# Bayesian Estimation of Two-Part Joint Models for a Longitudinal Semicontinuous Biomarker and a Terminal Event with R-INLA: Interests for Cancer Clinical Trial Evaluation

**Denis Rustand**
Biostatistic Team, Bordeaux Population Health Center,
ISPED, Centre INSERM U1219, Bordeaux, France.

**Janet van Niekerk**
King Abdullah University of Science and Technology,
CEMSE Division, Saudi Arabia.

**Håvard Rue**
King Abdullah University of Science and Technology,
CEMSE Division, Saudi Arabia.

**Christophe Tournigand**
Hopital Henri Mondor, Creteil, France

**Virginie Rondeau**
Biostatistic Team, Bordeaux Population Health Center,
ISPED, Centre INSERM U1219, Bordeaux, France.

**Laurent Briollais**
Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital,
Dalla Lana School of Public Health (Biostatistics),
University of Toronto, 600 University Ave., Ontario M5G 1X5, Canada

October 26, 2020

**Abstract**

Two-part joint model for a longitudinal semicontinuous biomarker and a terminal event has been recently introduced based on frequentist computation. The biomarker distribution is decomposed into a probability of positive value and the expected value among positive values. Shared random effects can represent the association structure between the biomarker and the terminal event. The computational burden increases compared to standard joint models with a single regression model for the biomarker. In this context, the frequentist estimation implemented in the R package **frailtypack** can be challenging for complex models (i.e., large number of parameters and dimension of the random effects). As an alternative, we propose a Bayesian estimation of two-part joint models based on the Integrated Nested Laplace Approximation (INLA) algorithm to alleviate the computational burden and be able to fit more complex models. Our simulation studies show that **R-INLA** reduces the computation time substantially as well as the variability of the estimates and improves the model convergence compared to **frailtypack**. We contrast the Bayesian and frequentist approaches in two randomized cancer clinical trials (GERCOR and PRIME studies), where **R-INLA** suggests a stronger association between the biomarker and the risk of event. Moreover, the Bayesian approach was able to characterize subgroups of patients associated with different responses to treatment in the PRIME study where **frailtypack** had convergence issues.

Our study suggests that the Bayesian approach using **R-INLA** algorithm enables broader applications of the two-part joint model to clinical applications.

# 1 Introduction

Estimation of joint models for longitudinal and time-to-event data were initially introduced using maximum likelihood estimation (Wulfsohn and Tsiatis (1997); Henderson et al. (2000); Song et al. (2002); Chi and Ibrahim (2006)). It was further developed within the Bayesian framework in situations where maximum likelihood estimation with asymptotic assumptions faces nonidentifiability issues. It allows flexible and more complex association structures and can handle multiple longitudinal outcomes (Andrinopoulou and Rizopoulos (2016)). Bayesian joint models can be fitted with the R package **JMbayes** (Rizopoulos et al. (2016)), which has been used in many biomedical researches (Lawrence Gould et al. (2015)), among other packages (e.g. **rstanarm**, Muth et al. (2018)). Bayesian estimation is usually based on MCMC techniques (Hanson et al. (2011); R. Brown and G. Ibrahim (2003); Xu and Zeger (2001); Rizopoulos and Ghosh (2011)), which can have slow convergence properties. The Integrated Nested Laplace Approximation (INLA) algorithm has been recently introduced as an alternative to MCMC techniques for latent Gaussian models (LGM) (Rue et al. (2009); Martins et al. (2013)). Many statistical models for spatial statistics, time series, etc., can be formulated as LGMs. A key feature of INLA is to provide approximations of the posterior marginals needed for Bayesian inference very efficiently and that still remain very accurate compare to MCMC methods (Rue et al. (2017)). By formulating complex joint models as LGMs, **R-INLA** can be used to fit these models as developed recently (Van Niekerk, Bakka, and Rue (2019); Van Niekerk, Bakka, Rue, and Schenk (2019)). For the two-part joint model, **R-INLA** is yet to be used for inference.

Two-part joint models (TPJMs) for a longitudinal semicontinuous biomarker and a terminal event have been recently introduced (Rustand, Briollais, Tournigand, and Rondeau (2020)) and applied to the joint analysis of survival times and repeated measurements of the Sum of the Longest Diameter of target lesions (SLD), which is a biomarker representative of tumor burden in cancer clinical trials. The TPJM uses a conditional two-part joint model that decomposes the biomarker distribution into a binary outcome (zero vs. positive value) fitted with a logistic mixed effects model and a continuous outcome (positive values only) fitted with either a linear mixed effect model on the log-transformed outcome or a lognormal mixed effects model (Rustand, Briollais, and Rondeau (2020)). The "conditional" form of the two-part model gives the effect of covariates on the mean biomarker value conditional on a positive value in the continuous part, an alternative marginal form has recently been proposed to get the effect of covariates on the (unconditional) mean of the biomarker. A drawback of the marginal two-part model is that it may lead to arbitrary heterogeneity and provide less interpretable estimates on the conditional mean of the biomarker among positive values (Smith et al. (2014)). In this article, we focus on the conditional two-part joint model, simply referred to as TPJM in what follows. The association with the survival model can be specified in terms of shared random effects, i.e., random effects that are shared between the different components of the models including the binary and continuous parts of the model and the survival component. An important limitation of such models is the estimation procedure that requires a numerical approximation of the random effects distribution, which can lead to long computation times and convergence issues with high-dimensional parameter settings and complex association structures between the different components of the TPJMs. In this article, we propose an efficient Bayesian estimation procedure for the TPJM which relies on INLA, as implemented in the R package **R-INLA**. The Bayesian inference is compared to the frequentist estimation of the TPJM available in the R package **frailtypack**. The remainder of the article is structured as follows: in Section 2, we describe the TPJM and introduce the frequentist and Bayesian estimations. In Section 3, we present a simulation study to assess the performance of these two estimation strategies in terms of bias, coverage probability, computation time and convergence rate. An application to two randomized clinical trials each comparing two treatment strategies in patients with metastatic colorectal cancer is proposed in Section 4 and we conclude with a discussion in Section 5.

# 2 Estimation of the conditional two-part joint model

## 2.1 Model specification

Let $Y_{ij}$ denote the biomarker measurement for individual $i (i = 1, ..., n)$ at visit $j (j = 1, ..., n_i)$, $T_i$ denotes the survival time and $\delta_i$ the censoring indicator for individual $i$. We use a logistic mixed effect model for the probability of a positive value of the biomarker and a lognormal mixed effect model for the conditional expected biomarker value. A proportional hazards survival model specifies the effect of covariates on survival time, adjusted for the individual heterogeneity captured in the biomarker model. The complete model is defined as follows:

$$\begin{cases} \eta_{Bij} = \text{Logit}(\text{Prob}(Y_{ij} > 0)) = \boldsymbol{X}_{Bij}^\top \boldsymbol{\alpha} + \boldsymbol{Z}_{Bij}^\top \boldsymbol{a}_i & \text{(Binary part)}, \\ \eta_{Cij} = \text{E}[Y_{ij} | Y_{ij} > 0] = \exp(\boldsymbol{X}_{Cij}^\top \boldsymbol{\beta} + \boldsymbol{Z}_{Cij}^\top \boldsymbol{b}_i) & \text{(Continuous part)}, \\ \lambda_i(t) = \lambda_0(t) \ \exp(\eta_{Si}) = \lambda_0(t) \ \exp\left(\boldsymbol{X}_i^\top \ \boldsymbol{\gamma} + \boldsymbol{a}_i^\top \boldsymbol{\varphi}_a + \boldsymbol{b}_i^\top \boldsymbol{\varphi}_b\right) & \text{(Survival part)}, \end{cases}$$

where $\boldsymbol{X}_{Bij}$, $\boldsymbol{X}_{Cij}$ and $\boldsymbol{X}_i$ are vectors of covariates associated to the fixed effects $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Similarly, $\boldsymbol{Z}_{Bij}$ and $\boldsymbol{Z}_{Cij}$ are vectors of covariates associated to the random effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ in the binary and continuous parts. These random effects are shared in the survival model, with association parameters $\boldsymbol{\varphi}_a$ and $\boldsymbol{\varphi}_b$, respectively. Therefore, the random effects accounts for both the association between the three components of the model and the correlation between the repeated measurements in the longitudinal process (observations are independent conditional on the random effects). The joint distribution assumes that the vectors of random effects underlies both the longitudinal and survival process, the joint distribution of the observed outcomes for individual $i$ is defined by

$$\begin{aligned} p(T_i, \delta_i, \boldsymbol{Y}_i | \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}) &= p(T_i, \delta_i | \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}) \prod_{j=1}^{n_i} p(Y_{ij} | \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}) \\ &= p(T_i, \delta_i | \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}) \prod_{j=1}^{n_i} p(Y_{ij} | Y_{ij} > 0; \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}) \ p(Y_{ij} > 0; \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}) \end{aligned}$$

with $\boldsymbol{\Theta}$ the full parameter vector, including the parameters for the binary, continuous and survival outcomes, the baseline hazard function and the random effects covariance matrix, such that the full conditional distribution is given by

$$p(\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Y} | \boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\Theta}) = \prod_{i=1}^{n} p(T_i, \delta_i, \boldsymbol{Y}_i | \boldsymbol{a}_i, \boldsymbol{b}_i; \boldsymbol{\Theta}).$$

