach has seen more of Markov chain Monte Carlo (MCMC) for parameter estimation, for example, Chen et al. (2004) presented multiple longitudinal markers as well as a cure structure for the survival component based on the promotion time cure rate model with MCMC Gibbs sampling. Chi and Ibrahim (2007) used MCMC adaptive rejection algorithm and an extra Metropolis step was used for parameter estimation in joint modelling of multivariate longitudinal component and cure survival component. He and Luo (2016) employed MCMC in their shared random effects joint model of a multilevel item response theory model for the multiple longitudinal outcomes, and a Cox's proportional hazard model with piecewise constant baseline hazards for the event time data. Alafchi et al. (2021) proposed a two-stage base model for joint modelling of multivariate longitudinal and multistate process. Maximum likelihood estimation was used for fixed effects coefficients in longitudinal and multistate model and empirical Bayes methods for random effects coefficients in longitudinal.

Many studies in the literature report 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, van Niekerk, Krainski, Rue, et al., 2024). The approximate Bayesian approach, INLA, introduced by Rue et al. (2009) is gaining usage for joint modelling as an alternative to MCMC with the Rue et al. (2009) discussing the advantages of INLA over MCMC. We refer to van Niekerk et al. (2021), Medina-Olivares et al. (2023), Rustand et al. (2023) and Rustand et al. (2024), Rustand, van Niekerk, Krainski, and Rue (2024), Alvares et al. (2024), Ekong et al. (2025) for more instances of INLA's applicability and suitability. Lázaro et al. (2020) presented implementation 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. (2019) showed that a joint model with a linear bivariate Gaussian association structure is a latent Gaussian model (LGM) and thus can be implemented using most existing packages for LGMs especially R-INLA and van Niekerk, Bakka, and Rue (2021) 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. Rustand, van Niekerk, Krainski, Rue, et al. (2024) presented joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations algorithm implemented in the R package R-INLA. They compared the INLA method to existing alternatives (MCMC and MCEM) via simulations applied to five longitudinal markers and included competing risks of death and transplantation in application to clinical trial on primary biliary cholangitis. Ekong et al. (2025) studied joint modelling of longitudinal and cure survival outcomes for univariate biomarker considering nonlinear trajectories with application to renal transplantation data which comprised with glomerular filtration rate of kidneys as biomarker and survival event of time to graft failure.

This study builds on the presentation of Ekong et al. (2025) and Rue et al. (2009) by investigating the impact of different priors parameters values on the posterior distributions using the same model. Also, the impact of prior specification for the joint model presented is investigated in the presence of data value missingness and outliers. We visualise the sensitivity of the joint model to prior specifications and assess the model stability and reliability. The main goal here is to see how the marginal posterior effects, missing values and outliers are impacted by the Gaussian prior for the fixed effects and association parameters, as well as those of the Inverse Wishart prior for the multivariate random effects. We considered application to real dataset of aortic valve replacement surgery from an observational study by Lim et al. (2008), on detecting effects of different heart valves, differing on type of tissue, implanted in the aortic position.

2. Methodology

Following the presentation of Ekong et al. (2025), 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 $f(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\beta};\boldsymbol{U})$. 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 η^L and hyperparameters θ^L , hence a structured additive model for the longitudinal component is given as follows:

$$g^{-1}\{E(y_i|\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{U})\} = \eta^L = \beta_0 + \boldsymbol{\beta X} + \sum_{i=1}^{n} f(u_i) + \boldsymbol{b}_i u_i + \boldsymbol{\epsilon} \quad (1)$$

where $f(u_i)$ is the latent random effect of covariate u_i for *i*-th patient, which could be spatial effects, temporal effects, patient or group-specific intercepts. $\boldsymbol{\beta}$ represent the linear fixed effects of the covariates $\boldsymbol{X}$, $\boldsymbol{b}_i$ is the vector of random effects of intercept and slope, where β_0 plus b_i 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 $f(u_i)$ which we take as cubic splines with internal knots at 1 and 4 years to account for nonlinear trajectories of the longitudinal outcome, $\boldsymbol{\epsilon}$ is the unstructured random effects.

2.2 Cure survival model component

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

$$S_u(t) = P(T_{im} > t \,|Z_i = 0)$$

$$S_c(t) = P(T_{im} > t \,|Z_i = 1) = 1$$

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

$$S(t) = P(T_{im} > t) = \pi + (1 - \pi)S_u(t) \quad (2)$$

Cure fraction π 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, 2014).

Covariates in the incidence model

Note that for a patient who has experienced the event ($\delta_i = 1$), we know $Z_i = 1$, but for a censored patient ($\delta_i = 0$), we do not observed Z_i, hence, 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

$$\text{logit}[\pi\,(\boldsymbol{\beta}_1)] = \boldsymbol{\beta}_1'\boldsymbol{x}_1 \equiv \pi(\beta_1) = \frac{\exp\{\boldsymbol{\beta}_1'\boldsymbol{x}_1\}}{1 + \exp\{\boldsymbol{\beta}_1'\boldsymbol{x}_1\}} \quad (3)$$

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

Covariates in the latency

For patients with $Z_i = 1$, the time to event is assumed to follow a parametric distribution. The Cox proportional hazards model is usually formulated in terms of the hazard function for the event time as

$$h_u(t\,|h_{u0}, \boldsymbol{\beta}_2) = \lim_{\Delta t \to \infty} \frac{P(t \le T_{im} < t + \Delta t\,|T \ge t)}{\Delta t} = h_{u0}(t)\exp\{\boldsymbol{\beta}_2' \boldsymbol{x}_2\} \quad (4)$$

where $h_{u0}(t)$ is the baseline hazard function that determines the shape of the hazard function. Model (4) can also be presented in terms of the survival function of T_{im} as

$$S_u(t|S_{u0}, \boldsymbol{\beta}_2) = [S_{u0}(t)]^{exp\{\boldsymbol{\beta}_2' \boldsymbol{x}_2\}} \quad (5)$$

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

2.3 Joint Model of Multivariate Longitudinal and Cure Survival Outcomes as LGMs

The modelling approach assumes a logistic distribution for the probability of cure in the incidence model in (3) and the Cox proportional hazard (4) for the survival time with a Weibull baseline hazard function $h_{u0}(t|\lambda,\ \alpha) = \lambda\alpha t^{\alpha-1}$ with λ and α as the scale and shape parameters respectively. γ is the association parameter estimating the strength of association between the survival and the longitudinal component, thus we define

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

The linear predictors of the joint model becomes

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

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

Here γ as a smooth function facilitates the joint estimation of the models associating the longitudinal trajectories and mixture cure process using the entire longitudinal predictors as shared random effect where each random effect's individual deviation is associated to an association parameter in the survival latency component.

2.3.1 Likelihood function of Joint Model

The likelihood of the longitudinal outcomes given the parameters $\boldsymbol{\beta}_0, \boldsymbol{\beta}, \theta^L, \boldsymbol{\eta}^L, f(\cdot), \boldsymbol{b}_i$ and $\boldsymbol{\epsilon}$ can be given as

$$\mathcal{L}^L(\boldsymbol{y}|\boldsymbol{\eta}^L) = \prod_{i=1}^{N_L} g\left(y_{im}\middle|\boldsymbol{\eta}_i^L(t)\right) \tag{6}$$

Given survival observations $\boldsymbol{d} = \{T_{im}, \delta_{im}, z_{im}\}$ and parameter vector $\boldsymbol{R} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \alpha, \lambda, \boldsymbol{\eta}^S, \theta^S, \gamma)$, the likelihood for the mixture cure survival becomes

$$\mathcal{L}^S(\boldsymbol{d}|\boldsymbol{R}) = \prod_{i=1}^{N} \mathcal{L}_i(\boldsymbol{R}|d) = \prod_{i=1}^{N} \eta_i^S \boldsymbol{R}^{z_i} \left(1 - \eta_i^S \boldsymbol{R}\right)^{1-z_i} h_{iu}(t_i|\boldsymbol{R})^{\delta_i(1-z_i)} S_{iu}(t_i|\boldsymbol{R})^{1-z_i} \tag{7}$$

