Received: 21 August 2020 | Revised: 18 June 2021 | Accepted: 22 June 2021

DOI: 10.1002/bimj.202000253

Biometrical Journal

RESEARCH ARTICLE

Robust joint modelling of longitudinal and survival data: Incorporating a time-varying degrees-of-freedom parameter

Lisa M. McFetridge[1] | **Özgür Asar**[2] | **Jonas Wallin**[3]

[1] Mathematical Sciences Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast, UK

[2] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Acıbadem Mehmet Ali Aydınlar University, İstanbul, Turkey

[3] Department of Statistics, Lund University, Lund, Sweden

Correspondence
Özgür Asar, Department of Biostatistics and Medical Informatics, Faculty of Medicine, Acıbadem Mehmet Ali Aydınlar University, 34752 İstanbul, Turkey.
Email: ozgur.asar@acibadem.edu.tr, ozgurasarstat@gmail.com

L. McFetridge and Ö. Asar are joint first authors.

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: Reference: EP/P026028/1

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

Monitoring of individual biomarkers has the potential of explaining the hazard of survival outcomes. In practice, these measurements are intermittently observed and are known to be subject to substantial measurement error. Joint modelling of longitudinal and survival data enables us to associate intermittently measured error-prone biomarkers with risks of survival outcomes and thus plays an important role in the analysis of medical data. Most of the joint models available in the literature have been built on the Gaussian assumption. This makes them sensitive to outliers. In this work, we study a range of robust models to address this issue. Of particular interest is the common occurrence in medical data that outliers can occur with different frequencies over time, for example, in the period when patients adjust to treatment changes. Motivated by the analysis of data gathered from patients with primary biliary cirrhosis, a new model with a time-varying robustness is introduced. Through both the motivating example and a simulation study, this research not only stresses the need to account for longitudinal outliers in the analysis of medical data and in joint modelling research but also highlights the bias and inefficiency from not properly estimating the degrees-of-freedom parameter. This work presents a number of methods in addition to the time-varying robustness, and each method can be fitted using the `R` package `robjm`.

KEYWORDS

degrees-of-freedom, longitudinal outliers, normal variance mixtures, robust joint model, *t*-distribution

1 | INTRODUCTION

Biomarkers are used as proxies of one's health and are known to be subject to substantial measurement error, due to biological and non-biological sources. In prospective studies, the biomarkers are repeatedly measured. These measurements are made at intermittent time points in practice rather than continuously, and times elapsed between successive measurements are typically unequal. It is of scientific interest to explain risks of survival outcomes by repeated measures of biomarker data. Two main obstacles are the intermittent nature of the biomarker data and inherent measurement error.

Joint modelling of longitudinal and survival outcomes enables us to associate repeated measures of biomarkers with the risks of survival events, while taking into account the aforementioned obstacles. The framework typically combines two sub-models: a mixed-effects model for repeated measures and a Cox model for survival data. The sub-models are linked with shared parameters. Reviews of relevant literature can be found in Tsiatis and Davidian (2004), Diggle et al. (2008), Rizopoulos (2012), McCrink et al. (2013), Asar et al. (2015) and Elashoff (2017).

The models introduced in this paper are motivated by the analysis of patients with liver cirrhosis. Utilising the Mayo Clinic Primary Biliary Cirrhosis (PBC) dataset, data are analysed from patients with PBC who participated in a randomised placebo-controlled trial of the drug D-penicillamine. Relevant biomarkers are observed over the study period until patients die or are censored. The link between the longitudinal and survival outcomes makes the data ideal for analysis via joint modelling.

However, a prevailing assumption in the literature is that both random-effects and error terms in the mixed effects sub-model follow Normal distributions. In most real-life problems, including the motivating example in this paper, the datasets include outliers for which the Normal assumption might be inadequate. Specifically, in prospective studies, two types of outliers may be present (Pinheiro et al., 2001):

(i) *b-outliers*: Outlying individuals within the population who do not conform to population trends. These are outliers within the longitudinal random effects.
(ii) *e-outliers*: Outlying observations within an individual's set of measurements that do not follow the individual's own trend over time. These are outliers within the longitudinal random error.

Only in recent years, studies have been undertaken to investigate the negative impacts that the presence of longitudinal outliers under the Normal assumption can have (Asar et al., 2021; Baghfalaki et al., 2013; Huang et al., 2010; Li et al., 2009; McCrink, 2014). Robust joint models replace the Normality assumption with *t*-distributional assumptions for the random terms. In doing so, the heavier tails down-weigh the detrimental impact of longitudinal outliers within the joint modelling framework. Initial research on robust joint modelling considered only the *e*-outliers (Huang et al., 2010; Li et al., 2009; Taylor et al., 2013) and both *e*- and *b*-outliers (Asar et al., 2021; Baghfalaki et al., 2014; McCrink, 2014; Song et al., 2012). Each of these studies found that the parameter estimates and corresponding standard errors are sensitive to the Normality assumptions when outliers are present, with *t*-distributional assumptions alleviating such bias. If utilised when outliers are not present, unbiased estimates are obtained, though Li et al. (2009) and Huang et al. (2010) noted that the robust joint models give slightly higher standard errors for the longitudinal parameters. This initial work, however, restrictively fixes the degrees-of-freedom at a constant chosen by the user, an assumption alleviated by Baghfalaki et al. (2013), Asar et al. (2021) and Baghfalaki et al. (2014), who utilised Bayesian approaches, and McCrink (2014), who utilised a frequentist approach, to allow estimation of the degrees-of-freedom to be dictated by the data. With estimated degrees of freedom, the *t*-distributional assumptions of the robust joint model will approximate normality in the absence of outliers.

All of the aforementioned research, however, assumes that the degrees-of-freedom is constant, unchanging over time, an assumption prevalent in the literature for robust mixed models alongside their robust joint model counterparts. For the PBC data that would therefore restrict the models to assuming that patients with highly fluctuating alkaline phosphatase levels (*e*-outliers) after starting the new drug do not stabilise during the period of observation and are prone to the same extent of outliers throughout the study, an unlikely assumption as illustrated in Figure 1.

To address this issue, we propose a robust joint model which can account for the situation where the impact of outliers can change over time. Varying frequency of outliers is a very likely scenario, for example, when patients are given new treatments, they typically take time to stabilise and adjust to these treatments. In medicine, this is known as the 'heterogeneity of treatment effects' and is an important concept in clinical practice that should therefore be reflected in the statistical literature and methodology (Kravitz et al., 2014). This period of adjustment means that patients are more prone to demonstrating responses which outlie from the expected and thus the patients' measurements may not be consistent with the population average and change over time as they grow accustomed to the new treatments. Incorrectly modelling such scenarios limits our ability to fully decipher the relationship between how patients' responses change over time, and the impact this has on their risk of an adverse event.