The likelihood contribution for the $i$th subject can be formulated as follows

$$\begin{aligned} L_i(\boldsymbol{\Theta} | \boldsymbol{Y}_i, T_i, \delta_i) = \int_{\boldsymbol{a}_i} \int_{\boldsymbol{b}_i} \prod_{j=1}^{n_i} & \exp\left(\boldsymbol{X}_{Bij}^\top \boldsymbol{\alpha} + \boldsymbol{Z}_{Bij}^\top \boldsymbol{a}_i\right)^{U_{ij}} \left(1 - \frac{\exp(\boldsymbol{X}_{Bij}^\top \boldsymbol{\alpha} + \boldsymbol{Z}_{Bij}^\top \boldsymbol{a}_i)}{1 + \exp(\boldsymbol{X}_{Bij}^\top \boldsymbol{\alpha} + \boldsymbol{Z}_{Bij}^\top \boldsymbol{a}_i)}\right) \\ & \times \left\{ \frac{1}{Y_{ij} \sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{(\log(Y_{ij}) - \mu_{ij})^2}{2\sigma_\epsilon^2}\right) \right\}^{U_{ij}} \\ & \times \lambda_i(T_i | \boldsymbol{a}_i, \boldsymbol{b}_i)^{\delta_i} \exp\left(-\int_0^{T_i} \lambda_i(t | \boldsymbol{a}_i, \boldsymbol{b}_i) \mathrm{d}t\right) p(\boldsymbol{a}_i, \boldsymbol{b}_i) \mathrm{d}\boldsymbol{b}_i \mathrm{d}\boldsymbol{a}_i, \end{aligned}$$

where $U_{ij} = I[Y_{ij} > 0]$ and $\lambda_i(t) = \lambda_0(t) \ \exp\{\boldsymbol{X}_{Si}(t)^\top \boldsymbol{\gamma} + \boldsymbol{a}_i^\top \boldsymbol{\varphi}_a + \boldsymbol{b}_i^\top \boldsymbol{\varphi}_b\}$.

## 2.2 Bayesian estimation of the TPJM

Define $\boldsymbol{D} \equiv \{T_i, \delta_i, Y_{ij} : i = 1, \cdots, n; j = 1, \cdots, n_i\}$ the observation variables. The goal of the Bayesian inference is to estimate the posterior distribution $\pi(\boldsymbol{\Theta} | \boldsymbol{D})$. The joint posterior distribution $\pi(\boldsymbol{\Theta} | \boldsymbol{D})$ is given by Bayes theorem as

$$\pi(\boldsymbol{\Theta} | \boldsymbol{D}) = \frac{p(\boldsymbol{D} | \boldsymbol{\Theta}) \pi(\boldsymbol{\Theta})}{\pi(\boldsymbol{D})} \propto p(\boldsymbol{D} | \boldsymbol{\Theta}) \pi(\boldsymbol{\Theta}),$$

where $p(\boldsymbol{D}|\boldsymbol{\Theta})$ is the likelihood and $\pi(\boldsymbol{\Theta})$ is the joint prior. The marginal likelihood $\pi(\boldsymbol{D}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{D}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}$ acts as a normalizing constant. The posterior marginal distribution of each parameter is then obtained by integrating out the other parameters of the model. In many cases, the posterior distribution is not analytically tractable and sampling-based methods like MCMC can be used. Approximate methods like INLA, provide exact approximations to the posterior at lower cost than sampling-based methods. The INLA methodology is based on the assumption that the statistical model is a latent Gaussian model, which we show in the next section for the TPJM.

## 2.3 Formulation of the TPJM as a latent Gaussian model

Let $\boldsymbol{u} \equiv (\boldsymbol{\eta}_B, \boldsymbol{\eta}_C, \boldsymbol{\eta}_S, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\varphi})$ be the set of latent Gaussian variables related to the TPJM, where $\boldsymbol{\lambda}$ is a vector of coefficients associated with a random walk order one or order two used to approximate the baseline hazard function $\lambda_0(t)$ of the survival model. Note that the first $\sum_{i=1}^{n} n_i + \sum_{i=1}^{n} n_i + n$ elements of $\boldsymbol{u}$ are the linear predictors of the TPJM and the rest of the elements are the latent unobserved variables. For that reason, the random field $\boldsymbol{u}$ is termed the latent field.
In particular, we assume $\boldsymbol{a}_i, \boldsymbol{b}_i|\boldsymbol{Q}_{ab} \sim \mathcal{N}(0, \boldsymbol{Q}_{ab}^{-1})$, $\boldsymbol{\alpha} \sim \mathcal{N}(0, \tau_\alpha \boldsymbol{I})$, $\boldsymbol{\beta} \sim \mathcal{N}(0, \tau_\beta \boldsymbol{I})$, $\boldsymbol{\gamma} \sim \mathcal{N}(0, \tau_\gamma \boldsymbol{I})$ and $\boldsymbol{\varphi} \sim \mathcal{N}(0, \tau_\varphi \boldsymbol{I})$. The coefficients of the baseline hazard $\boldsymbol{\lambda}$ are assumed to follow either a random walk one or random walk two model. These models are stochastic spline models with precision parameter $\tau_\lambda$. Thus, the latent field $\boldsymbol{u}$ is multivariate Gaussian with zero mean and precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_1)$, i.e.,

$$\boldsymbol{u}|\boldsymbol{\theta}_1 \sim \mathcal{N}(0, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_1)).$$

Note that $\boldsymbol{Q}(\boldsymbol{\theta}_1)$ is a sparse matrix indexed by a low dimension of parameters $\boldsymbol{\theta}_1$. This then implies that the latent field $\boldsymbol{u}$ is a Gaussian Markov random field (GMRF).

The distribution of the observation variables $\boldsymbol{D}$ is denoted by $\pi(\boldsymbol{D}|\boldsymbol{u}, \boldsymbol{\theta}_2)$ and depends on the set of hyper-parameters $\boldsymbol{\theta}_2$ that influence the likelihood. They are assumed to be conditionally independent over the $n$ individuals given the latent Gaussian random field $\boldsymbol{u}$ and hyper-parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,

$$\boldsymbol{D}|\boldsymbol{u}, \boldsymbol{\theta} \sim \prod_{i=1}^{n} p(\boldsymbol{d_i}|u_i, \boldsymbol{\theta}).$$

Thus, assuming a prior $\pi(\boldsymbol{\theta})$ for the hyper-parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, the posterior of $(\boldsymbol{u}, \boldsymbol{\theta})$ can be written as

$$\pi(\boldsymbol{u}, \boldsymbol{\theta}|\boldsymbol{D}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{u}|\boldsymbol{\theta}) \prod_{i=1}^{n} p(\boldsymbol{d}_i|u_i, \boldsymbol{\theta}),$$

$$\propto \pi(\boldsymbol{\theta})|\boldsymbol{Q}(\boldsymbol{\theta}_1)|^{n/2} \exp\Big[\frac{1}{2}\boldsymbol{u}^T \boldsymbol{Q}(\boldsymbol{\theta}_1)\boldsymbol{u} + \sum_{i=1}^{n} \log\{p(\boldsymbol{d}_i|u_i, \boldsymbol{\theta})\}\Big].$$

This construction then shows that the TPJM is in fact an LGM since the latent field is a Gaussian Markov field and each data contribution depends on only one element of the latent field.

The main aim of INLA is then to approximate the posterior marginals $\pi(u_i|\boldsymbol{D})$, $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ and $p(\theta_j|\boldsymbol{D})$.

## 2.4 INLA

The INLA methodology introduced by Rue and Held (2005) is a major contribution to achieving efficient Bayesian inference, especially for complex or large models. INLA uses a unique combination of Laplace Approximations and conditional distributions to approximate the joint posterior density as well as the marginals of the latent field and hyperparameters. It is thus not a sampling based method like MCMC and such.
For the sake of brevity, the INLA methodology can be presented in the following three steps:

1. Approximate
$$\pi(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{\pi(\boldsymbol{u}, \boldsymbol{\theta}|\boldsymbol{D})}{\pi(\boldsymbol{u}|\boldsymbol{\theta}, \boldsymbol{D})} \approx \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{u}|\boldsymbol{\theta})\pi(\boldsymbol{D}|\boldsymbol{u}, \boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{u}|\boldsymbol{\theta}, \boldsymbol{D})}\Big|_{u=u^*(\boldsymbol{\theta})},$$
where the Gaussian or Laplace approximation is used to approximate the denominator at the mode $u^*(\boldsymbol{\theta})$ of the latent field for a given configuration of $\boldsymbol{\theta}$.