The complete likelihood becomes

$$p(\boldsymbol{D}_i|\boldsymbol{\chi}_i, \boldsymbol{\theta}) = \int_{\boldsymbol{b}_i} \left[\prod_{i=1}^{N} \mathcal{L}_i(\boldsymbol{R}|d) \prod_{i=1}^{N_L} g\left(y_{im}\middle|\boldsymbol{\eta}_i^L(t)\right) p(\boldsymbol{b}_i)\right] d\boldsymbol{b}_i \tag{8}$$

where we define the latent field $\boldsymbol{\chi} = (\boldsymbol{\beta}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\eta}^S, \boldsymbol{\eta}^L, f(\cdot), \boldsymbol{b}_i, \lambda, \boldsymbol{\epsilon})$ and a vector of hyperparameters $\boldsymbol{\theta} = \left(\theta^L, \theta^S, \alpha, \tau^{-1}, \gamma_p\right)$. The aim is to present equation (8) as LGMs by showing its specific hierarchical structure. The first level of the hierarchy involves presenting the likelihood function given the latent field $\boldsymbol{\chi}$ and the vector of hyperparameters $\boldsymbol{\theta}$ as shown in equation (8).

2.3.2 Prior Specification settings

The next level of the hierarchy involves the conditional distribution of the latent field $\boldsymbol{\chi}$ 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 $\boldsymbol{Q}(\boldsymbol{\theta}_2)$, i.e. $\boldsymbol{\chi} \sim MVN(\boldsymbol{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2))$, this is given as

$$p(\boldsymbol{\chi}|\boldsymbol{\theta}) = (2\pi)^n |\boldsymbol{Q}(\boldsymbol{\theta}_2)|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\chi}'\boldsymbol{Q}(\boldsymbol{\theta}_2)\boldsymbol{\chi}\right)$$

Then at the final level of the hierarchy, a prior on the hyperparameter vector $p(\boldsymbol{\theta})$ can then be formulated for the set of hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, which could be non-normal. This enables us to assume normal prior for the vector of fixed effects for the longitudinal variable as $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau_i \boldsymbol{I})$, where $\boldsymbol{\beta} \in \boldsymbol{\chi}$ and $\tau_i \in \boldsymbol{\theta}$. To complete the model specification, we assume the inverse-Wishart prior distribution for the covariance matrix of the random effects and Gaussian priors for the fixed effects, while the penalised complexity prior PC(5) is assumed for the Weibull shape parameters of the baseline hazard function.

For the fixed effects and association parameter, the Gaussian distribution has density with mean μ and precision τ

$$\pi(\theta) = \left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{\tau}{2}(\theta - \mu)^2\right)$$

for continuous θ, with precision τ and mean μ.

For random effects, given vector of random effects (b_i, d_i) that are iid bivariate Normals $\begin{pmatrix} b_i \\ d_i \end{pmatrix} \sim N(\mathbf{0}\ ,\ \boldsymbol{W}^{-1})$; $\boldsymbol{W}^{-1} = \begin{pmatrix} 1/\tau_a & \rho/\sqrt{\tau_a \tau_b} \\ \rho/\sqrt{\tau_a \tau_b} & 1/\tau_b \end{pmatrix}$, with $\boldsymbol{W}^{-1}$ as covariance matrix, τ_a, τ_b and ρ are hyperparameters, with ρ as correlation, τ_a and τ_b are the marginal precisions.

The precision matrix $\boldsymbol{W}$ is Wishart distributed

$$\boldsymbol{W} \sim \text{Wishart}_p(r, \boldsymbol{G}^{-1})$$

with density

$$\pi(\boldsymbol{W}) = c^{-1}|\boldsymbol{W}|^{(r-(p+1))/2} \exp\left\{-\frac{1}{2}\text{Trace}(\boldsymbol{W}\boldsymbol{G})\right\},\ \ r > p+1$$

and

$$c = 2^{(rp)/2}|\boldsymbol{G}|^{-r/2}\pi^{(p(p-1))/4}\prod_{j=1}^{p}\Gamma\big((r+1-j)/2\big).$$

We consider the first prior settings for the fixed effects parameters, mean and precision, mean intercept and precision intercepts and association parameter (mean and precision) respectively as (0, 0.01), (0, 0.01), (0, 0.01), and (10, 1) for the prior on the precision of matrix which follows a Wishart distribution for the random effects. The second prior settings is respectively (0, 0.16), (0, 0.16), (0, 0.16) and (10, 1) for the prior of random eff the prior on the precision of matrix of random effects. The third prior settings is (0, 0.16), (0, 0.16), (0, 0.16) and (100, 1) for the prior of random eff the prior on the precision of matrix of random effects. The forth settings is (0, 0.001), (0, 0.001), (0, 0.001) and (100, 1) for the prior of random eff the prior on the precision of matrix of random effects.

2.3.3 Posterior Estimation using Integrated Laplace Approximation

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

$$p(\boldsymbol{\chi}, \boldsymbol{\theta}|\boldsymbol{D}) \propto\ p(\boldsymbol{\theta})p(\boldsymbol{\chi}|\boldsymbol{\theta})\prod_{\boldsymbol{i}} p(\boldsymbol{D}_i|\boldsymbol{\chi}, \boldsymbol{\theta})$$

$$\propto\ p(\boldsymbol{\theta})|\boldsymbol{Q}(\boldsymbol{\theta}_2)|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\chi}'\boldsymbol{Q}(\boldsymbol{\theta}_2)\boldsymbol{\chi} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \log(\boldsymbol{D}_i|\boldsymbol{\chi}_i, \boldsymbol{\theta})\right) \qquad (9)$$

Within this framework the joint posterior density (9) and subsequently the marginal posterior densities, $p(\boldsymbol{\chi}_i|\boldsymbol{D})$; $i = 1\ , \ldots , n$ and $p(\boldsymbol{\theta}|\boldsymbol{D})$ can be efficiently and accurately calculated using the integrated Laplace approximation methodology developed by Rue et al. (2009). The marginal posterior densities becomes

$$p(\boldsymbol{\chi}_i|\boldsymbol{D}) = \int p(\boldsymbol{\chi}_i, \boldsymbol{\theta}|\boldsymbol{D})\, d\boldsymbol{\theta} = \int p(\boldsymbol{\chi}_i, \boldsymbol{\theta}|\boldsymbol{D})p(\boldsymbol{\theta}|\boldsymbol{D})\, d\boldsymbol{\theta}$$

and

$$p(\theta_i|\boldsymbol{D}) = \int p(\boldsymbol{\theta}|\boldsymbol{D})\, d\boldsymbol{\theta}_{-\boldsymbol{j}}$$

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

$$p(\boldsymbol{\theta}_i, \boldsymbol{\chi}_i|\boldsymbol{D}_i) = \frac{p(\boldsymbol{D}_i|\boldsymbol{\theta}_i, \boldsymbol{\chi}_i)p(\boldsymbol{\theta}_i, \boldsymbol{\chi}_i)}{p(\boldsymbol{D}_i)} \propto p(\boldsymbol{D}_i|\boldsymbol{\theta}_i, \boldsymbol{\chi}_i)p(\boldsymbol{\chi}_i|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)$$

where $p(\boldsymbol{\chi}_i|\boldsymbol{\theta}_i)$ and $p(\boldsymbol{\theta}_i)$ are prior distributions and the focus is on approximating the multidimensional integral from the marginal likelihood $p(\boldsymbol{D}_i|\boldsymbol{\theta}_i, \boldsymbol{\chi}_i)$ 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, van Niekerk, Krainski, Rue, et al. 2024).