In addition to the introduction of this new methodology for time-varying degrees-of-freedom, this paper also investigates and contrasts this new model with the three time-invariant robust joint modelling approaches currently found within the literature. These time-invariant approaches, described in full in Section 3, impose different restrictive assumptions on the relationship between parameters. The impact of these various assumptions and, consequently of not properly estimating the degrees-of-freedom parameter, is explored through a simulation study. To the knowledge of the authors, no work

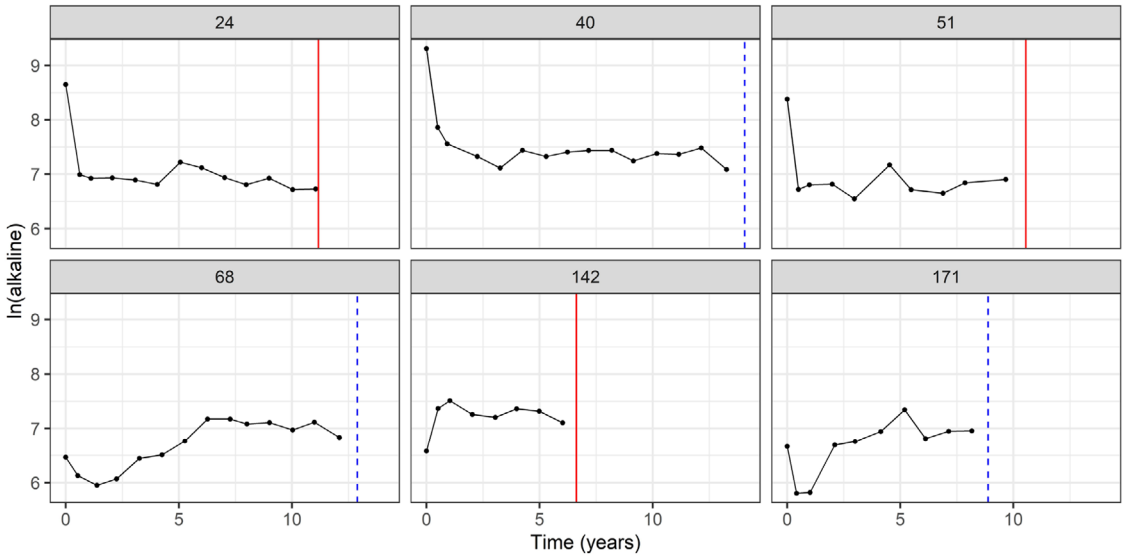


FIGURE 1 Longitudinal profile plots for six individuals from the PBC data. The observed data are given in black. The solid red line indicates the time of death, and the dashed blue line indicates the time of censoring

has been undertaken to date that has contrasted these time-invariant approaches in a simulation study. The potentially detrimental impact of the different assumptions is then demonstrated in the analysis of the PBC data via the different robust joint modelling approaches under investigation in this paper.

A more detailed discussion on the motivational example is given in Section 2. The rest of this paper is organised as follows. In Section 3, the methodology for the various joint modelling approaches investigated in the paper is presented, including the introduction of a robust joint model with time-varying degrees-of-freedom. Section 4 presents a simulation study comparing the approaches assuming an underlying structure of time-varying degrees-of-freedom. Analysis of the motivating example is presented in Section 5 followed by discussions in Section 6.

2 | MOTIVATING EXAMPLE: PBC DATA

The motivating example arises from data collected from 312 patients with primary biliary cirrhosis , a rare autoimmune disease which results in increased levels of alkaline phosphatase in the blood and leads to cirrhosis of the liver and fatality. This association with patients' alkaline levels and survival lends these data to the use of joint models. As illustrated in Figure 1, alkaline levels tend to fluctuate and differ across the population in their patient-specific trends over time and thus there is a great potential that PBC patients will either have outlying observations from their own trends (e-outliers) or be an outlier themselves from the population (b-outlier).

The PBC dataset was collected from the Mayo Clinic trial conducted between 1974 and 1984, which aimed to assess the effectiveness of the D-penicillamine drug (Murtaugh et al., 1994). Of the 312 patients, 158 were randomly assigned to the drug and 154 to placebo. Median baseline age was 48.8 years (minimum = 26.28, maximum = 78.44). In total, 1885 repeated alkaline observations were collected over the 10-year trial period, with a median of 5 (minimum = 1, maximum = 16) observations per individual. Individuals were observed for a minimum of 0.1 years, and a maximum of 14.3 years. One hundred forty (45%) patients died during the study period with the remaining patients being right censored.

Figures 2 and 3 show the tail behaviour of the standardised conditional residuals and random effects, respectively, based on a joint model fitted for the PBC data with Gaussian assumptions for both random effects and error terms. Degrees-of-freedom estimates for Figure 2 were obtained by fitting a univariate t distribution such that $t(\mu, \sigma, \tau)$, where μ is the location parameter, σ the scale, τ the degrees-of-freedom. Time (in years) was divided into bins based on 20% percentiles of the follow-up time. The quantile–quantile plot for the standardised conditional residuals (left panel of Figure 2) indicates heavier tails than Gaussian. Figure 2 further indicates that the degrees-of-freedom parameter of the t-distribution for the residuals varies over time, an anticipated result based on the known fluctuations in alkaline levels for PBC patients adjusting to the new treatment. Such changes over time will be accounted for through the time-varying degrees-of-freedom

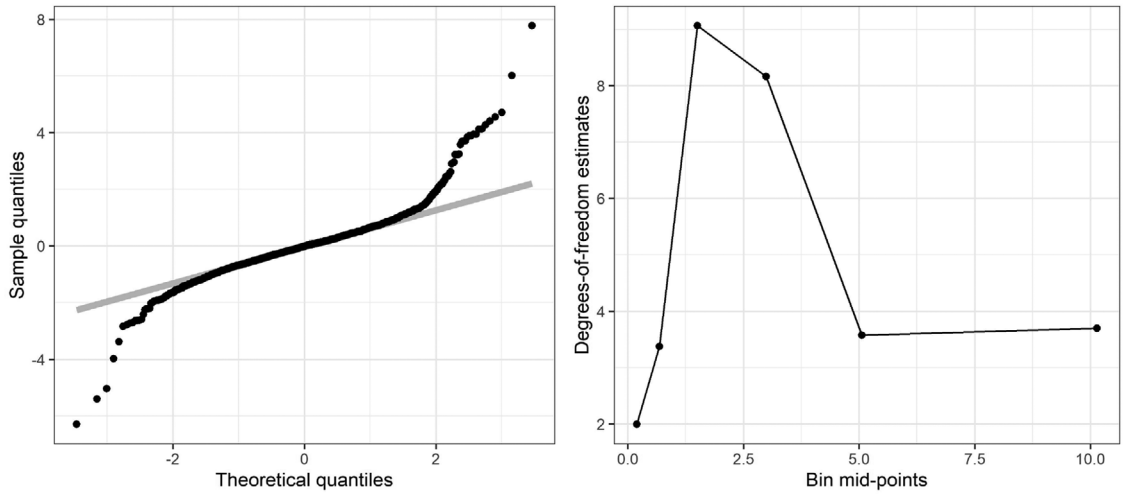


FIGURE 2 Plots to inspect the standardised conditional residuals obtained from the joint model with Gaussian assumptions for the PBC dataset. Details of calculations are provided in Section 2. Left: The quantile–quantile plot against standard Normal distribution; right: degrees-of-freedom parameter estimates