2. Approximate

$$\pi(u_j|\boldsymbol{\theta}, \boldsymbol{D}) \propto \frac{\pi(\boldsymbol{u}, \boldsymbol{\theta}|\boldsymbol{D})}{\pi(\boldsymbol{u}_{-j}|u_j, \boldsymbol{\theta}, \boldsymbol{D})},$$

using a Gaussian approximation (option 1), or in a similar way as mentioned in step 1 (option 2) or by expanding the numerator and denominator up to a third order Taylor series expansion and then applying a Laplace approximation (option 3).

3. Use numerical integration to approximate

$$\pi(u_j|\boldsymbol{Y}) \approx \sum_{h=1}^{H} \tilde{\pi}(u_j|\boldsymbol{\theta}_h^*, \boldsymbol{Y}) \tilde{\pi}(\boldsymbol{\theta}_h^*|\boldsymbol{Y}) \Delta_h,$$

from steps 1 and 2. The integration points $\{\boldsymbol{\theta}_1^*, ..., \boldsymbol{\theta}_H^*\}$ are selected from a rotation using polar coordinates and based on the density at these points.

### 2.5 Priors for the hyperparameters

From the formulation of the TPJM as an LGM, the prior for the hyperparameters, $\pi(\boldsymbol{\theta})$, should be specified. This prior can assume any form while keeping the TPJM an LGM. Amidst the debate about priors, Simpson et al. (2017) proposed a framework to construct principled priors for hyperparameters, namely penalizing complexity (PC) priors. These priors are derived based on the distance from a complex model to a simpler (base) model, with a user-defined parameter that informs the strength of contraction towards the simpler model. This parameter defines whether the PC priors are vague, weakly informative, or strongly informative based on the departure from the base model measured by the Kullback-Leibler distance. It is based on the principle of parsimony, simplifying the interpretation of the results by ensuring that the priors do not overfit.

In our case we have various precision hyperparameters, $\{\boldsymbol{Q}_{ab}, \tau_\alpha, \tau_\beta, \tau_\gamma, \tau_\varphi, \tau_\lambda\}$. We assign weakly informative priors to the fixed effects such that $\tau_\alpha = \tau_\beta = \tau_\gamma = \tau_\varphi = 10^{-3}$. We thus need to formulate priors for the elements of $\boldsymbol{Q}_{ab}$ and $\tau_\lambda$. For all these hyperparameters (precision and correlation parameters), we assume the respective PC priors as given in Simpson et al. (2017).
As illustration we give the details for the precision of the first element of $\boldsymbol{a}$, $\tau_{a_0}$. The PC prior is derived as

$$\pi(\tau_{a_0}) = \frac{\rho}{2} \tau_{a_0}^{-3/2} \exp(-\rho \tau_{a_0}^{-1/2}),$$

with the user-defined scaling parameter $\rho = -\frac{\ln(v)}{w}$. This parameter is chosen based on the desired tail behaviour (or strength of contraction towards the base model $\sigma_{a_0} = \tau_{a_0}^{-1/2} = 0$) in the sense that $v$ and $w$ are such that

$$P[\sigma_{a_0} > w] = v, \quad w > 0, 0 < v < 1.$$

Larger values of $v$ and $w$ results in higher prior density away from the base model, whereas smaller values of $v$ places more density closer to the base model.

## 3 Simulation study

### 3.1 Settings

We designed simulations studies to compare the performances of **R-INLA** and **frailtypack** in terms of bias of the parameter estimates, coverage probabilities, computation time and convergence rates. The main factor driving the performance is the model complexity defined by the number of parameters. In particular, the number of correlated random effects defines the dimension of the integration that needs to be numerically approximated. We propose two simulation scenarios based on the results obtained from the real data analyses. The first scenario includes a random intercept in the binary and continuous parts of the TPJM that are correlated. The second simulation scenario includes an additional random-effect for the individual deviation from the mean slope in the continuous part, thus 3 correlated random effects. For each scenario, we generate

1000 datasets with 200 individuals each, corresponding to a small sample size commonly seen in randomized clinical trials. We first sample the positive longitudinal biomarker repeated measurements from a log-normal distribution and include the zero values sampled from a binomial distribution. The relationship between the probability of zero value and the positive values is given by the correlated random effects. Survival times for the terminal event are generated with an exponential baseline hazard function with a scale of 0.2, an administrative censoring is assumed to occur at the end of the follow-up (4 years). The rate of zeros is 8% (SD=1%), which is in between what we observed in our two real datasets (12% of zeros in application 1 and 4% in application 2). A zero value observation corresponds to a patient who experienced a complete disappearance of his/her target lesions and thus is extremely informative about treatment effect.

The baseline hazard function in the survival part of the model is approximated by a random walk model with **R-INLA** (Martino et al. (2011)) such that for $m$ bins of the time axis,

$$\lambda_k - \lambda_{k-1} \sim N(0, \tau_\lambda),$$

where the PC prior (see Section 2.5) is used as the prior for $\tau_\lambda$.
The random walk order one model is a stochastic smoothing spline that smooths based on first order differences. The number of bins are not influential (as opposed to knots of other splines) since an increase in bins only results in an estimate closer to the stochastic model. In the simulations and applications, we use the random walk order two model that provides a smoother spline since the smoothing is then done on the second order. See Van Niekerk et al. (2020) for more details on the use of these random walk models as Bayesian smoothing splines. This approximation is different with **frailtypack** that uses cubic M-splines with 5 knots. A penalization ensures that the baseline hazard is smooth (a smoothing parameter is chosen using an approximate cross-validation criterion from a separate Cox model). The Levenberg-Marquardt algorithm, a robust Newton-like algorithm maximizes the log-likelihood function with **frailtypack** (Marquardt (1963)). The convergence of the algorithm depends on three conditions: The difference between the log-likelihood, the estimated coefficients and the gradient of the log-likelihood of two consecutive iterations must be under $10^{-3}$. We use a Monte-Carlo approximation for the approximation of the integrals over the random effects in the likelihood function, with 1000 integration points which is a reasonable tradeoff between the precision of the approximation and computation time. The simulation studies are performed with 80 CPUs, **frailtypack** uses Message Passing Interface (MPI) for parallel computation while the conjunction of **R-INLA** with the **PARDISO** library provides a high performance computing environment with parallel computing support (Schenk and Gärtner (2004)). In practice, the 80 CPUs are only useful to reduce the computation time with **frailtypack** because the computation time with INLA is very low regardless of the number of threads because of the small sample size and number of hyperparameters.

## 3.2 Results

In the results, we are comparing a Bayesian and frequentist method and for this we have to keep in mind that each has a different criteria for evaluation of the method. Frequentist bias is used to evaluate the results from **frailtypack** while the plausibility of the result based on 95% credible intervals are used to evaluate the results from **R-INLA** (Hespanhol et al. (2019)). However we are interested in the Bayesian approximation of the MLE (i.e. non informative priors) and therefore provide an interpretation in this context.

### 3.2.1 Scenario 1: Two correlated random effects

The fixed effects parameters from the binary and continuous parts are properly estimated with both algorithms, with similar precision and coverage probabilities close to the expected 95% level. The treatment effect in the survival part ($\gamma = 0.2$) is associated to a larger variability with **frailtypack** ($\hat{\gamma} = 0.22$, SD=0.35, CP=96%) compared to **R-INLA** ($\hat{\gamma} = 0.19$, SD=0.27, CP=96%). The true value of the random intercept in the binary part ($\sigma_a = 1$) is within the 95% credible interval with **R-INLA** ($\hat{\sigma}_a = 0.86$, SD=0.19), with a slightly lower mean value compared to **frailtypack** ($\hat{\sigma}_a = 0.97$, SD=0.22). The random intercept in the continuous part is found similar with both algorithms with same variability but the correlation ($corr_{ab} = 0.5$) has a reduced variability with **R-INLA** ($\hat{corr}_{ab} = 0.48$, SD=0.10) compared to **frailtypack** ($\hat{corr}_{ab} = 0.51$, SD=0.15). The main difference observed is the value of the parameters for the association of the random effects with the risk of event, which links the biomarker to the terminal event. The association of the random

intercept from the binary part ($\varphi_a = 1$) has much lower variability with **R-INLA** ($\hat{\varphi}_a = 1.00$, SD=0.11, CP=99%) and is unbiased with good coverage with **frailtypack** ($\hat{\varphi}_a = 1.08$, SD=0.82, CP=93%). The association of the random intercept from the continuous ($\varphi_b = 1$) part is biased upwards with **frailtypack** with large variability ($\hat{\varphi}_b = 1.33$, SD=1.13, CP=92%) while **R-INLA** recovers the true value ($\hat{\varphi}_b = 1.05$, SD=0.15, CP=99%). This could be due to the small sample size problems that cause more issues for the frequentist framework. Although **R-INLA** gives good estimates with small variability for these parameters, the coverage probabilities are higher than the expected 95%. The computation times are much lower with **R-INLA** (14 seconds per model, SD=1) compared to **frailtypack** (66 seconds per model, SD=26). Finally, all models converged with **R-INLA** while 11% of the 1000 models did not reach convergence with **frailtypack**.