We consider the Laplace transformation using a second-order Taylor series expansion for the integral of the density function $p(\boldsymbol{\chi})$ by taking the form of (Blangiardo and Cameletti, 2015)

$$\int_{-\infty}^{\infty} p(\boldsymbol{\chi})\, d\boldsymbol{\chi} = \int_{-\infty}^{\infty} \exp(\log p(\boldsymbol{\chi}))\, d\boldsymbol{\chi} = \int_{-\infty}^{\infty} \exp\big(g(\boldsymbol{\chi})\big)\, d\boldsymbol{\chi} \qquad (10)$$

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

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

$$\boldsymbol{\chi}^* = \text{argmax}_{\,\boldsymbol{\chi}} g(\boldsymbol{\chi}),$$

then

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

for $g(\boldsymbol{\chi})$ to be approximated as

$$g(\boldsymbol{\chi}) \approx g(\boldsymbol{\chi}^*) + 0.5(\boldsymbol{\chi} - \boldsymbol{\chi}^*)'\mathbf{H}g(\boldsymbol{\chi}^*)(\boldsymbol{\chi} - \boldsymbol{\chi}^*)$$

where $\mathbf{H}g(\boldsymbol{\chi}^*)$ is the Hessian of $g(\boldsymbol{\chi}^*)$, and equation (10) can be written as

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

$$= \exp\big(g(\boldsymbol{\chi}^*)\big) \int_{-\infty}^{\infty} \exp\big(0.5(\boldsymbol{\chi} - \boldsymbol{\chi}^*)'\mathbf{H}g(\boldsymbol{\chi}^*)(\boldsymbol{\chi} - \boldsymbol{\chi}^*)\big)\, d\boldsymbol{\chi}$$

$$= \exp\big(g(\boldsymbol{\chi}^*)\big) \int_{-\infty}^{\infty} \exp\big(-0.5(\boldsymbol{\chi} - \boldsymbol{\chi}^*)'\{-\mathbf{H}g(\boldsymbol{\chi}^*)\}(\boldsymbol{\chi} - \boldsymbol{\chi}^*)\big)\, d\boldsymbol{\chi}$$

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

$$\int_{-\infty}^{\infty} (2\pi)^{-\frac{nm}{2}}|\mathbf{H}g(\boldsymbol{\chi}^*)|^{-\frac{1}{2}} \exp\big(-0.5(\boldsymbol{\chi} - \boldsymbol{\chi}^*)'\{-\mathbf{H}g(\boldsymbol{\chi}^*)\}(\boldsymbol{\chi} - \boldsymbol{\chi}^*)\big)\, d\boldsymbol{\chi}$$

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

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

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

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

The conditional posterior distribution of $p(\boldsymbol{\chi}, \boldsymbol{\theta}|\boldsymbol{D})$ is defined from the joint posterior distribution in Equation (9) as

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

which can be rewritten as, ignoring elements with $\boldsymbol{\chi}$

$$p(\boldsymbol{\chi}|\boldsymbol{\theta}, \boldsymbol{D}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\chi}'\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{\chi} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} g_i(\boldsymbol{\chi}_i)\right) \tag{11}$$

Refer to Rue et al. (2009) and Ekong et al. (2025) for more details on the approximation procedures, such as Gaussian, Laplace and simplified Laplace approximations, as well as further discussions on the approximations error and its asymptotics with practical issues.

3 Results and discussion

3.1 Descriptive analyses of the aortic valve replacement surgery data

The aortic valve replacement surgery data is an observational study on detecting effects of different heart valves, differing on type of tissue, implanted in the aortic position carried out by Lim et al. (2008). The data consists of 300 patients who underwent aortic valve replacement from 1991 to 2001 with at least a year of follow-up with a total of 1,273 serial echocardiographic measurements. Patients with two or more procedures were censored from the time point of the second procedure to ensure that they were analysed only once. Demographic, operative, and mortality data were obtained from individual hospital notes, death certificates, and autopsy reports. Details of the dataset can be found in Lim et al. (2008).

The version of the aortic valve replacement surgery data (n = 256) used in this study was obtained from the R package joineRML (Hickey et al., 2018) and for the sake of comparison of results we selected variables used in the analysis in Lim et al. (2008) for our own analysis and they include *hs*, the type of implanted aortic prosthesis: Homograft or Stentless valve; *sex*, gender of patient (0 = Male and 1 = Female); *time*, observed time point, with surgery date as the time origin (years); *fuyrs*, maximum follow up time, with surgery date as the time origin (years); *status*, censoring indicator (1 = died and 0 = lost at follow up); *size*, size of the valve (millimeters); *lv*, preoperative left ventricular ejection fraction (1 = good, 2 = moderate and 3 = poor); *grad*, valve gradient at follow-up visit; *lvmi*, left ventricular mass index (standardised) at follow-up visit; and *ef*, ejection fraction at follow-up visit. The longitudinal biomarker of interest is *grad*. The data has biomarker, *grad*, with 359 missing values (36.3% missing) and about 64 outliers and these are represented in Fig. 1 showing the plot of percentage of missing values and boxplot including outliers. The dataset with missing and outlier values is denoted as Data A.

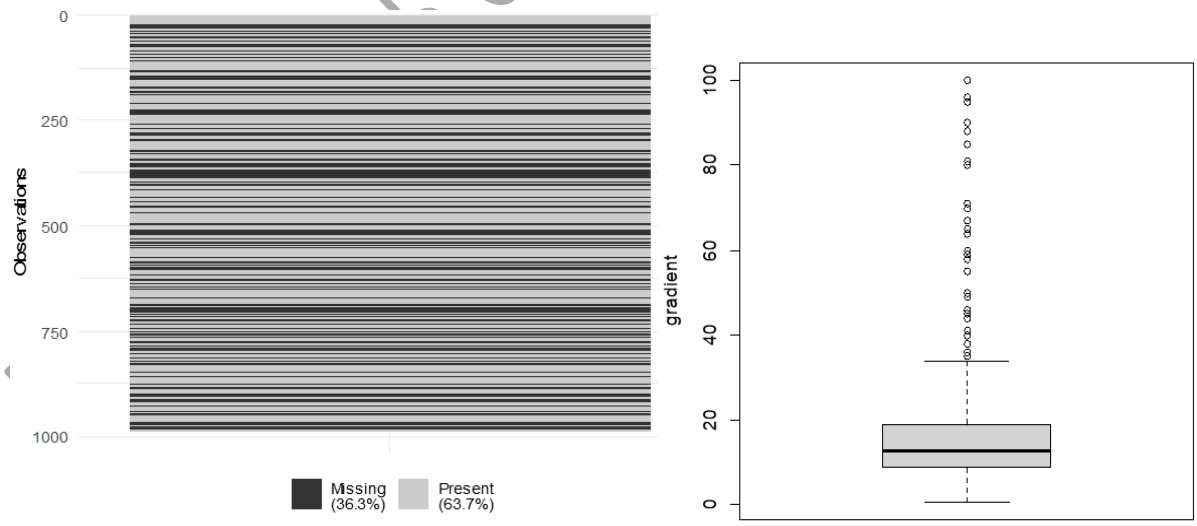

Fig. 1 Plot of percentage of missing values and boxplot showing outliers

In order to treat the missing and outlier values in the dataset, we apply data imputation by replacing the missing values with the mean of the dataset without the outliers and replacing the outliers with the inter-quantile range by method of capping. Hence, the dataset without missing values and outliers is also used in the analysis and is denoted as Data B.

We examined the survival curves for the two covariates of type of implanted aortic prosthesis received and gender of patient, *hs* and *sex*. From Fig. 2 which shows the different survival curves estimated by type of prosthesis, it can be observed that survival seemed to lower quickly for patients who received stentless valve replacement before flattening at 50% rate at less than two years after treatment. The patients with homograft valve replacement showed slower decrease in survival rate before flattening at aver 70% at less than two years after treatment. This information suggest the possibility of cure fraction for some patients who may not observe the failure event of death at the end of the follow-up period and or beyond.