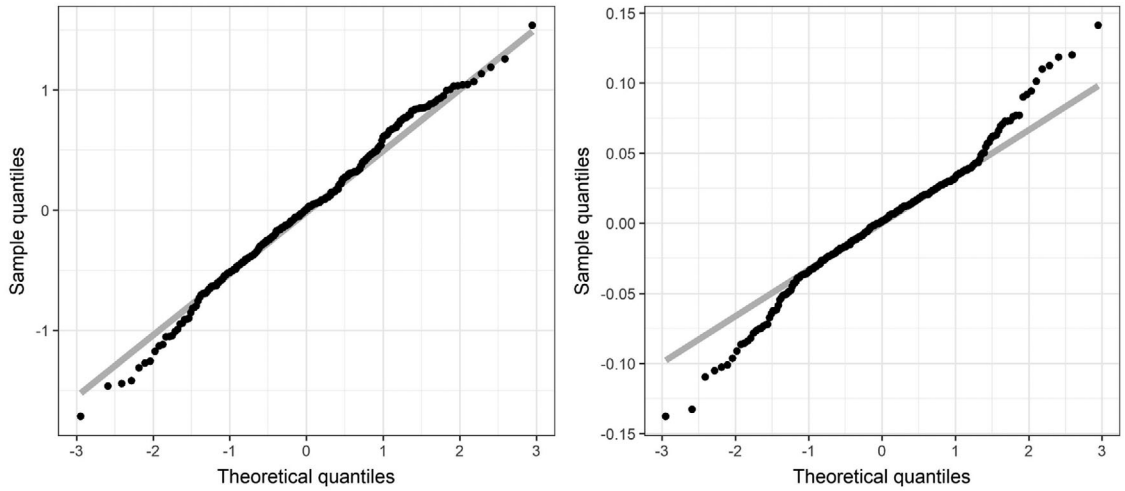


FIGURE 3 Plots to inspect the distributions of the random-intercept (left) and random-slope (right) terms against standard Normal for the PBC dataset

formulation (Approach 4) introduced in Section 3.4. Figure 3 indicates that normality may be a reasonable assumption for the random effects.

3 | APPROACHES FOR ROBUST JOINT MODELLING

3.1 | Notation

Before introducing the models, we present the general notation used throughout the paper. Let Y_{ij} denote the jth ($j \in \{1, \ldots, m_i\}$) repeated measurement belonging to subject ith ($i \in \{1, \ldots, n\}$) collected at time t_{ij}. We denote $t_i = \{t_{i1}, \ldots, t_{im_i}\}$ the set of the follow-up times at which Y_{ij}s are collected, $a_i = \{a_{i1}, \ldots, a_{ik}\}$ baseline covariate information and S_i survival time. Here, S_i is subject to right censoring and defined as $S_i = \min(S_i^*, C_i)$, where S_i^* is the true survival time and C_i censoring time for subject i. C_i is defined as $C_i = \min(C, D_i)$, where C is the study end-time and D_i is the drop-out time for subject i. Hence, to complete the survival information, we introduce an additional random variable, E_i, defined by $E_i = I(S_i^* \leq C_i)$, where $I(\cdot)$ being an indicator function.

3.2 | Joint modelling of longitudinal and survival data

The framework for the so-called joint models for longitudinal and survival data under the shared-parameter paradigm is given by

$$\begin{aligned} Y_{ij} &= Y_i^*(t_{ij}) + Z_{ij}, \\ &= \boldsymbol{x}_{ij}^\top \boldsymbol{\alpha} + \boldsymbol{d}_{ij}^\top \boldsymbol{B}_i + Z_{ij}, \end{aligned} \tag{1}$$

$$h_i(t) = h_0(t) \exp\left(\boldsymbol{c}_i^\top \boldsymbol{\omega} + f(\mathcal{Y}_i^*(t); \boldsymbol{\eta})\right). \tag{2}$$

The framework allows for biomarker values (the longitudinal data) to be measured with error, that is, the observed data, Y_{ij}, is composed of underlying continuous-time signal at time t_{ij}, $Y_i^*(t_{ij})$, and noise, Z_{ij}. The signal is de-composed into fixed effects, $\boldsymbol{x}_{ij}^\top \boldsymbol{\alpha}$, and random effects, $\boldsymbol{d}_{ij}^\top \boldsymbol{B}_i$. Here, $\boldsymbol{x}_{ij}$ is a $p \times 1$ matrix structured by a_i and t_i. $\boldsymbol{\alpha}$ is a $p \times 1$ matrix of regression coefficients as in multiple linear regression. $\boldsymbol{d}_{ij}$ is a $q \times 1$ matrix, that is typically a subset of $\boldsymbol{x}_{ij}$. $\boldsymbol{B}_i$ are subject-specific coefficients that take into account heterogeneity between subjects. $h_i(t)$ is the hazard of survival event for subject i at time t. $h_0(t)$ is the baseline hazard that can be specified using hazard functions of parametric distributions, for example Weibull; left un-specified as in Cox (1972); assumed to be piece-wise constant; or be expressed in terms of splines, for example, natural cubic or B-splines. $\boldsymbol{c}_i$ is a $g \times 1$ matrix with elements from a_i and $\boldsymbol{\omega}$ is a $g \times 1$ vector of regression coefficients. $f(\mathcal{Y}_i^*(t); \boldsymbol{\eta})$ is the term for taking into account the association between hazard of survival event and features of a biomarker process, with $f(\cdot)$ being a known function. A popular choice is to use the current value parametrisation such that $f(\mathcal{Y}_i^*(t); \boldsymbol{\eta}) = \eta Y_i^*(t)$. For other choices, see Hickey et al. (2016).

The classical model assumptions for the random effects and error components are zero-mean Gaussian distributions such that

$$\begin{aligned} \mathbf{B}_i &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\ Z_{ij} &\sim \mathcal{N}\left(0, \sigma^2\right), \\ \mathbf{B}_i \perp Z_{ij},\ & Z_{ij} \perp Z_{ij'} \text{ for } j \neq j', \end{aligned}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the random effects and σ is the standard deviation of the measurement error. In what follows, we extend the assumptions beyond the Gaussian.