Table 1: Simulations with two correlated random effects

| Package | | R-INLA Est. (SE) | frailtypack Est. (SE) |
|---|---|---|---|
| **Binary part** (SLD>0 versus SLD=0) | | | |
| intercept | $\alpha_0 = 4$ | 3.96 (0.35) [94%] | 4.02 (0.38) [95%] |
| time (year) | $\alpha_1 = -0.5$ | -0.51 (0.11) [95%] | -0.51 (0.12) [95%] |
| treatment (B/A) | $\alpha_2 = -0.5$ | -0.49 (0.45) [96%] | -0.50 (0.47) [95%] |
| time:treatment (B/A) | $\alpha_3 = 0.5$ | 0.50 (0.18) [94%] | 0.51 (0.18) [95%] |
| **Continuous part** ($E[Y_{ij}\|Y_{ij} > 0]$) | | | |
| intercept | $\beta_0 = 2$ | 2.00 (0.05) [95%] | 2.00 (0.06) [92%] |
| time (years) | $\beta_1 = -0.3$ | -0.30 (0.01) [95%] | -0.30 (0.01) [95%] |
| treatment (B/A) | $\beta_2 = -0.3$ | -0.30 (0.08) [94%] | -0.30 (0.09) [91%] |
| time:treatment (B/A) | $\beta_3 = 0.3$ | 0.30 (0.02) [94%] | 0.30 (0.02) [95%] |
| residual S.E. | $\sigma_\varepsilon = 0.3$ | 0.30 (0.01) [89%] | 0.30 (0.01) [94%] |
| **Death risk** | | | |
| treatment (B/A) | $\gamma_1 = 0.2$ | 0.19 (0.27) [96%] | 0.22 (0.35) [96%] |
| **Association** | | | |
| Intercept (binary part) | $\varphi_a = 1$ | 1.00 (0.11) [99%] | 1.08 (0.82) [93%] |
| Intercept (continuous part) | $\varphi_b = 1$ | 1.05 (0.15) [99%] | 1.33 (1.13) [92%] |
| **Random effects** | | | |
| intercept (binary part) | $\sigma_a = 1$ | 0.86 (0.19) | 0.97 (0.22) |
| intercept (continuous part) | $\sigma_b = 0.5$ | 0.50 (0.03) | 0.50 (0.03) |
| | $corr_{ab} = 0.5$ | 0.48 (0.10) | 0.51 (0.15) |
| **Computation time** | | | |
| 80 CPUs - Intel Xeon E5-4627 v4 2.60 GHz | | 14 sec. (1) | 66 sec. (26) |
| **Convergence rate** | | 100% | 89% |

### 3.2.2 Scenario 2: Three correlated random effects

With an additional random-effect compared to scenario 1, the fixed effects parameters are still properly estimated with **R-INLA**. The coverage probabilities are low with **frailtypack** for the slope and treatment by slope parameters in the continuous part ($\beta_1 = -0.3$ and $\beta_3 = 0.3$), while the parameter estimates remain unbiased. The variability for these two parameters is lower with **R-INLA** ($\hat{\beta}_1 = -0.30$, SD=0.06, CP=94% and $\hat{\beta}_3 = 0.30$, SD=0.08, CP=95%) compared to **frailtypack** ($\hat{\beta}_1 = -0.25$, SD=0.11, CP=46% and $\hat{\beta}_3 = 0.29$, SD=0.14, CP=44%). As observed in the first scenario, the treatment effect in the survival model has lower variability with **R-INLA** ($\hat{\gamma}_1 = 0.20$, SD=0.30, CP=95%), moreover the coverage probability for this parameter is lower than expected with **frailtypack** ($\hat{\gamma}_1 = 0.24$, SD=0.49, CP=84%). The random effects covariance structure estimations is similar than in scenario 1, where the posterior mean from **R-INLA** is slightly lower than the value of the random intercept in the binary part ($\hat{\sigma}_a = 0.86$, SD=0.15) with lower variability for the standard deviations and correlation terms overall. The association parameters ($\varphi_a = 1$, $\varphi_{b_0} = 1$, $\varphi_{b_1} = 1$) are recovered well and have much lower variability with **R-INLA** ($\hat{\varphi}_a = 1.03$, SD=0.13, CP=98%, $\hat{\varphi}_{b_0} = 1.07$, SD=0.14, CP=98%, $\hat{\varphi}_{b_1} = 1.07$, SD=0.14, CP=98%) compared to **frailtypack** ($\hat{\varphi}_a = 0.87$, SD=1.86, CP=91%, $\hat{\varphi}_{b_0} = 1.03$, SD=1.82, CP=89%, $\hat{\varphi}_{b_1} = 1.44$, SD=1.70, CP=91%), but still with conservative coverage probabilities. Computation times remain much lower with **R-INLA** (19 seconds per model, SD=2) compared to **frailtypack** (159 seconds per model, SD=52) for which the time increased substantially when adding the third random-effect. Moreover, the convergence rate of the model is reduced with **frailtypack** for this scenario (82%), because the model complexity increased while it remains 100% with **R-INLA**.

Table 2: Simulations with three correlated random effects

| Package | | R-INLA Est. (SE) | frailtypack Est. (SE) |
|---|---|---|---|
| **Binary part** (SLD>0 versus SLD=0) | | | |
| intercept | $\alpha_0 = 4$ | 3.95 (0.35) [94%] | 4.03 (0.39) [94%] |
| time (year) | $\alpha_1 = -0.5$ | -0.52 (0.12) [94%] | -0.51 (0.12) [95%] |
| treatment (B/A) | $\alpha_2 = -0.5$ | -0.51 (0.47) [95%] | -0.50 (0.50) [94%] |
| time:treatment (B/A) | $\alpha_3 = 0.5$ | 0.51 (0.18) [95%] | 0.50 (0.18) [96%] |
| **Continuous part** ($E[Y_{ij}\|Y_{ij} > 0]$) | | | |
| intercept | $\beta_0 = 2$ | 2.00 (0.05) [96%] | 1.99 (0.06) [88%] |
| time (years) | $\beta_1 = -0.3$ | -0.30 (0.06) [94%] | -0.25 (0.11) [46%] |
| treatment (B/A) | $\beta_2 = -0.3$ | -0.30 (0.08) [96%] | -0.29 (0.09) [87%] |
| time:treatment (B/A) | $\beta_3 = 0.3$ | 0.30 (0.08) [95%] | 0.29 (0.14) [44%] |
| residual S.E. | $\sigma_\varepsilon = 0.3$ | 0.29 (0.01) [88%] | 0.30 (0.01) [96%] |
| **Death risk** | | | |
| treatment (B/A) | $\gamma_1 = 0.2$ | 0.20 (0.30) [95%] | 0.24 (0.49) [84%] |
| **Association** | | | |
| Intercept (binary part) | $\varphi_a = 1$ | 1.03 (0.13) [98%] | 0.87 (1.86) [91%] |
| Intercept (continuous part) | $\varphi_{b_0} = 1$ | 1.07 (0.14) [98%] | 1.03 (1.82) [89%] |
| Slope (continuous part) | $\varphi_{b_1} = 1$ | 1.07 (0.14) [98%] | 1.44 (1.70) [91%] |
| **Random effects** | | | |
| intercept (binary part) | $\sigma_a = 1$ | 0.86 (0.15) | 1.07 (0.24) |
| intercept (continuous part) | $\sigma_{b_0} = 0.5$ | 0.50 (0.03) | 0.50 (0.04) |
| slope (continuous part) | $\sigma_{b_1} = 0.5$ | 0.49 (0.03) | 0.58 (0.10) |
| | $corr_{ab_0} = 0.5$ | 0.47 (0.10) | 0.48 (0.16) |
| | $corr_{ab_1} = 0.5$ | 0.46 (0.12) | 0.58 (0.16) |
| | $corr_{b_0b_1} = -0.2$ | -0.19 (0.09) | -0.14 (0.19) |
| **Computation time** | | | |
| 80 CPUs - Intel Xeon E5-4627 v4 2.60 GHz) | | 19 sec. (2) | 159 sec. (52) |
| **Convergence rate** | | 100% | 82% |

### 3.3 Conclusions

Our method comparison suggests that the frequentist approach, implemented in **frailtypack**, reaches some limitations when fitting the more complex TPJMs, compared to the Bayesian approach implemented in **R-INLA**. Convergence rates are lower and estimation of the association parameters is highly variable with **frailtypack**. However, a representation of the baseline survival curves estimated under both scenarios is displayed in Figure 1. The median of the estimated survival curves is slightly lower than the true survival with **R-INLA** although the credible interval contains the true curve, while the point estimate from **frailtypack** is closer to the true curve but yields much larger confidence intervals.