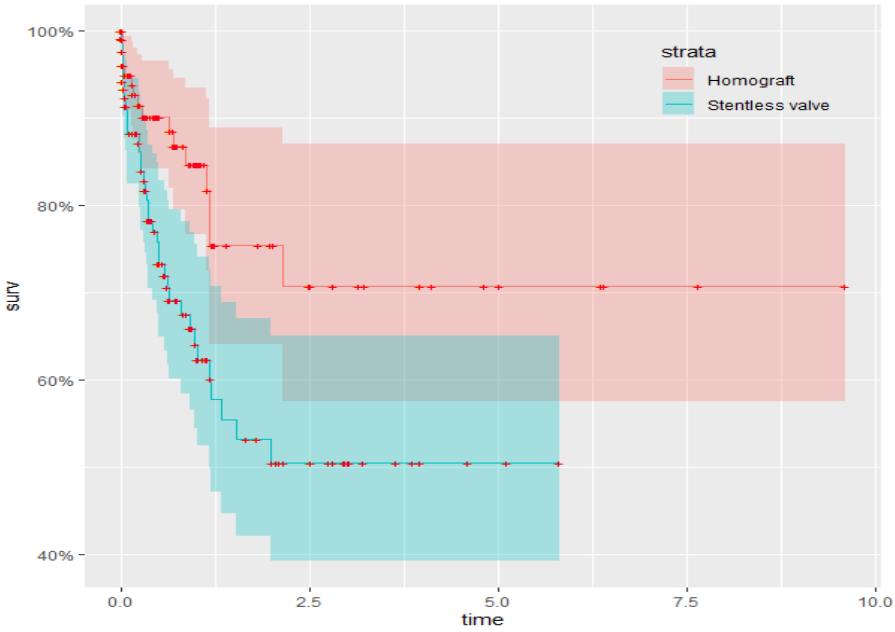

Fig. 2 Survival estimates of patients according to treatment covariate *hs*

3.2 Joint cure modelling of aortic valve replacement surgery data

Our approach here is to fit the joint model for the aortic valve replacement surgery data using the LGM framework as implemented with INLA by considering the possibilities of cure proportion in the survival component of the modelling. The modelling formulation as described in Section 2 is presented for the dataset thus.

Longitudinal component with spline trajectories:

$$\text{grad}_i(t) = (\beta_{10} + b_{i10}) + (b_{i11} + \beta_{1time})\text{time}_i F_1(t) + \beta_{1hs}\text{hs}_i + \beta_{1sex}\text{sex}_i + \beta_{1size}\text{size}_i + \beta_{1lv1}\text{lv}_i + \beta_{1lv2}\text{lv}_i + \varepsilon_{i1}(t) = \eta_{i1}(t) + \varepsilon_i(t)$$

Latency model component:

$$h_i(t\,|h_0) = (\lambda\alpha t^{\alpha-1})\exp\{\gamma_{\text{grad}_S}(\eta_{i1}(t)) + (\beta_{2hs}\text{hs}_i + \beta_{2sex}\text{sex}_i + \beta_{2size}\text{size}_i + \beta_{2lv1}\text{lv}_i + \beta_{2lv2}\text{lv}_i)\}$$

Incidence model component:

$$\text{logit}[\pi_i] = \beta_{30} + \beta_{3hs}\text{hs}_i + \beta_{3sex}\text{sex}_i + \beta_{3size}\text{size}_i + \beta_{3lv1}\text{lv}_i + \beta_{3lv2}\text{lv}_i$$

where $\varepsilon_i(t)$ are independent Gaussian measurement errors for the longitudinal outcome, $F_1(t)$ is natural cubic spline basis functions with internal knots at 2 years. For priors specification, as noted earlier, we assume a Gaussian prior for the latent field $\boldsymbol{\chi}$ with precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ conditioned on $\boldsymbol{\theta}$, which we also assume prior distributions $p(\boldsymbol{\theta})$ 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 parameter, α, is assumed to follow the penalised complexity prior PC(5). The inverse-Wishart prior is assumed for the covariance matrix of the random effects and Gaussian priors for all the fixed effects.

For the sensitivity analysis, we consider the first prior settings, the Gaussian prior for the fixed effects parameters, mean and precision, mean intercept and precision intercepts and the association parameter, γ_{grad_S}, (mean and precision) respectively as (0, 0.01), (0, 0.01), (0, 0.01), and the hyperparameters values in the inverse-Wishart prior are (10, 1) on the precision of matrix of the Wishart distribution for the random effects. The second prior settings is respectively (0, 0.16), (0, 0.16), (0, 0.16) and (10, 1) fixed effects parameters, mean and precision, mean intercept and precision intercepts, the association parameter and the prior of random effects precision matrix. The third prior settings follows the same order as (0, 0.16), (0, 0.16), (0, 0.16) and (100, 1) for the prior of random eff the prior on the precision of matrix of random effects. Similarly, the forth settings is (0, 0.001), (0, 0.001), (0, 0.001) and (100, 1) for the prior of random eff the prior on the precision of matrix of random effects. The posterior distributions for the different prior settings are used in the joint model for the data with missing values and outliers (Data A) and the data without missing values and outliers (Data B) and the plots of the posteriors are presented in what follows for random effects and association parameter and we also compared the models under the two data scenarios and prior settings using Bayesian model diagnostics to see the best situation and how sensitive the joint model approach is to priors, missing values and outliers in datasets.

The results are represented via plots of the prior distributions against the posterior distributions for all the prior settings for the random effects and the association parameters and are shown in Fig. 3 to Fig. 8. Starting with Fig. 3, we saw that the first prior setting was not sensitive to the missing values and outliers for the random intercept and longitudinal trajectory terms but had impact on their interaction term, meaning that the missing value and outliers influenced their posteriors for this prior setting. Hence, this prior for the joint model was robust to the effect of missing values and outliers in datasets. Prior two setting showed more robustness to sensitivity, missing values and outliers in data as both posteriors of the random effects were similar for Data A and Data B as

shown in Fig. 4, except for the intercept random effect term. The case was even most encouraging for prior three and four, shown in Fig. 5 and Fig. 6, as these priors had almost no impact on the marginal posteriors of the random effects parameters for Data A and Data B, meaning that missing values and outliers do not influence our model given this prior settings, thus supporting its robustness.