3.3 | Robust joint modelling

The robustness of the joint model will be determined by the distributions of $\mathbf{B}_i$ and Z_{ij}. The tail of the density for $\mathbf{B}_i$ determines robustness against b-outliers, and the tail of the density for Z_{ij} determines the robustness against e-outliers. We consider symmetric robust distributions for both the random-effects and error components through normal-variance mixtures such that

$$\begin{aligned} \mathbf{B}_i &= \boldsymbol{\Sigma}^{1/2} \sqrt{V_i^B} \mathbf{B}_i^*, \\ Z_{ij} &= \sigma \sqrt{V_{ij}^Z} Z_{ij}^*, \end{aligned}$$

where $\mathbf{B}_i^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{q\times q})$, $Z_{ij}^* \sim \mathcal{N}(0,1)$ and $\mathbf{B}_i^* \perp Z_{ij}^*$. This formulation is flexible and includes widely used distributions as special cases (Asar et al., 2020). The tail behaviour is determined by the distribution of V_i^B and V_{ij}^Z. The special case of $V_i^B = 1$ and $V_{ij}^Z = 1$ recovers the Gaussian joint model (see Section 3.2). In this study, we specifically consider inverse Gamma distribution ($\mathcal{IG}$) with equal shape and scale parameters for the V terms which results a t-distribution.

Ideally, one can use the posterior distribution of V_i^B or V_{ij}^Z to detect outliers. If the posterior distribution of V_i^B (V_{ij}^Z) is concentrated at large values, this should indicate a b-outlier for subject i (e-outlier for observation j of patient i). However, in much of the robust modelling literature dependence between V_i^B and V_{ij}^Z, for example, see Pinheiro et al. (2001), and V_{ij}^Z and $V_{ij'}^Z$, for example, Baghfalaki et al. (2013), have been introduced (mainly to simplify the inferential procedure). These dependencies, however, make the aforementioned interpretation on outlier detection impossible.

In what follows, we will set three robust joint model formulations that are available in the literature and discuss implications. We then will propose a new approach to the joint modelling with time-varying degrees-of-freedom parameter for the error term.

3.3.1 | Approach 1

McCrink (2014) considered

$$V_i^B = V_i,$$
$$V_{ij}^Z = V_i,$$
$$V_i \sim \mathcal{IG}(\gamma/2, \gamma/2).$$

The mixing variable is equal for both random effects and error components. Note that under this assumption $\mathbf{B}_i$ and $\mathbf{Z}_i$ will be jointly multivariate t with the single degrees-of-freedom parameter, γ, and the properties, $\mathbf{B}_i \perp Z_{ij}$, and $Z_{ij} \perp Z_{ij'}$ for $j \neq j'$, do not hold. These properties only hold conditionally such that $\mathbf{B}_i$ and Z_{ij} and Z_{ij} and $Z_{ij'}$ are conditionally independent given V_i, due to the following specifications: $\mathbf{B}_i^* \perp Z_{ij}^*$ and $Z_{ij}^* \perp Z_{ij'}^*$.

3.3.2 | Approach 2

Baghfalaki et al. (2013) considered a robust joint model by setting

$$V_i^B = V_i^B \sim \mathcal{IG}(\phi/2, \phi/2),$$
$$V_{ij}^Z = V_i^Z \sim \mathcal{IG}(\delta/2, \delta/2).$$

Under this approach, $\mathbf{B}_i$ and $\mathbf{Z}_i$ will be multivariate t, with separate degrees-of-freedom parameters, ϕ and δ, respectively; the property of $\mathbf{B}_i \perp Z_{ij}$ holds, whereas $Z_{ij} \perp Z_{ij'}$ for $j \neq j'$ does not. As Z_{ij} and $Z_{ij'}$ (for $j \neq j'$) share a common V_i^Z, this approach can be seen as a random-effects approach on the variance of Z. Note that Z_{ij} and $Z_{ij'}$ are conditionally independent given V_i^Z.

3.3.3 | Approach 3

Asar et al. (2021) considered the setting of

$$V_i^B = V_i^B \sim \mathcal{IG}(\phi/2, \phi/2),$$
$$V_{ij}^Z = V_{ij}^Z \sim \mathcal{IG}(\delta/2, \delta/2),$$

for joint modelling. Under this approach, the dependence between Z_{ij} and $Z_{ij'}$ that was present in Approach 2 has been removed. This is the most natural approach compared to the previous two, as all of the outliers are decoupled such that $\mathbf{B}_i \perp Z_{ij}$ and $Z_{ij} \perp Z_{ij'}$ and will be used as a base for our time-varying degrees-of-freedom formulation.

3.4 | Approach 4: Time-varying degrees-of-freedom formulation

By building on Approach 3, we consider time-varying degrees-of-freedom parameter, $\delta(t)$, for Z_{ij}. In practice, we need to discretise time, that is, $\delta(t) = \delta(t_{ij})$. This approach considers

$$V_{ij}^Z \sim \mathcal{IG}(\delta_{ij}/2, \delta_{ij}/2),$$
$$\delta_{ij} = \exp(\delta_0 + \boldsymbol{a}_{ij}^\top \boldsymbol{\beta}).$$

Here, $\boldsymbol{a}_{ij}$ is a $s \times 1$ vector to be specified by natural cubic splines or B-splines, with $\boldsymbol{\beta}$ associated coefficients. Note that one obtains time invariant degrees-of-freedom as $\delta_{ij} = \delta = \exp(\delta_0)$, when $\boldsymbol{\beta} = \mathbf{0}$.

3.5 | Priors

We only give details about the prior distributions that are assigned to the parameters based on Approach 4, since Approaches 1–3 are just special cases. We set weakly informative prior distributions for the parameters. Elements of α are assigned zero-mean Cauchy prior with scale of 5, $C(0, 5)$, only α_0 was given $C(0, 20)$. $\boldsymbol{\Sigma}$ is re-written as $\mathbf{R\Omega R}$, with $\mathbf{R}$ being a diagonal matrix of scale parameters of B_h $(h = 1, \ldots, q)$, and $\boldsymbol{\Omega}$ a correlation matrix (Stan Development Team, 2019). Elements of $\mathbf{R}$ are given half-Cauchy, $C_+(0, 5)$, whereas elements of $\boldsymbol{\Omega}$ are given Lewandowski–Kurowicka–Joe (LKJ) distribution with the parameter of 2, $\mathcal{LKJ}(2)$. The degrees-of-freedom parameter for $\mathbf{B}_i$, ϕ, and δ_0 of Z_{ij} are given uniform priors, between 2 and 100. Elements of $\boldsymbol{\beta}$ are given $C(0, 5)$. σ is given $C_+(0, 5)$. Log-transformed elements of $h_0(t)$ and elements of $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$ are given $C(0, 5)$.