## 4 Application

We applied the Bayesian TPJM to two cancer clinical trials, the GERCOR and the PRIME studies. A comparison with **frailtypack** is provided only for the GERCOR data since this approach did not converge on the PRIME study. We used the same parameterizations for **R-INLA** and **frailtypack** as detailed in the simulation studies. In the context of a Bayesian approximation of the MLE, we provide indications of the p-value for both **frailtypack** and **INLA** to ease the interpretation and the comparison of the results.

Table 3: Description of the GERCOR and PRIME study datasets

| Study | GERCOR | | PRIME | |
|---|---|---|---|---|
| Treatment | arm A FOLFIRI/FOLFOX6 | arm B FOLFOX6/FOLFIRI | arm A FOLFOX4 | arm B Panitumumab/FOLFOX4 |
| Number of patients enrolled | 109 | 111 | 593 | 590 |
| Number of patients for the analysis | 101 | 104 | 223 | 219 |
| number of repeated measurements of the SLD | 748 | 727 | 1192 | 1081 |
| Number of zero values (%) | 118 (16.2%) | 56 (7.5%) | 47 (3.8%) | 52 (4.6%) |
| Number of death (%) | 83 (82.2%) | 82 (78.8%) | 164 (73.5%) | 164 (74.9%) |
| Median OS (years) | 1.8 (1.4-2.3) | 1.8 (1.5-2.2) | 1.7 (1.5-1.9) | 1.4 (1.3-1.7) |
| KRAS exon 2 at codons 12 and 13 | | | | |
| Nonmutated | | | 132 (59.2%) | 128 (58.4%) |
| Mutated | | | 91 (40.8%) | 91 (41.6%) |
| Not available | 101 (100%) | 104 (100%) | | |

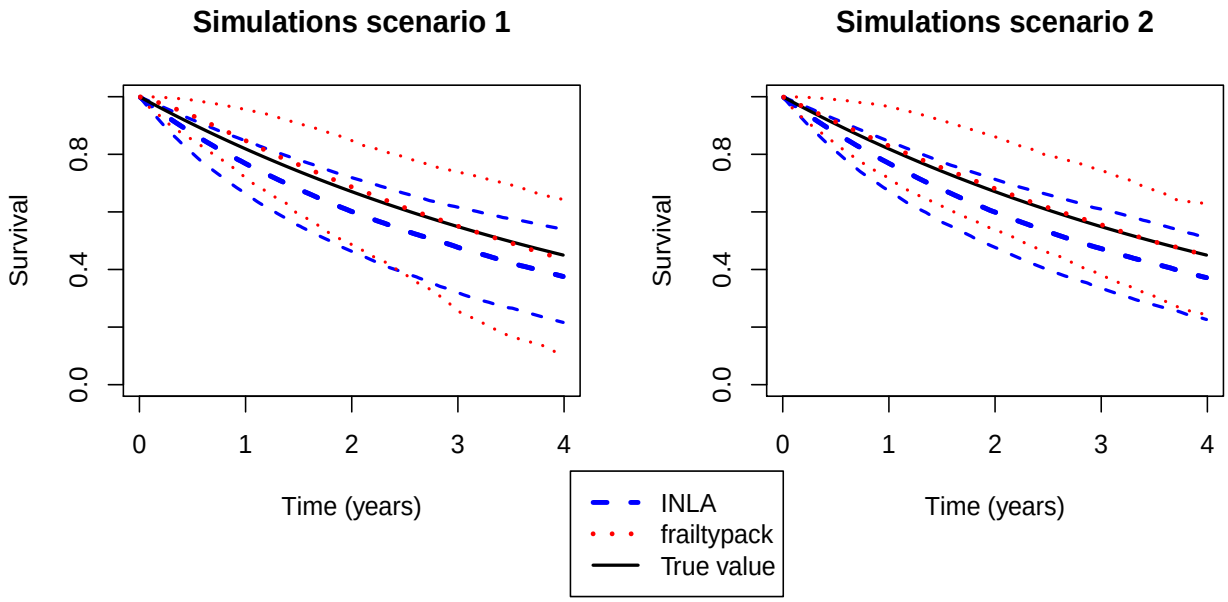


Figure 1: Quantiles at 2.5%, 50% and 97.5% of the baseline survival curves estimated with **frailtypack** and **R-INLA** in the simulations (Section 3).

## 4.1 GERCOR study

### 4.1.1 Description

It is a randomized clinical trial investigating two treatment strategies that included a total of 220 patients with metastatic colorectal cancer. The reference strategy (arm A) corresponds to FOLFIRI (irinotecan) followed by FOLFOX6 (oxaliplatin) while arm B involves the reverse sequence. Patients were randomly assigned from December 1997 to September 1999 and the date chose to assess overall survival was August 30, 2002. Complete data are available on 205 individuals for data analysis. Among them, 165 (80%) died during the follow-up. There are 1475 repeated measurements for the biomarker, 174 of which are zero values (12%). A summary of the dataset structure is given in Table 3. Our model uses death as the terminal event and the repeated SLD measurements (in centimeters) as the semicontinuous biomarker. Additional baseline covariates collected at the start of the study are also included, including performance status (0/1/2), lung metastatic site (Y/N), previous adjuvant radiotherapy (Y/N), previous surgery (no surgery/curative/palliative) and metastases (metachronous/synchronous). The first analysis of this dataset (Tournigand et al. (2004)) did not find any significant difference between the two treatment strategies using classic survival analysis methods (i.e. log-rank tests). A trivariate joint model has been applied to this study for the simultaneous analysis of the longitudinal SLD, recurrent events (progression of lesions not included in the SLD or new lesions) and the terminal event (Król et al. (2018)). A flexible mechanistic model using ordinary differential equation was proposed to fit the biomarker dynamics. The results shows a greater decline of the SLD for treatment arm A compared to treatment arm B. Moreover, the model finds a strong association between the biomarker model and the risk of terminal event. However the interpretation of this treatment effect is difficult due to the non-linear transformation of the outcome (Box-Cox) and the non-linear mechanistic model. Finally, a conditional two-part joint model was recently proposed (Rustand et al. (2020)), showing a significant treatment effect on the positive values of the biomarker (and no treatment effect on the probability of zero value). The model was able to show that when taking into account this treatment

effect on the biomarker, the risk of terminal event is not significantly different between the treatment arms. However the interpretation of the treatment effect on the biomarker value could have been impacted by a logarithm transformation used on the outcome. Instead, we use a GLM with log link function here for the continuous part of the biomarker.

#### 4.1.2 Results

As presented in Table 4, the fixed effects parameter estimates in the binary, continuous and survival parts are quite similar between the frequentist and Bayesian approaches, with a slightly lower variability for the parameters estimated with **R-INLA**. Treatment arm A is associated with a significant reduction in the SLD conditional on a positive value with **R-INLA** ($\hat{\beta}_1 = -0.33$, SE=0.07) and **frailtypack** ($\hat{\beta}_1 = -0.35$, SE=0.08), compared to treatment arm B. This is because the treatment by time interaction balances out the negative slope obtained, i.e., with **R-INLA** ($\hat{\beta}_9 = 0.30$, SE=0.10) and with **frailtypack** ($\hat{\beta}_9 = 0.33$, SE=0.10). Therefore, conditional on a positive value of the SLD, treatment arm A is associated with a reduction of $\sim 26\%$ of the SLD per year $(1 - \exp(-0.30))$ while treatment arm B is associated with no change over time. The hazard ratio of treatment arm B versus treatment arm A that evaluates the change in the risk of death was similar between **R-INLA** (HR=1.30, CI $0.84 - 1.92$) and **frailtypack** (HR=1.26, CI $0.85 - 1.79$). The main difference between **R-INLA** and **frailtypack** is in the estimation of the parameters for the association between the two-part model for the biomarker and the survival model. There is a positive and significant association between the random intercept ($\hat{\varphi}_a = 0.11, SE = 0.03$) from the binary part, the random intercept ($\hat{\varphi}_{b_0} = 0.66, SE = 0.07$) and the random slope ($\hat{\varphi}_{b_1} = 0.83, SE = 0.26$) from the continuous part and the risk of event with **R-INLA**. This association has a slightly lower effect size and much larger variability with **frailtypack** ($\hat{\varphi}_a = 0.13, SE = 0.13$, $\hat{\varphi}_{b_0} = 0.46, SE = 0.37$ and $\hat{\varphi}_{b_1} = 0.55, SE = 0.60$), so that the effects are not significant. This is in line with our simulation results (scenario 1) where the association structure was estimated with better precision with **R-INLA**. The computation time is much longer with **frailtypack** compared to **R-INLA**, the latter fits the data in less than a minute. The computation time for **frailtypack** increases quickly with the sample size and the model complexity (number of parameters and dimension of the random effects). The model was estimated in 60 minutes with **frailtypack** with 8 CPUs and this reduces to 10 minutes with 80 CPUs. The differences found in the association structure estimates is important when assessing the relationship between the biomarker dynamics and the risk of event. For instance, let's assume a clinician is interested in the top 15% patients who had the largest SLD increase during follow-up compared to the average patient. Their random effect $b_{1i}$ should be higher than 1 standard deviation, that is from Table 4, $b_{1i} > 0.51$ with **R-INLA** (respectively $b_{1i} > 0.52$ with **frailtypack**). Conditional on $b_{1i} > 0.51$ (respectively $b_{1i} > 0.52$), the mean values of the random effects can be derived by sampling from a conditional multivariate normal distribution with correlation matrix given in Table 4. These conditional means are 2.35, $-0.14$ and 0.77 for $a$, $b_0$ and $b_1$, respectively (2.25, $-0.20$ and 0.80 with **frailtypack**). Therefore, these top 15% individuals increase their chance to have the terminal event (i.e., to die) measured by an hazard ratio of HR $= \exp(0.11 * 2.35 + 0.66 * (-0.14) + 0.83 * 0.77) = \exp(0.81) =$ 2.24, CI=$1.46 - 3.51$, compared to a patient who has an average longitudinal SLD profile. **Frailtypack** underestimates this risk as we obtain HR $= \exp(0.13 * 2.25 + 0.46 * (-0.20) + 0.55 * 0.80) = \exp(0.64) = 1.90$, CI=$1.21 - 3.05$. The confidence intervals were obtained by sampling parameters from the Hessian matrix with **frailtypack** and the credible intervals from the posterior distribution of the parameters with **R-INLA**.