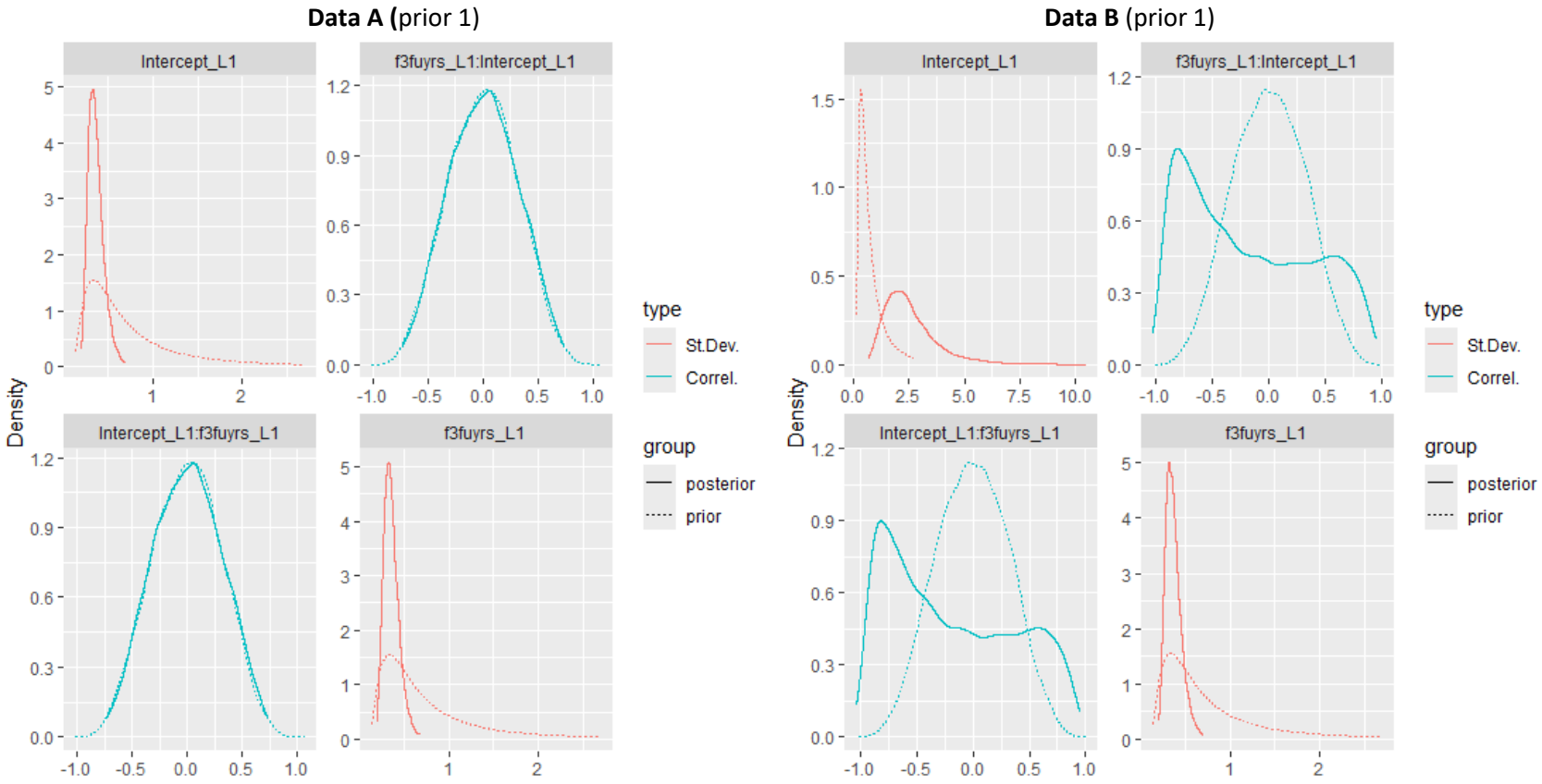

Fig. 3 Comparison of posterior distributions of random effects for Data A and Data B with prior setting one

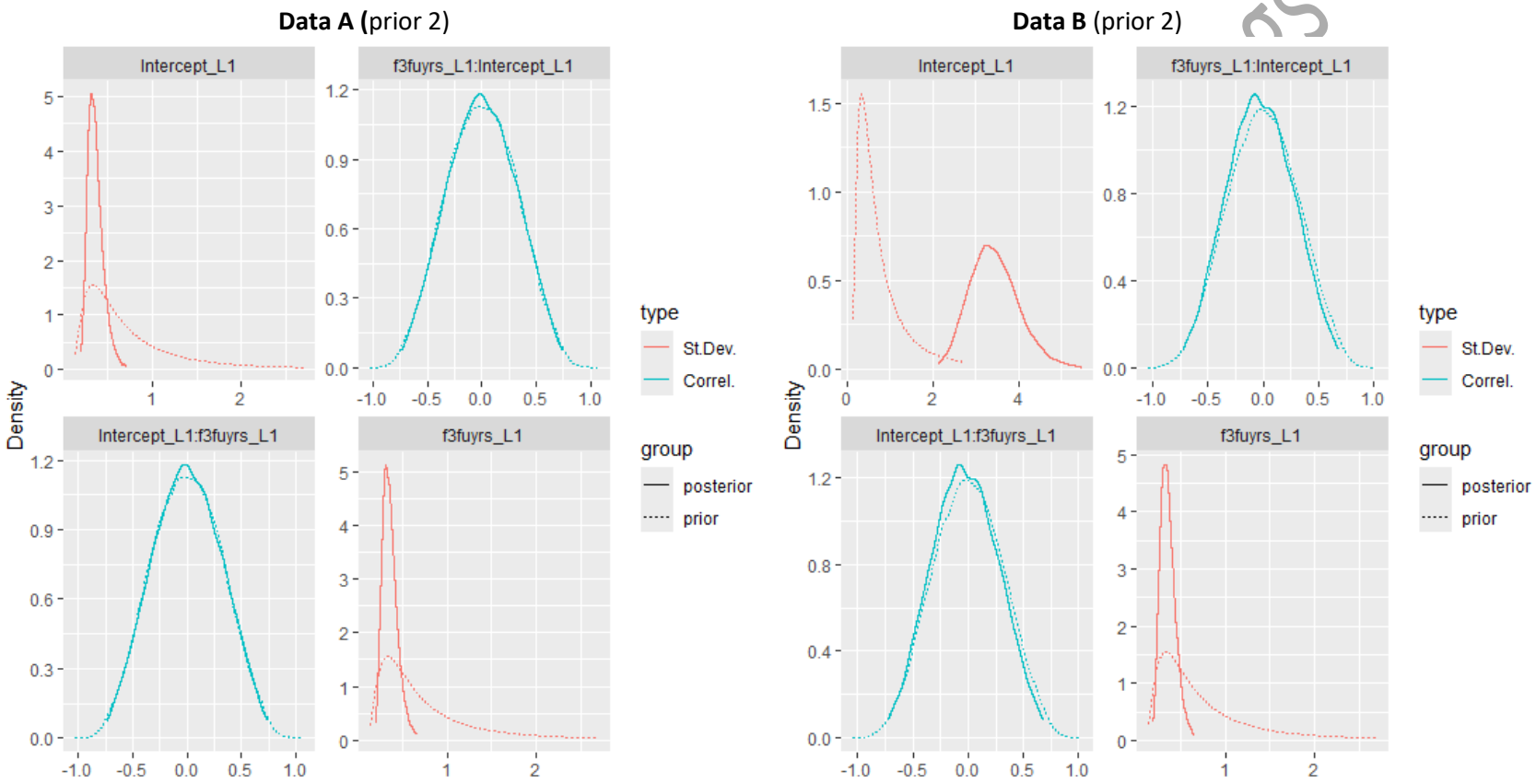

Fig. 4 Comparison of posterior distributions of random effects for Data A and Data B with prior setting two