3.6 | Inference

For inference, we use Markov Chain Monte Carlo (MCMC) methods to sample from the joint posterior densities of the parameters and latent variables given data. We specifically consider the Hamiltonian Monte Carlo (HMC) (Gelman et al., 2013; Neal, 2011) as implemented in the No-U-Turn Sampler (NUTS) (Hoffman & Gelman, 2014) for each of the Approaches 1–4 as well as the standard joint model. For further details of HMC and NUTS, we refer the reader to these references. We do not present the details of either the likelihood function, nor HMC and NUTS, since, whereas the former is quite straightforward (e.g. see Asar et al., 2021), the second can be followed from the cited references.

Bespoke `R` (R Core Team, 2020) codes to fit the joint model under Approaches 1–4 are available from the `robjm` package (https://github.com/ozgurasarstat/robjm) that internally uses the HMC sampling engine `Stan` (Carpenter et al., 2017) through the `R` package `Rstan` (Stan Development Team, 2018).

4 | SIMULATION STUDY

A simulation study was conducted to investigate the effects of time-varying degrees-of-freedom in the estimation of parameters for the joint modelling approaches discussed in Section 3. A sample size of $n = 250$ individuals was considered with 200 datasets being simulated under the assumption of time-varying degrees-of-freedom for the longitudinal residuals. Note that we consider 200 as the number of replications and 250 as the number of subjects mainly because of the computational cost.

Data were generated under Approach 4. The longitudinal sub-model was given by

$$\begin{aligned} Y_{ij} &= Y_i^*(t_{ij}) + Z_{ij}, \\ &= \alpha_1 + t_{ij}\alpha_2 + B_{1i} + t_{ij}B_{2i} + Z_{ij}, \end{aligned} \tag{3}$$

where each individual has an average of 20 observations between time points 0 and 5. A random-intercept $(\alpha_1 + B_{1i})$ and random-slope $(\alpha_2 + B_{2i})$ model was assumed to replicate what is most commonly used in the joint modelling literature. The assumed underlying structure for the time-varying degrees-of-freedom is illustrated in Figure 4. Survival data were generated from the following model:

$$h_i(t) = h_0(t) \exp\left(X_i\omega + Y_i^*(t)\eta\right), \tag{4}$$

where $X \sim$ Bernoulli(0.5), $h_0(t)$ specified by a Weibull baseline hazard, that is, $h_0(t) = \lambda\nu t^{\nu-1}$. There is a final truncation time of 5 after which non-informative right-censoring occurs. The true values for the unknown parameters are given in Table 1.

Each robust model and the model with Gaussian assumptions for $\mathbf{B}_i$ and Z_{ij}, called the standard joint model, were estimated utilising the `robjm` package with four chains each of length 2000 with the first 1000 iterations considered as

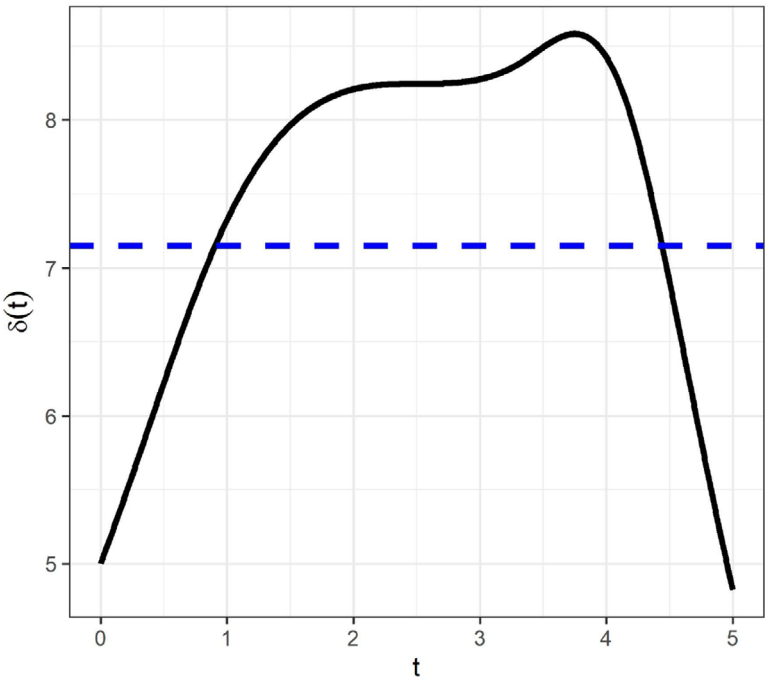


FIGURE 4 Black line gives the underlying degrees-of-freedom assumption for the simulation study. The dashed blue line gives the estimated degrees-of-freedom based on Approach 3 from the simulation study

TABLE 1 Table showing the simulation results based on data assuming time-varying degrees-of-freedom

		Standard joint			Approach 1			Approach 2			Approach 3			Approach 4		
	True	Mean	CI	Cov.	Mean	CI	Cov.	Mean	CI	Cov.	Mean	CI	Cov.	Mean	CI	Cov.
α_1	1.00	0.992	0.329	0.930	0.986	0.291	0.935	0.994	0.247	0.920	0.995	0.247	0.910	0.995	0.246	0.920
α_2	0.40	0.391	0.233	0.940	0.389	0.209	0.950	0.396	0.175	0.935	0.396	0.176	0.930	0.396	0.175	0.930
$\boldsymbol{\Sigma}(1,1)^*$	0.60	1.703	0.638	0.495	1.234	0.523	0.005	0.622	0.377	0.955	0.619	0.373	0.970	0.616	0.372	0.955
$\boldsymbol{\Sigma}(2,1)^*$	0.25	0.673	0.363	0.590	0.497	0.287	0.070	0.251	0.184	0.960	0.251	0.183	0.955	0.252	0.183	0.955
$\boldsymbol{\Sigma}(2,2)^*$	0.30	0.823	0.328	0.490	0.608	0.265	0.000	0.311	0.188	0.960	0.311	0.188	0.965	0.310	0.187	0.965
γ^{**}	—	—	—	—	20.719	—	—	—	—	—	—	—	—	—	—	—
ϕ	3.00	—	—	—	—	—	—	3.169	2.285	0.930	3.163	2.286	0.940	3.159	2.270	0.940
σ^{2***}	0.25	0.351	—	—	0.322	0.044	0.000	0.325	0.043	0.000	0.250	0.042	0.965	0.259	0.044	0.825
δ^{***}	5.00	—	—	—	—	—	—	30.350	—	—	7.149	—	—	—	—	—
δ_0	5.00	—	—	—	—	—	—	—	—	—	—	—	—	5.801	6.160	0.950
β_1	0.50	—	—	—	—	—	—	—	—	—	—	—	—	0.536	2.285	0.960
β_2	0.50	—	—	—	—	—	—	—	—	—	—	—	—	0.668	2.648	0.945
β_3	0.50	—	—	—	—	—	—	—	—	—	—	—	—	0.612	2.508	0.935
β_4	0.50	—	—	—	—	—	—	—	—	—	—	—	—	0.614	2.41	0.935
β_5	0.50	—	—	—	—	—	—	—	—	—	—	—	—	0.693	2.921	0.940
β_6	-0.25	—	—	—	—	—	—	—	—	—	—	—	—	0.152	2.400	0.895
λ	0.04	0.040	0.036	0.950	0.040	0.036	0.955	0.040	0.036	0.950	0.040	0.036	0.945	0.040	0.036	0.950
ν	1.20	1.207	0.433	0.950	1.206	0.434	0.955	1.210	0.430	0.945	1.210	0.431	0.950	1.211	0.432	0.950
ω	0.50	0.494	0.709	0.945	0.494	0.709	0.945	0.494	0.708	0.940	0.494	0.709	0.935	0.494	0.713	0.950
η	0.30	0.308	0.139	0.955	0.309	0.140	0.945	0.306	0.137	0.960	0.307	0.138	0.950	0.307	0.138	0.950