### 4.2 PRIME study

#### 4.2.1 Description

The Panitumumab Randomized Trial in Combination with Chemotherapy for Metastatic Colorectal Cancer to Determine Efficacy (PRIME) study is a more challenging application for fitting the TPJM because it includes information about the KRAS mutation status (exon 2 codons 12/13), which has been shown to relate to outcome with regard to treatment received in metastatic colorectal cancer patients (Van Cutsem et al. (2008); Normanno et al. (2009); Bokemeyer et al. (2008)). It is therefore an important risk modifier and clinicians are interested to assess treatment by mutation interaction in order to tailor treatment to patients' genetic risk (Marabelle et al. (2020)). This dataset is freely available on ProjectDataSphere.org.

Table 4: Application of the Bayesian and frequentist two-part joint models with shared random effects to the GERCOR study with the R packages **R-INLA** and **frailtypack**

| Package | | R-INLA Est. (SE) | frailtypack Est. (SE) |
|---|---|---|---|
| **Binary part** (SLD>0 versus SLD=0) | | | |
| intercept | $\alpha_0$ | 5.05*** (0.68) | 5.10*** (0.69) |
| time (year) | $\alpha_1$ | -2.11*** (0.39) | -2.14*** (0.41) |
| treatment (B/A) | $\alpha_2$ | -1.17 (0.69) | -1.24 (0.69) |
| PS (1 vs. 0) | $\alpha_3$ | 1.83** (0.57) | 1.97*** (0.58) |
| PS (2 vs. 0) | $\alpha_4$ | 1.72 (1.08) | 1.72 (1.18) |
| Previous_radio (Y/N) | $\alpha_5$ | 0.82 (0.70) | 0.85 (0.72) |
| Lung (Y/N) | $\alpha_6$ | 1.84** (0.67) | 2.14** (0.75) |
| time:treatment (B/A) | $\alpha_7$ | 0.30 (0.46) | 0.31 (0.47) |
| **Continuous part** ($E[Y_{ij}\|Y_{ij}>0]$) | | | |
| intercept | $\beta_0$ | 1.99*** (0.16) | 2.04*** (0.23) |
| time (years) | $\beta_1$ | -0.33*** (0.07) | -0.35*** (0.08) |
| treatment (B/A) | $\beta_2$ | -0.28** (0.10) | -0.35*** (0.10) |
| PS (1 vs. 0) | $\beta_3$ | 0.42*** (0.11) | 0.42*** (0.11) |
| PS (2 vs. 0) | $\beta_4$ | 0.53** (0.17) | 0.55** (0.17) |
| Previous_surgery (curative) | $\beta_5$ | -0.53** (0.20) | -0.61** (0.23) |
| Previous_surgery (palliative) | $\beta_6$ | 0.00 (0.15) | -0.02 (0.19) |
| Previous_radio (Y/N) | $\beta_7$ | -0.25* (0.12) | -0.22 (0.13) |
| Metastases (metachronous vs. synchronous) | $\beta_8$ | 0.43** (0.17) | 0.46** (0.16) |
| time:treatment (B/A) | $\beta_9$ | 0.30** (0.10) | 0.33** (0.10) |
| residual S.E. | $\sigma_\varepsilon$ | 0.39*** (0.00) | 0.42*** (0.01) |
| **Death risk** | | | |
| treatment (B/A) | $\gamma_1$ | 0.24 (0.21) | 0.21 (0.19) |
| PS (1 vs. 0) | $\gamma_2$ | 0.81*** (0.22) | 0.78*** (0.21) |
| PS (2 vs. 0) | $\gamma_3$ | 1.59*** (0.34) | 1.58*** (0.33) |
| Previous_surgery (curative) | $\gamma_4$ | -0.93* (0.42) | -0.97* (0.42) |
| Previous_surgery (palliative) | $\gamma_5$ | -0.51 (0.30) | -0.53 (0.30) |
| Metastases (metachronous vs. synchronous) | $\gamma_6$ | 0.95** (0.35) | 0.99** (0.34) |
| **Association** | | | |
| Intercept (binary part) | $\varphi_a$ | 0.11*** (0.03) | 0.13 (0.13) |
| Intercept (continuous part) | $\varphi_{b_0}$ | 0.66*** (0.07) | 0.46 (0.37) |
| Slope (continuous part) | $\varphi_{b_1}$ | 0.83** (0.26) | 0.55 (0.60) |
| **Random effects** | | | |
| intercept (binary part) | $\sigma_a$ | 2.67 | 2.81 |
| intercept (continuous part) | $\sigma_{b_0}$ | 0.67 | 0.71 |
| slope (continuous part) | $\sigma_{b_1}$ | 0.51 | 0.52 |
| | $corr_{ab_0}$ | 0.49 | 0.55 |
| | $corr_{ab_1}$ | 0.57 | 0.53 |
| | $corr_{b_0b_1}$ | -0.14 | -0.18 |
| **Computation time (Intel Xeon E5-4627 v4 2.60 GHz)** | | | |
| 8 CPUs | | < 1 minute | 60 minutes |
| 80 CPUs | | < 1 minute | 10 minutes |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The PRIME study is a randomized clinical trial that compares the efficacy and safety of panitumumab (anti-EGFR) in combination with FOLFOX4 (chemotherapy) with those of FOLFOX4 alone in the first-line treatment of patients, according to KRAS exon 2 status (Wild type or Mutant type). Between August 2006 and February 2008, 1183 patients were randomly assigned to receive treatment arm A (FOLFOX4 alone) or treatment arm B (panitumumab + FOLFOX4). The data for analysis includes a subset of 442 patients (i.e., 741 excluded from the publicly available dataset). There are 2372 repeated measurements of the SLD, 99 of which are zero values (4%). The small rate of zero measurements in the SLD distribution leads to a large variability in the binary part, however zeros corresponds to patients with a complete shrinkage of their target lesions, which is a very relevant information for clinicians about treatment effect. The number of individual repeated measurements for this biomarker varies between 1 and 24 with a median of 5. The death rate is 74%, corresponding to 328 deaths. Summary statistics of the dataset are given in Table 3. Additional baseline covariates collected at the start of the study are also included, including metastases to liver at study entry (Y/N), the number of baseline metastases sites (1/2/3/4+), age (<60/60-69/>=70) and baseline ECOG performance status (0/1/2). We used a global backward selection procedure for each component of the model to select the covariates to include in the final joint model. The conclusions of the study are presented in Douillard et al. (2013) and show the importance of taking into account the mutation status when assessing treatment effect. Among patients without mutated KRAS, treatment arm B was associated with a slightly significant reduced risk of death compared to treatment arm A. For patients with mutated KRAS, treatment arm B was associated with a non-significant increase in the risk of death compared

to treatment arm A.

### 4.2.2 Results

As presented in Table 5, in the binary part of the TPJM, the intercept is very large ($\hat{\alpha}_0 = 21.32$, SE=3.99), corresponding to a high probability of positive value at baseline. This probability is increased for patients with mutated KRAS ($\hat{\alpha}_3 = 5.63$, SE=5.73) and patients receiving treatment arm B ($\hat{\alpha}_2 = 2.41$, SE=4.70). The slope parameter with time is negative and significant ($\hat{\alpha}_1 = -10.43$, SE=1.95), meaning that patients without mutated KRAS and receiving treatment A have a higher odds of zero SLD value over time, i.e., complete response to treatment. This odds decreases, but not significantly, among patients with either mutated KRAS ($\hat{\alpha}_6 = 1.68$, SE=2.90) or receiving treatment B ($\hat{\alpha}_5 = 1.99$, SE=2.24) and is slightly attenuated in patients with both mutated KRAS and receiving treatment arm B ($\hat{\alpha}_7 = -1.38$, SE=5.32).