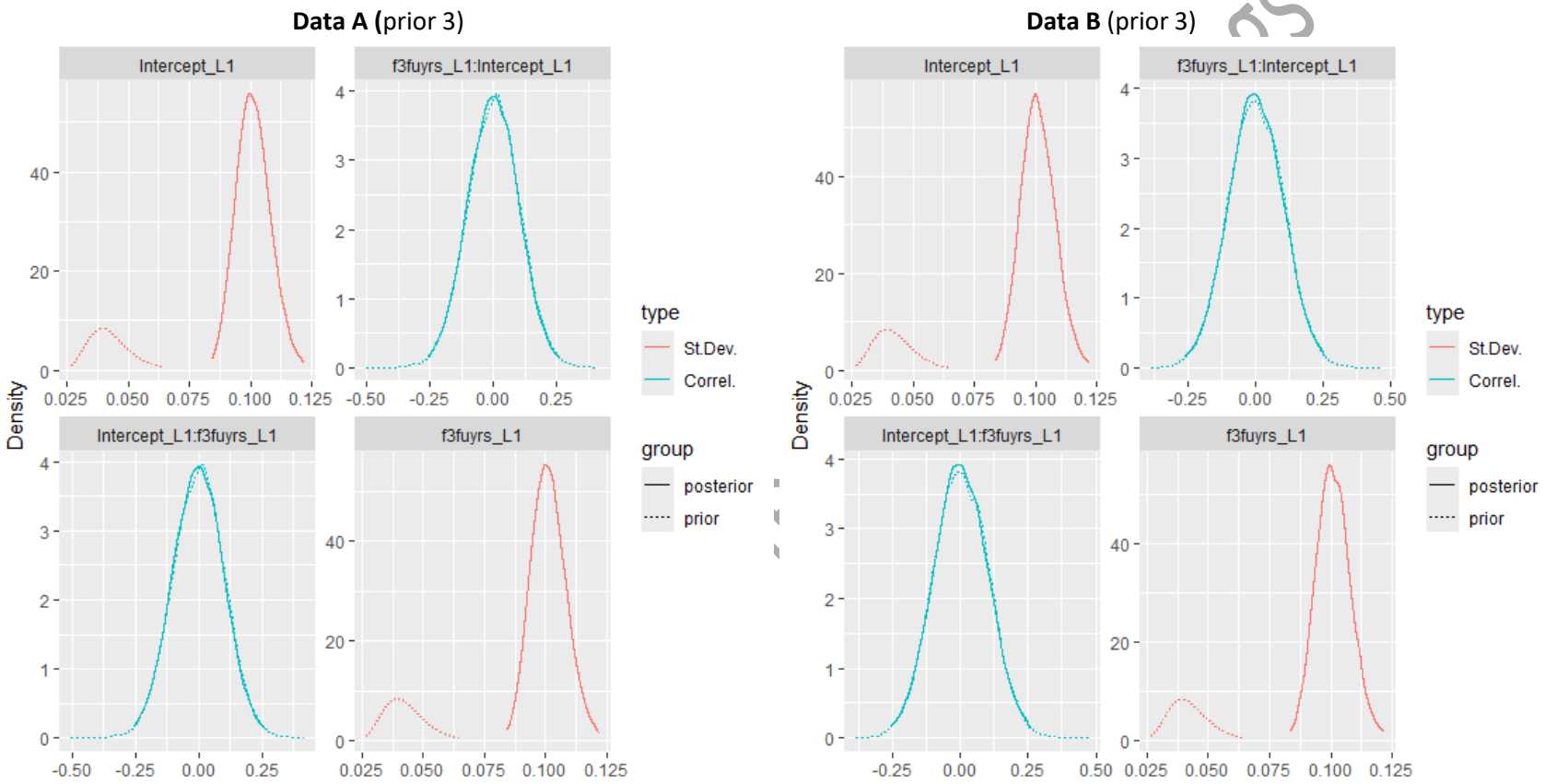

Fig. 5 Comparison of posterior distributions of random effects for Data A and Data B with prior setting three

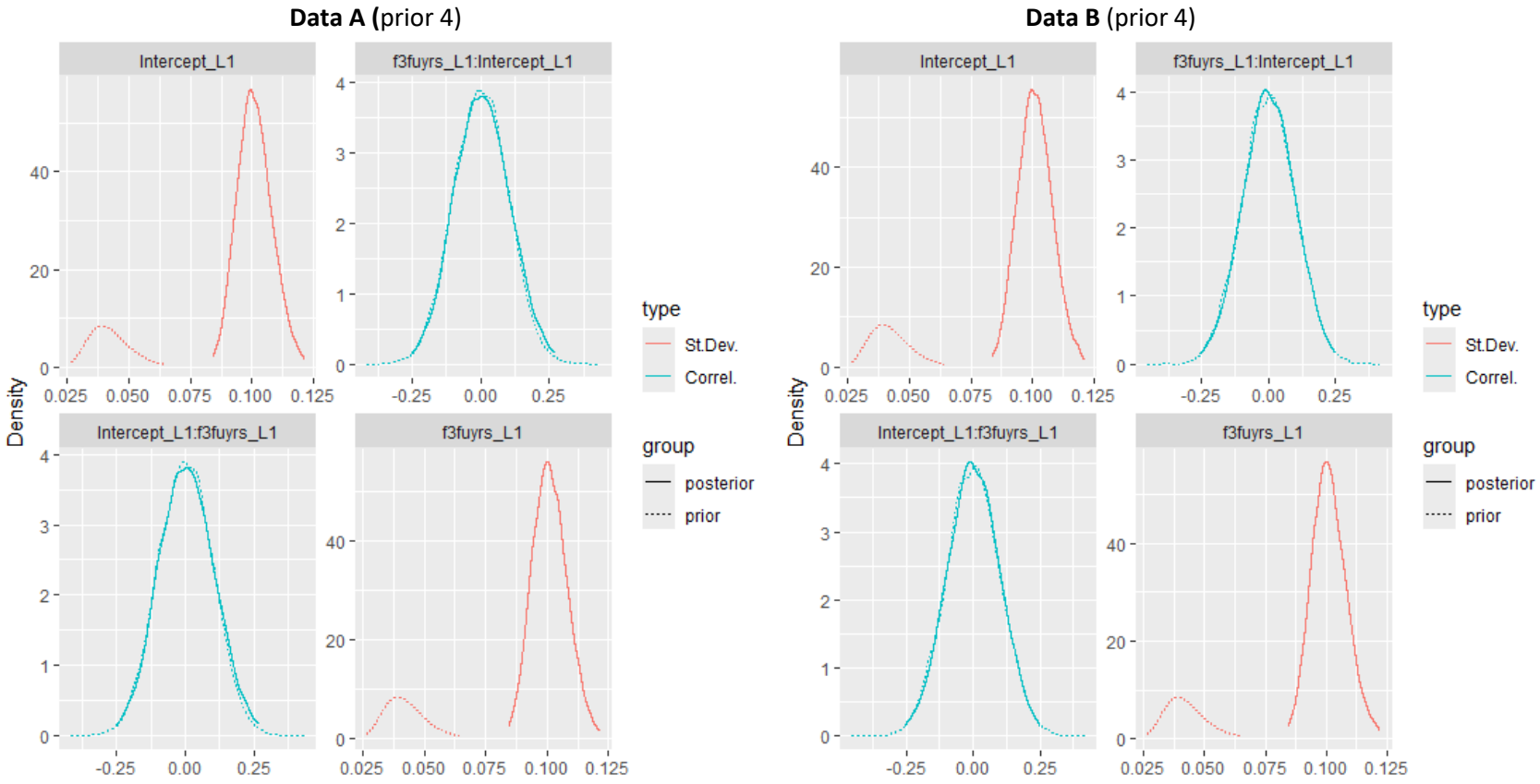

Fig. 6 Comparison of posterior distributions of random effects for Data A and Data B with prior setting four