Abbreviations: CI, credibility intervals; Cov., coverage.

*For standard joint model, true values for these parameters were considered as 1.8 ($0.6 * \frac{3}{3-2}$), 0.75 ($0.3 * \frac{3}{3-2}$) and 0.9 ($0.3 * \frac{3}{3-2}$).

**For this parameter, CI and Cov. were not calculated, since the truth is unknown.

***For these parameters, CI and Cov. were not calculated, since the truths are time varying.

warm-up. The averages of posterior means, width of the corresponding 95% credibility intervals (CI) and coverage (Cov.) of 200 replications for each model are collated in Table 1.

To make $\boldsymbol{\Sigma}$, the variance of the random effects, comparable between the standard and robust joint models (Approaches 1–4), the relationship $\boldsymbol{\Sigma} \times \frac{\phi}{\phi-2}$ should be utilised for the true parameters of the standard joint model, since whereas $\boldsymbol{\Sigma}$ is the variance–covariance matrix for the Normal model, it is $\boldsymbol{\Sigma} \times \frac{\phi}{\phi-2}$ for the time-varying model that was used to simulate data. For ϕ of Model 1, CI and Cov. were not calculated, since the truth is unknown, as we simulate data from Approach 4. Similarly, for σ^2 of the Normal model and δ of Approaches 2 and 3, CI and Cov. were not calculated, since the truths are time varying.

For all the models, including the standard joint model, fixed effects for both the longitudinal and survival models and baseline hazard parameters demonstrate similar biases and coverages. In terms of averages of the credibility intervals for α_1 and α_2, the standard joint model produces larger values compared to the other models. Coverages for the elements of $\boldsymbol{\Sigma}$ are lower than expected for the standard joint model, and almost 0 for Approach 1, whereas they are at the expected level for Approaches 2–4. Approaches 1 and 2 produce biased and 0 coverage results for σ^2. This highlights the need to remove the dependence between Z_{ij} and $Z_{ij'}$. Approach 3 provides similar estimates and levels of coverage compared to Approach 4. Therefore, whilst Approach 3 has the inability to fully capture the time-varying nature of the degrees-of-freedom for the residuals, this appears to have a limited impact on the estimation of the other parameters.

5 | ILLUSTRATIVE DATA EXAMPLE: PBC DATA ANALYSIS

5.1 | Models fit

The work undertaken in this paper is motivated by the analysis of data from PBC patients. Due to the association between patients' alkaline levels and their survival, a model of the following form was fitted to the data:

$$\begin{aligned} Y_{ij} &= Y_i^*(t_{ij}) + Z_{ij} \\ &= \alpha_1 + \text{age}_i\alpha_2 + t_{ij}\alpha_3 + \text{I}(\text{drug}_i = \text{D-penicillamine})\alpha_4 + B_{i1} + t_{ij}B_{i2} + Z_{ij}, \end{aligned} \tag{5}$$

$$h_i(t) = h_0(t)\exp\left(\text{age}_i\omega_1 + \text{I}(\text{drug}_i = \text{D-penicillamine})\omega_2 + Y_i^*(t)\eta\right), \tag{6}$$

where $Y = \log(\text{alkaline})$, age is baseline age, t follow-up time and $h_0(t)$ Weibull baseline hazard function.

Based on the initial fitting of the models, the results for ϕ, the degrees-of-freedom parameter for the random effects distribution for Approaches 2–4, indicated that the Normal assumption might be reasonable for $\mathbf{B}_i$. As heavy tails were indicated for the residuals, only the Z term was therefore assumed to be t-distributed. In the context of the PBC data, this would suggest that, whilst individuals seem to demonstrate outlying observations from their own trends over the period of the study (e-outliers) and, thus do appear to need a period of adjustment to the new treatment, the individual average trends over time appear to be relatively consistent in their behaviour compared to that of the population and thus does not suggest the presence of b-outliers. Uncoupling the relationship between e- and b-outliers, as is achieved in Approaches 3 and 4, would therefore likely be required to fully capture the aspects of the PBC data being analysed.

To investigate the presence of time-varying degrees-of-freedom, for Approach 4 we considered the number of knots to range from 1 to 5, with the knots put into the empirical quantiles of the time variable. We build nine models in total:

- Normally distributed $\boldsymbol{B}_i$, Normally distributed Z_{ij} (nor-nor),
- t-distributed $\boldsymbol{B}_i$ and t-distributed Z_{ij} based on Approach 1 (t-t-mod1),
- Normally distributed $\boldsymbol{B}_i$ and t-distributed Z_{ij} based on Approach 2 (nor-t-mod2),
- Normally distributed $\boldsymbol{B}_i$ and t-distributed Z_{ij} based on Approach 3 (nor-t-mod3),
- Normally distributed $\boldsymbol{B}_i$ and t-distributed Z_{ij} based on Approach 4, with number of knots to range from 1 to 5 (nor-t-tv-1knot, nor-t-tv-2knots, nor-t-tv-3knots, nor-t-tv-4knots, nor-t-tv-5knots).

For each approach, four chains, each length of 6000, were started from random initials with half of each chain considered as warm-up. Convergence was checked through traceplots and R-hat statistics (Brooks & Gelman, 1997). Posterior summaries, specifically the 2.5%, 50% and 97.5% percentiles, are displayed in Table 2. Traceplots for the Approach 4 with 5 knots are presented in Figure 5.