In the continuous part of the TPJM, patients with the wild type KRAS status and in treatment arm A are associated with a decrease in the SLD value over time conditional on a positive SLD value ($\hat{\beta}_1 = -1.71$, SE=0.09). This reduction of SLD over time is attenuated in patients with mutated KRAS ($\hat{\beta}_{12} = 1.22$, SE=0.14) or receiving treatment B ($\hat{\beta}_{11} = 0.98$, SE=0.13). However, this reduction is accentuated in patients with the KRAS mutation receiving treatment B ($\hat{\beta}_{13} = -1.16$, SE=0.20).

In the survival part, the model shows no significant difference between treatment arms for the risk of death ($\hat{\gamma}_1 = 0.10$, SE=0.16). Besides, patients with mutated KRAS have similar risk of death compared to patients with the wild type ($\hat{\gamma}_2 = 0.21$, SE=0.17), so do patients with mutated KRAS receiving treatment B ($\hat{\gamma}_3 = 0.04$, SE=0.23). The random effect from the binary part and the random slope from the continuous part are not associated to the risk of death ($\hat{\varphi}_a = 0.00$, SE=0.01 and $\hat{\varphi}_{b1} = 0.13$, SE=0.15) but the random intercept from the continuous part ($\hat{\varphi}_{b0} = 0.36$, SE=0.09) have a positive and highly significant association with the risk of event. This means that conditional on a positive value, the individual deviation from the mean baseline value of the SLD is predictive of the risk of event. Similarly to the GERCOR study, we can compare the top 15% patients with the smallest SLD at baseline to the average patient, their risk of death is reduced by 32% (HR=0.68, CI=$0.61 - 0.78$).

In conclusion, we did not find a direct effect of treatment B vs. A on the risk of death, in line with the initial study (Douillard et al. (2013)). However, the analysis of the continuous part of the TPJM suggests that the subgroup of patients with the KRAS mutation receiving treatment B, has a larger decrease of the SLD over time, which in turn is associated to lower risk of death (i.e., through the association with the random effects). Therefore suggesting a possible indirect association of treatment B vs. A on the risk of death.

# 5 Discussion

In this article, we developed a Bayesian estimation approach based on the INLA algorithm for two-part joint models for a longitudinal semicontinuous biomarker and a terminal event. We also provided a comparison with a frequentist alternative approach previously implemented into the **frailtypack** package, using small sample sizes as seen in cancer clinical trial evaluation. The frequentist estimation raises several limitations both in terms of model complexity and computation time. The Bayesian estimation proposed in the R package **R-INLA** has been recently introduced to fit complex joint models (Van Niekerk, Bakka, Rue, and Schenk (2019)) but to our knowledge, has never been proposed for TPJMs. Accounting for the semicontinuous nature of the biomarker, i.e. the SLD, and being able to fit joint models with more complex association structures between the biomarker and the terminal event, can be quite relevant in clinical applications by providing critical insights into the direct and indirect effect of a treatment on the event of interest. This was illustrated in our simulations and applications to two randomized cancer clinical trials.

In our simulation studies, the estimation with **R-INLA** was found superior to **frailtypack** in terms of computation time and precision of the fixed effects estimation. The point estimates from **frailtypack** yielded closer results to the true values of the random effects' standard deviations, the residual error term and the baseline hazard function than the posterior mean from **R-INLA**, even though **R-INLA** recovered all parameters well based on the estimated credible intervals.

Our first application to the GERCOR randomized clinical trial investigating two treatment lines to treat

Table 5: Application of the Bayesian two-part joint model with shared random effects to the PRIME study with the R package **R-INLA**

| Package | | R-INLA Est. (SE) |
|---|---|---|
| **Binary part** (SLD>0 versus SLD=0) | | |
| intercept | $\alpha_0$ | 21.32*** (3.99) |
| time (year) | $\alpha_1$ | -10.43*** (1.95) |
| treatment (B/A) | $\alpha_2$ | 2.41 (4.70) |
| kras (MT/WT) | $\alpha_3$ | 5.63 (5.73) |
| treatment (B/A):kras (MT/WT) | $\alpha_4$ | 3.66 (8.43) |
| time:treatment (B/A) | $\alpha_5$ | 1.99 (2.24) |
| time:kras (MT/WT) | $\alpha_6$ | 1.68 (2.90) |
| time:treatment (B/A):kras (MT/WT) | $\alpha_7$ | -1.38 (5.32) |
| **Continuous part** $E[Y_{ij}\|Y_{ij} > 0]$ | | |
| intercept | $\beta_0$ | 2.41*** (0.17) |
| time (years) | $\beta_1$ | -1.71*** (0.09) |
| treatment (B/A) | $\beta_2$ | -0.17 (0.10) |
| kras (MT/WT) | $\beta_3$ | -0.20 (0.11) |
| liver metastases (Y/N) | $\beta_4$ | 0.63*** (0.14) |
| ECOG (symptoms but ambulatory vs. fully active) | $\beta_5$ | 0.19* (0.08) |
| ECOG (in bed less than 50% of the time vs. fully active) | $\beta_6$ | 0.52** (0.18) |
| baseline metastases sites (2 vs. 1) | $\beta_7$ | 0.08 (0.12) |
| baseline metastases sites (3 vs. 1) | $\beta_8$ | 0.26* (0.12) |
| baseline metastases sites (4+ vs. 1) | $\beta_9$ | 0.19 (0.12) |
| treatment (B/A):kras (MT/WT) | $\beta_{10}$ | 0.19 (0.15) |
| time:treatment (B/A) | $\beta_{11}$ | 0.98*** (0.13) |
| time:kras (MT/WT) | $\beta_{12}$ | 1.22*** (0.14) |
| time:treatment (B/A):kras (MT/WT) | $\beta_{13}$ | -1.16*** (0.20) |
| residual S.E. | $\sigma_\varepsilon$ | 0.31 (0.01) |
| **Death risk** | | |
| treatment (B/A) | $\gamma_1$ | 0.10 (0.16) |
| kras (MT/WT) | $\gamma_2$ | 0.21 (0.17) |
| treatment (B/A):kras (MT/WT) | $\gamma_3$ | 0.04 (0.23) |
| age (60-69 vs. <60) | $\gamma_4$ | 0.08 (0.13) |
| age (70+ vs. <60) | $\gamma_5$ | 0.22 (0.14) |
| liver metastases (Y/N) | $\gamma_6$ | 0.03 (0.23) |
| ECOG (symptoms but ambulatory vs. fully active) | $\gamma_7$ | 0.30** (0.12) |
| ECOG (in bed less than 50% of the time vs. fully active) | $\gamma_8$ | 0.81** (0.26) |
| baseline metastases sites (2 vs. 1) | $\gamma_9$ | 0.12 (0.20) |
| baseline metastases sites (3 vs. 1) | $\gamma_{10}$ | 0.32 (0.20) |
| baseline metastases sites (4+ vs. 1) | $\gamma_{11}$ | 0.43* (0.21) |
| **Association** | | |
| Intercept (binary part) | $\varphi_a$ | 0.00 (0.01) |
| Intercept (continuous part) | $\varphi_{b_0}$ | 0.36*** (0.09) |
| Slope (continuous part) | $\varphi_{b_1}$ | 0.13 (0.15) |
| **Random effects** | | |
| intercept (binary part) | $\sigma_a$ | 11.17 |
| intercept (continuous part) | $\sigma_{b_0}$ | 0.74 |
| slope (continuous part) | $\sigma_{b_1}$ | 0.74 |
| | $corr_{ab_0}$ | 0.03 |
| | $corr_{ab_1}$ | 0.84 |
| | $corr_{b_0b_1}$ | -0.13 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

metastatic colorectal cancer shows some differences between the two estimation approaches. In line with our simulations, the variability of the parameter estimates is reduced with **R-INLA**, in particular for the association parameters between the biomarker and the survival outcome where **R-INLA** concluded to a strong association unlike **frailtypack**, which found a non-significant association with an attenuated effect size. The second application to the PRIME study reflects upon the concept that treatment response might depend on genetic alterations or tumor biomarker status (DNA/RNA/protein features). There is now a great interest in identifying subgroups of patients with specific patterns of responses however most methods provide an average effect of covariates. Instead, our model can distinguish complete responders (i.e. SLD=0) from partial responders (i.e. SLD >0). This leads also to an increase in model complexity as additional covariates and random effects are included in each submodel of the TPJM. The frequentist approach proposed in **frailtypack** can have convergence issues or sometimes cannot be fitted at all as this was the case for the PRIME study. Only the Bayesian approach could be used in that situation. Interestingly, the analysis of the continuous part of the TPJM suggested that the subgroup of patients with the KRAS mutation receiving treatment B, has a much larger decrease of the SLD over time (compared to the KRAS mutation group alone or treatment B alone), which in turn is associated to lower risk of death (i.e., through the association

with the random effects). Therefore suggesting a possible indirect association of treatment B vs. A on the risk of death. To our knowledge, this subgroup was not initially identified in the original study and could motivate further investigations of the interaction of KRAS mutation and anti-EGFR therapies to treat advanced colorectal cancer. Clinicians are also including additional information on other tumour gene mutations status that we could also include in future analyses.