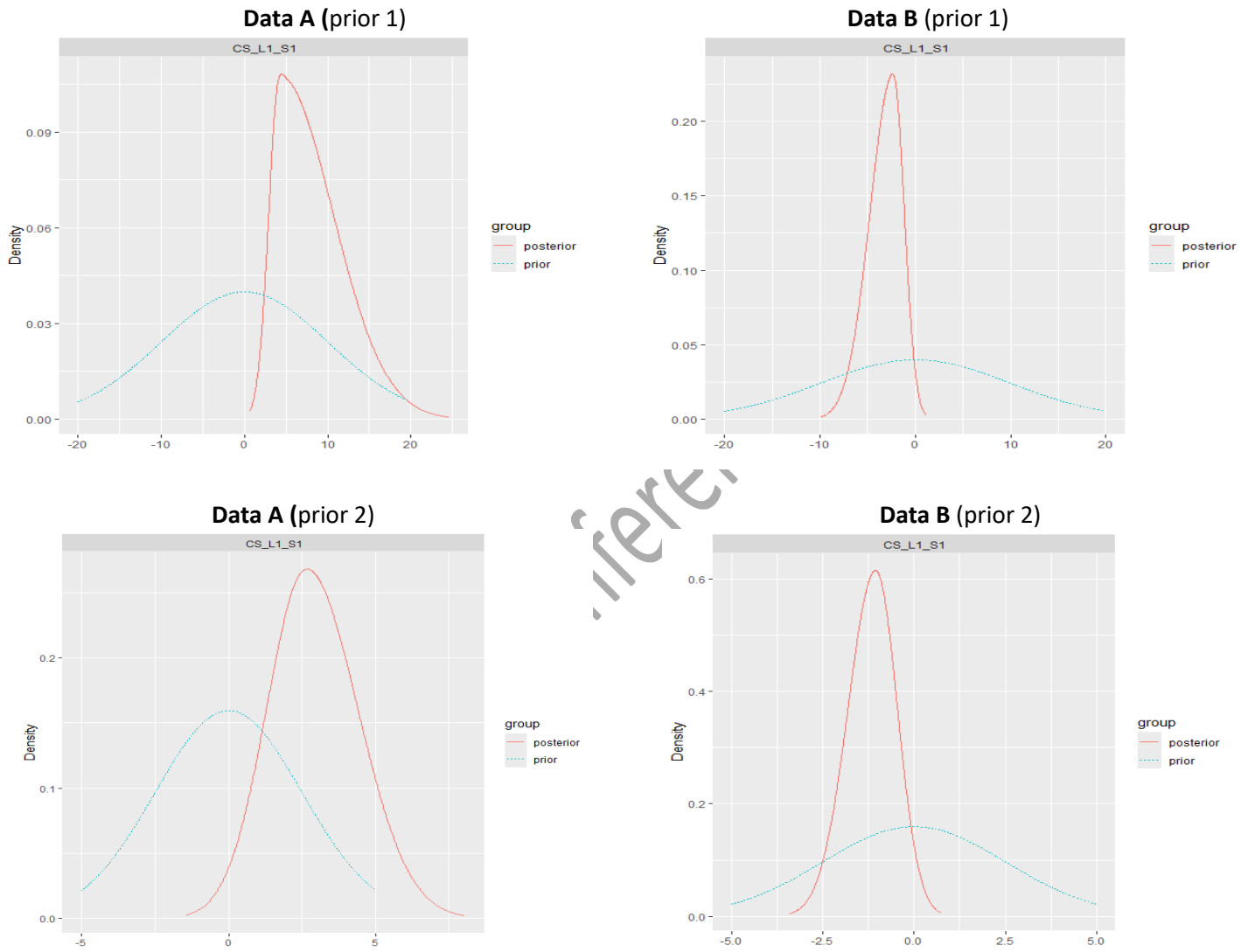

Fig. 7 Comparison of posterior distributions of association parameter for prior settings one and two with Data A and Data B

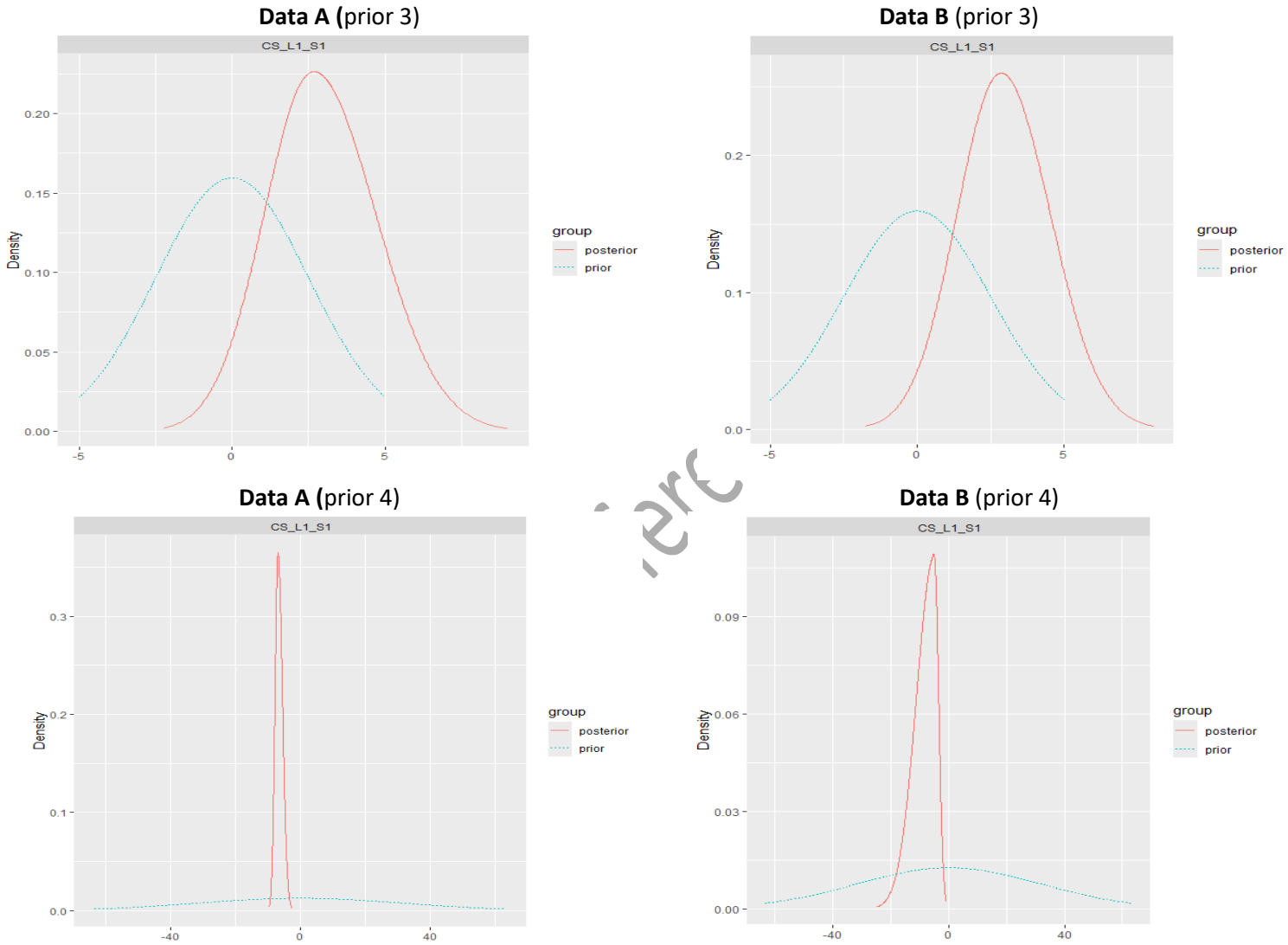

Fig. 8 Comparison of posterior distributions of association parameter for prior settings three and four with Data A and Data B

For marginal posterior of the association parameter for the four prior settings in Data A and Data B, we saw from Fig. 7 that prior settings one and two had different impacts on the marginal posteriors of the association parameter in the data with missing values and outliers against data without them, meaning that these priors were not robust to the missing values and outliers in estimating the association term. However, prior settings three and four showed similar posterior distributions for the association parameter in both Data A and Data B as seen in Fig. 8, meaning these priors had limited impact on these posteriors and showed robustness to missing values and outliers as in the case of the random effects parameters.

Table 1: Comparison of different Prior specifications on the two datasets

Data A.	*Prior 1*	*Prior 2*	*Prior 3*	*Prior 4*
DIC	1349.625	1357.548	1357.909	1345.592
WAIC	1189.869	1197.313	1197.667	1186.181
Loglike	-4902.484	-4894.183	-4894.039	-4914.956
Associa.	8.2788	2.9589	3.0018	-6.4966
Data B.	*Prior 1*	*Prior 2*	*Prior 3*	*Prior 4*
DIC	3543.856	3542.586	3613.034	3590.417
WAIC	3385.403	3384.444	3452.953	3430.888
Loglike	-6028.639	-6024.494	-6024.549	-6039.454
Associa.	-3.3233	-1.1818	2.9959	-8.6742

In Table 1 we see that all the model comparison diagnostics are in favour of the joint model with Data A which included missing values and outliers, since their values for all four prior settings were lower than those of Data B without missing values and outliers, and this results shows that the joint modelling approach is robust to missing values and outliers, while they are not sensitive to prior specification given missingness and outliers. However, we also observed from the results for Data A that the fourth prior specification was best, since it had the lowest values of these diagnostics. Whereas in the case of Data B, prior settings two was preferred before prior setting one, then prior setting four and three in that order. We also looked at the association parameter that measures the strength of association between the longitudinal and cure survival components and we saw that for Data A, prior settings one and four reported a strong association between both components, howbeit in different directions. Prior setting four reported a strong negative association for Data B, and hence the fourth prior setting showed strong association notwithstanding the missingness and outliers in data and this favoured association value indicated that higher values of the biomarker, the implanted valve gradient, at follow-up were pointer to lower risk of the event happening.