TABLE 2 Posterior summaries of the joint model parameters fitted to the PBC dataset. Values in the upper rows for each parameter are the 50% percentiles, whereas the brackets in the lower rows are the 2.5% and 97.5% percentiles, respectively

	nor-nor	mod1	nor-t-mod2	nor-t-mod3	nor-t-tv-1knot	nor-t-tv-2knots	nor-t-tv-3knots	nor-t-tv-4knots	nor-t-tv-5knots
α_1	7.851	7.468	7.479	7.515	7.521	7.517	7.521	7.526	7.529
	(7.259, 7.907)	(7.159, 7.779)	(7.161, 7.792)	(7.205, 7.840)	(7.191, 7.848)	(7.194, 7.832)	(7.206, 7.856)	(7.210, 7.863)	(7.197, 7.842)
α_2	−0.008	−0.007	−0.006	−0.007	−0.007	−0.007	−0.007	−0.007	−0.007
	(−0.014, −0.001)	(−0.013, −0.001)	(−0.012, 0.001)	(−0.013, −0.001)	(−0.014, −0.0003)	(−0.013, −0.001)	(−0.014, −0.001)	(−0.014, −0.001)	(−0.013, −0.001)
α_3	−0.046	−0.041	−0.040	−0.035	−0.035	−0.035	−0.035	−0.035	−0.035
	(−0.058, −0.034)	(−0.051, −0.030)	(−0.052, −0.028)	(−0.046, −0.024)	(−0.046, −0.024)	(−0.046, −0.023)	(−0.046, −0.024)	(−0.046, −0.023)	(−0.046, −0.023)
α_4	−0.066	−0.020	−0.059	−0.072	−0.073	−0.071	−0.072	−0.070	−0.073
	(−0.185, 0.054)	(−0.133, 0.095)	(−0.184, 0.062)	(−0.199, 0.051)	(−0.200, 0.049)	(−0.193, 0.054)	(−0.198, 0.050)	(−0.192, 0.051)	(−0.192, 0.046)
Σ_{11}	0.339	0.230	0.320	0.349	0.357	0.359	0.360	0.359	0.358
	(0.286, 0.406)	(0.186, 0.284)	(0.268, 0.384)	(0.295, 0.414)	(0.301, 0.425)	(0.304, 0.426)	(0.303, 0.428)	(0.303, 0.428)	(0.302, 0.427)
Σ_{12}	−0.018	−0.009	−0.012	−0.015	−0.016	−0.016	−0.016	−0.016	−0.016
	(−0.026, −0.011)	(−0.015, −0.004)	(−0.019, −0.005)	(−0.022, −0.009)	(−0.023, −0.009)	(−0.023, −0.009)	(−0.024, −0.010)	(−0.023, −0.010)	(−0.024, −0.010)
Σ_{22}	0.004	0.003	0.004	0.005	0.005	0.005	0.005	0.005	0.005
	(0.003, 0.005)	(0.002, 0.004)	(0.003, 0.006)	(0.004, 0.006)	(0.004, 0.006)	(0.004, 0.006)	(0.004, 0.006)	(0.004, 0.006)	(0.004, 0.006)
σ^2	0.107	0.047	0.040	0.021	0.023	0.024	0.024	0.024	0.024
	(0.099, 0.115)	(0.040, 0.054)	(0.034, 0.047)	(0.018, 0.025)	(0.020, 0.027)	(0.021, 0.028)	(0.021, 0.028)	(0.021, 0.029)	(0.021, 0.029)
γ	—	3.516	—	—	—	—	—	—	—
		(2.810, 4.394)							
δ	—	—	2.759	2.091	—	—	—	—	—
			(2.207, 3.451)	(1.824, 2.432)					
δ_0	—	—	—	—	1.449	1.184	1.130	1.122	1.130
					(1.235, 1.704)	(0.994, 1.410)	(0.940, 1.361)	(0.931, 1.360)	(0.935, 1.367)
β_1	—	—	—	—	2.001	−0.527	1.358	1.308	1.432
					(1.432, 2.625)	(−1.408, 0.339)	(0.782, 2.009)	(0.591, 2.118)	(0.664, 2.035)
β_2	—	—	—	—	−0.543	2.946	0.202	1.376	1.175
					(−1.459, 0.459)	(2.076, 3.969)	(−0.673, 1.031)	(0.699, 2.128)	(0.391, 2.032)
β_3	—	—	—	—	—	1.380	2.566	0.242	1.329
						(0.079, 3.056)	(1.720, 3.549)	(−0.816, 1.241)	(0.694, 2.024)
β_4	—	—	—	—	—	—	1.210	2.414	0.314
							(−0.192, 3.202)	(1.516, 3.694)	(−0.703, 1.352)
β_5	—	—	—	—	—	—	—	1.763	2.263

(Continues)

TABLE 2 (Continued)

	nor-nor	mod1	nor-t-mod2	nor-t-mod3	nor-t-tv-1knot	nor-t-tv-2knots	nor-t-tv-3knots	nor-t-tv-4knots	nor-t-tv-5knots
								(0.107, 4.256)	(1.321, 3.556)
β_6	—	—	—	—	—	—	—	—	1.909
									(0.161, 4.574)
$\log(\lambda)$	−11.276	−11.427	−12.129	−11.952	−11.901	−11.911	−11.861	−11.854	−11.872
	(−14.850, −7.844)	(−14.814, −8.168)	(−15.584, −8.869)	(−15.215, −8.761)	(−15.300, −8.641)	(−15.256, −8.773)	(−15.178, −8.694)	(−15.200, −8.618)	(−15.242, −8.679)
$\log(\nu)$	0.223	0.226	0.234	0.222	0.221	0.220	0.219	0.220	0.221
	(0.064, 0.373)	(0.068, 0.372)	(0.076, 0.380)	(0.069, 0.366)	(0.065, 0.367)	(0.062, 0.367)	(0.065, 0.364)	(0.062, 0.362)	(0.066, 0.363)
ω_1	0.051	0.052	0.053	0.053	0.053	0.053	0.053	0.053	0.053
	(0.033, 0.070)	(0.035, 0.070)	(0.036, 0.072)	(0.035, 0.071)	(0.035, 0.072)	(0.036, 0.072)	(0.035, 0.071)	(0.036, 0.071)	(0.035, 0.072)
ω_2	−0.157	−0.121	−0.144	−0.137	−0.137	−0.136	−0.136	−0.133	−0.139
	(−0.519, 0.206)	(−0.460, 0.229)	(−0.508, 0.212)	(−0.488, 0.217)	(−0.492, 0.220)	(−0.498, 0.228)	(−0.494, 0.215)	(−0.483, 0.217)	(−0.490, 0.215)
η	0.799	0.805	0.895	0.873	0.872	0.869	0.865	0.862	0.865
	(0.391, 1.216)	(0.423, 1.193)	(0.511, 1.293)	(0.502, 1.244)	(0.491, 1.253)	(0.497, 1.255)	(0.487, 1.245)	(0.482, 1.250)	(0.488, 1.248)

TABLE 3 WAIC, LPML and computational times

Model	−2* LPPD	2 * ρ_{WAIC}	WAIC	LPML	Time (h)
nor-nor	1949.325	489.323	2438.648	−1374.996	0.8
mod1	886.645	812.729	1699.374	−1045.701	1.8
nor-t-mod2	769.040	917.524	1686.564	−1038.587	1.5
nor-t-mod3	−394.382	1261.797	867.415	−649.473	1.8
nor-t-tv-1knot	−279.065	1185.973	906.908	−678.852	4.3
nor-t-tv-2knots	−241.282	1143.800	902.518	−662.119	6.2
nor-t-tv-3knots	−251.507	1148.744	897.237	−657.711	5.6
nor-t-tv-4knots	−246.941	1145.244	898.303	−664.230	3.1
nor-t-tv-5knots	−245.131	1150.149	905.018	−664.834	2.9

Abbreviations: LPML, log psuedo-marginal likelihood; LPPD, log posterior predictive density; WAIC, Watanebe Information Criterion.