Our work has several limitations. Our applications focused on clinical trials of very advanced cancers, which often have high death rates and small proportions of complete responses (i.e. SLD=0). In situations where we have a higher proportion of complete responders, the relative performances of **R-INLA** vs. **frailtypack** could be different. The conclusions might be different for different settings (i.e. with higher zero rate and reduced censoring). For instance a meta-analysis evaluating the responses among non-Hodgkin's Lymphoma patients estimated complete response rates (i.e. SLD= 0) ranging from 1.2% to 84% in the different pooled clinical trials (Mangal et al. (2018)). We also notice that the two models estimated with **R-INLA** and **frailtypack** are not completly comparable because of the difference in the approximation of the baseline hazard function. Besides the shared random effects, other association structures have also been proposed such as the current value association, i.e., it uses the current level of the biomarker, and is available in **frailtypack**. For the TPJMs, the current value of the biomarker is defined as $\mathrm{E}[Y_{ij}] = Prob(Y_{ij} > 0)\mathrm{E}[Y_{ij}|Y_{ij} > 0]$, which is non linear. It cannot be directly defined as part of the latent gaussian model and more work is warranted to be included in **R-INLA**. It would be also interesting to consider a Bayesian development for the marginal TPJM we recently proposed (Rustand, Briollais, and Rondeau (2020)). Finally, the definition of the hyperparameter prior distributions are an important aspect of Bayesian estimation. In this work, the PC priors provided a general setting for the priors since they provide a natural avenue to incorporate knowledge from the practitioner about the expected size of the parameter and they are constructed to be proper and avoid overfitting. Alternative prior choices for the hyperparameters can be used in **R-INLA** if the practitioner posseses motivation for it from expert or prior knowledge.

The reduction in the computation times with **R-INLA** was beyond our expectations. It improves drastically the applicability of the Bayesian estimation for complex models such as the TPJMs and other families of joint models, such as a bivariate joint model for recurrent events and a terminal event or a trivariate joint model for a longitudinal biomarker, recurrent events and a terminal event, which are currently availble in **frailtypack**. Finally, **R-INLA** can accommodate multiple longitudinal outcomes while **frailtypack** is currently limited to a single longitudinal outcome.

## Acknowledgements

This publication is based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere, LLC. Neither Project Data Sphere, LLC nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this publication.

## References

Andrinopoulou, E.-R. and D. Rizopoulos (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statistics in medicine 35*(26), 4813–4823.

Bokemeyer, C., I. Bondarenko, J. Hartmann, F. De Braud, C. Volovat, J. Nippgen, C. Stroh, I. Celik, and P. Koralewski (2008). Kras status and efficacy of first-line treatment of patients with metastatic colorectal cancer (mcrc) with folfox with or without cetuximab: The opus experience. *Journal of Clinical Oncology 26*(15_suppl), 4000–4000.

Chi, Y.-Y. and J. G. Ibrahim (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics 62*(2), 432–445.

Douillard, J.-Y., K. S. Oliner, S. Siena, J. Tabernero, R. Burkes, M. Barugel, Y. Humblet, G. Bodoky,

D. Cunningham, J. Jassem, et al. (2013). Panitumumab–folfox4 treatment and ras mutations in colorectal cancer. *New England Journal of Medicine 369*(11), 1023–1034.

Hanson, T. E., A. J. Branscum, and W. O. Johnson (2011). Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime data analysis 17*(1), 3–28.

Henderson, R., P. Diggle, and A. Dobson (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics 1*(4), 465–480.

Hespanhol, L., C. S. Vallio, L. M. Costa, and B. T. Saragiotto (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian journal of physical therapy 23*(4), 290–301.

Król, A., C. Tournigand, S. Michiels, and V. Rondeau (2018). Multivariate joint frailty model for the analysis of nonlinear tumor kinetics and dynamic predictions of death. *Statistics in Medicine 37*(13), 2148–2161.

Lawrence Gould, A., M. E. Boye, M. J. Crowther, J. G. Ibrahim, G. Quartey, S. Micallef, and F. Y. Bois (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the dia bayesian joint modeling working group. *Statistics in medicine 34*(14), 2181–2195.

Mangal, N., A. H. Salem, M. Li, R. Menon, and K. J. Freise (2018). Relationship between response rates and median progression-free survival in non-hodgkin's lymphoma: A meta-analysis of published clinical trials. *Hematological Oncology 36*(1), 37–43.

Marabelle, A., M. Fakih, J. Lopez, M. Shah, R. Shapira-Frommer, K. Nakagawa, H. C. Chung, H. L. Kindler, J. A. Lopez-Martin, W. H. Miller Jr, et al. (2020). Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 keynote-158 study. *The Lancet Oncology 21*(10), 1353–1365.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics 11*(2), 431–441.

Martino, S., R. Akerkar, and H. Rue (2011). Approximate bayesian inference for survival models. *Scandinavian Journal of Statistics 38*(3), 514–528.

Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2013). Bayesian computing with inla: new features. *Computational Statistics & Data Analysis 67*, 68–83.

Muth, C., Z. Oravecz, and J. Gabry (2018). User-friendly bayesian regression modeling: A tutorial with rstanarm and shinystan. *Quantitative Methods for Psychology 14*(2), 99–119.

Normanno, N., S. Tejpar, F. Morgillo, A. De Luca, E. Van Cutsem, and F. Ciardiello (2009). Implications for kras status and egfr-targeted therapies in metastatic crc. *Nature reviews Clinical oncology 6*(9), 519.

R. Brown, E. and J. G. Ibrahim (2003). A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics 59*(2), 221–228.

Rizopoulos, D. et al. (2016). The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software 72*(i07).

Rizopoulos, D. and P. Ghosh (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine 30*(12), 1366–1380.

Rue, H. and L. Held (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Rue, H., S. Martino, and N. Chopin (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.

Rue, H., A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren (2017). Bayesian computing with inla: A review. *Annual Review of Statistics and Its Application 4*(1), 395–421.

Rustand, D., L. Briollais, and V. Rondeau (2020). A marginal two-part joint model for a longitudinal biomarker and a terminal event with application to advanced head and neck cancers. (Under submission).

Rustand, D., L. Briollais, C. Tournigand, and V. Rondeau (2020). Two-part joint model for a longitudinal semicontinuous marker and a terminal event with application to metastatic colorectal cancer data. *Biostatistics*. kxaa012.

Schenk, O. and K. Gärtner (2004). Solving unsymmetric sparse systems of linear equations with pardiso. *Future Generation Computer Systems 20*(3), 475–487.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, S. H. Sørbye, et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical science 32*(1), 1–28.

Smith, V. A., J. S. Preisser, B. Neelon, and M. L. Maciejewski (2014). A marginalized two-part model for semicontinuous data. *Statistics in Medicine 33*(28), 4891–4903.

Song, X., M. Davidian, and A. A. Tsiatis (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics 58*(4), 742–753.

Tournigand, C., T. André, E. Achille, G. Lledo, M. Flesh, D. Mery-Mignard, E. Quinaux, C. Couteau, M. Buyse, G. Ganem, B. Landi, P. Colin, C. Louvet, and A. de Gramont (2004). Folfiri followed by folfox6 or the reverse sequence in advanced colorectal cancer: A randomized gercor study. *Journal of Clinical Oncology 22*(2), 229–237.

Van Cutsem, E., I. Lang, G. D'haens, V. Moiseyenko, J. Zaluski, G. Folprecht, S. Tejpar, O. Kisker, C. Stroh, and P. Rougier (2008). Kras status and efficacy in the first-line treatment of patients with metastatic colorectal cancer (mcrc) treated with folfiri with or without cetuximab: The crystal experience. *Journal of Clinical Oncology 26*(15_suppl), 2–2.

Van Niekerk, J., H. Bakka, and H. Rue (2019). Joint models as latent gaussian models-not reinventing the wheel. *arXiv:1901.09365*.

Van Niekerk, J., H. Bakka, and H. Rue (2020). Stable non-linear generalized bayesian joint models for survival-longitudinal data. *Sankhya A (Accepted)*.

Van Niekerk, J., H. Bakka, H. Rue, and L. Schenk (2019). New frontiers in bayesian modeling using the inla package in r. *arXiv:1907.10426*.

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

Xu, J. and S. L. Zeger (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 50*(3), 375–387.