The output of the joint modelling approach using the fourth prior settings, being the best prior specification reported in Table 1, with the mean, standard deviation and 95% credible interval, is

presented in Table 2. The results showed that the random effect was not significant for the valve gradient over time, whereas the significant predictors for valve gradient were gender, the type of implanted aortic prosthesis and preoperative left ventricular ejection fraction. The valve gradient was higher in female patients with male reference, decreased in patients with moderate preoperative left ventricular ejection fraction and increased in patients with poor preoperative left ventricular ejection fraction both with good preoperative left ventricular ejection fraction as reference. It can be seen that the spline function was significant in the time effect in capturing the longitudinal profile of valve gradient.

Table 2: Posterior mean, standard deviation and 95% Confidence intervals of aortic valve replacement surgery data from fourth prior settings

Fixed	mean	sd	Low 95% CI	Up 95% CI
β_{10}	36.685	9.015	19.016	54.355
β_{1time}	-3.877	1.303	-6.431	-1.324
β_{1hs}	3.826	1.803	0.292	7.359
β_{1sex}	2.281	1.590	-0.836	5.397
β_{1size}	-0.776	0.377	-1.516	-0.036
β_{1lv1}	-1.040	1.528	-4.035	1.955
β_{1lv2}	3.034	2.606	-2.074	8.143
σ_{e1}	302.571	16.942	270.802	337.375
Random				
σ^2_{b10}	0.010	0.002	0.008	0.014
σ^2_{b11}	0.010	0.002	0.008	0.014
$cov_{b10,b11}$	0.000	0.001	-0.002	0.002
Latency				
α	0.779	0.087	0.626	0.969
λ	0.020	0.083	-0.087	0.085
β_{2hs}	0.836	0.374	0.103	1.570
β_{2sex}	0.305	0.323	-0.329	0.939
β_{2size}	-0.017	0.075	-0.164	0.131
β_{2lv1}	0.240	0.301	-0.350	0.829
β_{2lv2}	0.553	0.411	-0.253	1.359
Incidence				
β_{30}	0.145	1.698	-3.185	3.475
β_{3hs}	-0.177	0.339	-0.841	0.487
β_{3sex}	-0.508	0.310	-1.117	0.101
β_{3size}	-0.012	0.072	-0.154	0.130
β_{3lv1}	0.047	0.281	-0.504	0.598
β_{3lv2}	-0.046	0.472	-0.972	0.880
Association				
γ_{grad_S}	-6.497	1.144	-8.554	-4.072

The latency model reports that the type of implanted aortic prosthesis, sex, and the Weibull shape parameter and preoperative left ventricular ejection fraction were also significant in the conditional

failure time latency model, as evidenced in the survival curves plots in Figure 1 to 3, however we see that there is only mild variation in gender at the onset after the surgery as the curve for both male and females flatten together over time. For the incidence model, only gender was significant, where the cure variable has a negative log-odds coefficients for female patients. The association parameter was significant and strong in the link between the survival component and the longitudinal trajectory of valve gradient. The case of negative association parameter indicates that higher values of the longitudinal outcome implies that there was a reduction in the probability in the risk of the event.

4 Conclusion

This paper presented the modelling of longitudinal outcomes and mixture cure survival under shared random effect using latent Gaussian modelling approach, involving evaluating the posterior distribution of the resulting Bayesian modelling using integrated Laplace approximation (INLA) introduced by Rue et al. (2009). The focus here was the sensitivity of the joint modelling approach for longitudinal outcome with nonlinear trajectory and survival cure component with data missingness and outliers. The joint cure modelling approach was applied to the aortic valve replacement surgery data to study the effects of different heart valves on valve gradient (grad) and the risk of death after aortic valve replacement surgery. We compared four prior settings for the Gaussian distribution for the fixed effects and association parameters, as well as the inverse-Wishart distribution for the random effects.

Prior parameter values for Gaussian prior for the fixed effects parameters, mean and precision, mean intercept and precision intercepts and the association parameter, (mean and precision) respectively given as (0, 0.16), (0, 0.16), (0, 0.16), (100, 1) and (0, 0.001), (0, 0.001), (0, 0.001), (100, 1) for priors three and four had limited impact on their posteriors. The prior settings three and four had almost no impact on the marginal posteriors of the random effects parameters for Data A and Data B, meaning that missing values and outliers do not influence our model given this prior settings, thus supporting its robustness. These prior settings three and four also showed similar posterior distributions for the association parameter in both Data A and Data B, again meaning that these priors had limited impact on their posteriors and showed robustness to missing values and outliers as in the case of the random effects parameters.

Bayesian inference has always been concerned with the effects of prior settings on posterior estimates of model parameters and these effects have often been studied via sensitivity analyses and the case of joint modelling longitudinal and survival cure outcomes is presented in this study, where latent Gaussian model leading to INLA for parameter estimation. This study contributes to the literature on joint modelling using approximate Bayesian inference with INLA as a time efficient alternative to Markov Chain Monte Carlo (MCMC).

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

Alafchi, B., Mahjub, H., Tapak, L., Roshanaei, G., and Amirzargar, M. A. (2021). Two-Stage Joint Model for Multivariate Longitudinal and Multistate Processes, with Application to Renal Transplantation Data. *Journal of Probability and Statistics*, *2021*, 1–10. https://doi.org/10.1155/2021/6641602

Alsefri, M., Sudell, M., García-Fiñana, M., and 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

Alvares, D., van Niekerk, J., Krainski, E. T., Rue, H., and Rustand, D. (2024). Bayesian survival analysis with INLA. *Statistics in Medicine*, *43*(20), 3975–4010. https://doi.org/10.1002/sim.10160

Blangiardo, M., and Cameletti, M. (2015). *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley. https://doi.org/10.1002/9781118950203

Chen, M. H., Ibrahim, J. G., and Sinha, D. (2004). A new joint model for longitudinal and survival data with a cure fraction. *J Multivar Anal*, *91*(1), 18–34. https://doi.org/10.1016/j.jmva.2004.04.005

Chi, Y. Y., and Ibrahim, J. G. (2007). Bayesian approaches to joint longitudinal and survival models accommodating both zero and nonzero cure fractions. *Statistica Sinica*, *17*(2), 445–462.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Ekong, A. H., Olayiwola, M. O., Dawodu, A. G., and Osinuga, A. I. (2025). Latent Gaussian Approach to Joint Modelling of Longitudinal and Mixture Cure Outcomes. *Computational Journal of Mathematical and Statistical Sciences*, *4*(1), 72–95. https://doi.org/10.21608/cjmss.2024.303748.1061

He, B., and Luo, S. (2016). Joint modeling of multivariate longitudinal measurements and

survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research*, *25*(4), 1346. https://doi.org/10.1177/0962280213480877

Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2018). 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

Kalbfleisch, J. D., and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley. https://doi.org/10.1002/9781118032985

Lázaro, E., Armero, C., and 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

Lim, E., Ali, A., Theodorou, P., Sousa, I., Ashrafian, H., Chamageorgakis, T., Duncan, A., Henein, M., Diggle, P., and Pepper, J. (2008). Longitudinal Study of the Profile and Predictors of Left Ventricular Mass Regression after Stentless Aortic Valve Replacement. *The Annals of Thoracic Surgery*, *85*(6), 2026–2029. https://doi.org/10.1016/j.athoracsur.2008.02.023

Medina-Olivares, V., Calabrese, R., Crook, J., and Lindgren, F. (2023). 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

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

Rue, H., Martino, S., and 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

Rustand, D., van Niekerk, J., Krainski, E. T., and Rue, H. (2024). *Joint Modeling of Multivariate Longitudinal and Survival Outcomes with the R package INLAjoint*.

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

Rustand, D., van Niekerk, J., Rue, H., Tournigand, C., Rondeau, V., and 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

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

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

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

van Niekerk, J., Bakka, H., Rue, H., and Schenk, O. (2021). New Frontiers in Bayesian Modeling Using the INLA Package in R. *Journal of Statistical Software*, *100*(2). https://doi.org/10.18637/jss.v100.i02