Computational times are presented in Table 3. These times are based on a 64-bit desktop computer with 16 GB RAM and AMD Ryzen 7 1800X eight core processor 3.60 GHz running Windows 10.

5.2 | Model comparison

For model comparison, we consider the Watanebe Information Criterion (WAIC; Gelman et al., 2014; Watanebe, 2010) and the log psuedo-marginal likelihood (LPML; Dey et al., 1997; Gelman et al., 2014). The formulae that we used to calculate these measures are given in the Appendix. Lower values of WAIC and higher values of LPML indicate better model fit. WAIC and LPML values are presented in Table 3. In terms of both WAIC and LPML, Approach 3 is the best fitting model. Amongst the models based on Approach 4, the one with three knots seems to be the best fitting model. Both Approaches 3 and 4 seem to outperform the other models.

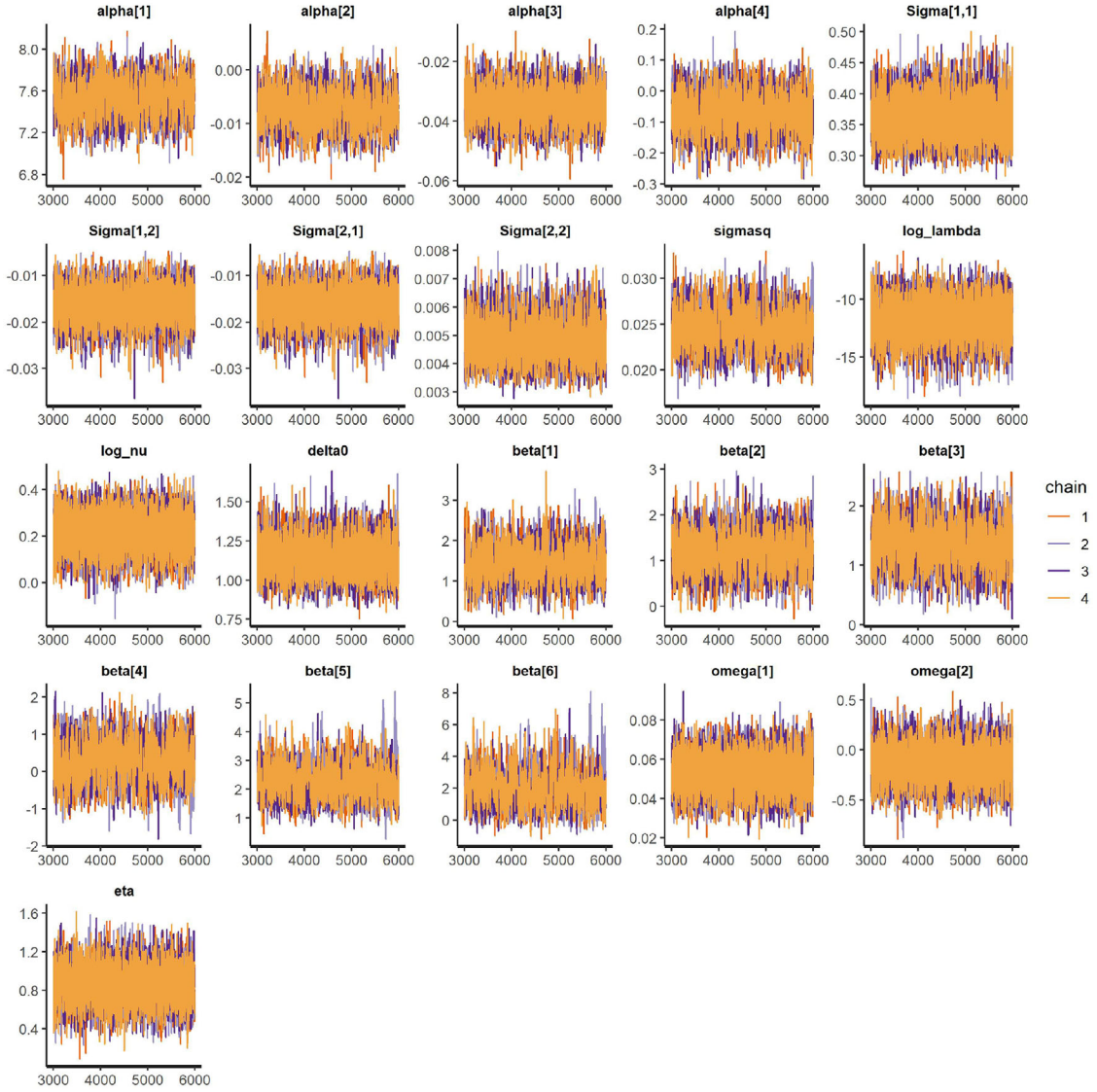


FIGURE 5 Traceplots of the estimated parameters from Approach 4 with five knots

5.3 | Interpretation

Similar to the findings of the simulation study in Section 4, in general, results based on Approaches 3 and 4 are quite similar. This is particularly evident in the regression coefficients from the longitudinal and survival sub-models, $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$, respectively. Given the limited change in estimated parameters between the two approaches, it is therefore not unexpected to find high levels of agreement when comparing the posterior summaries of the predictions for the longitudinal random effects, as illustrated in Figure 6. This consistency between Approaches 3 and 4 is in spite of the difference in assumption regarding the degrees-of-freedom parameter, as illustrated in the right-hand panel of Figure 7, which gives the degrees-of-freedom estimates for Approaches 3 and 4 with three knots based on the PBC data. The PBC analysis further highlights the robustness of Approach 3 to time-varying degrees-of-freedom structure in the measurement error term, Z, that is demonstrated in Section 4.

There are differences, however, in the results of Approaches 3 and 4 and the other models, with some parameters being overestimated (such as σ^2) whereas some were underestimated. For example, the association parameter, η, appears to be affected by the choice of model. It is underestimated by the standard joint model, which was overcome by Approach 1 up to some extent. In particular, contrasting Approaches 1 and 2 show the effect of alleviating the restrictive assumption

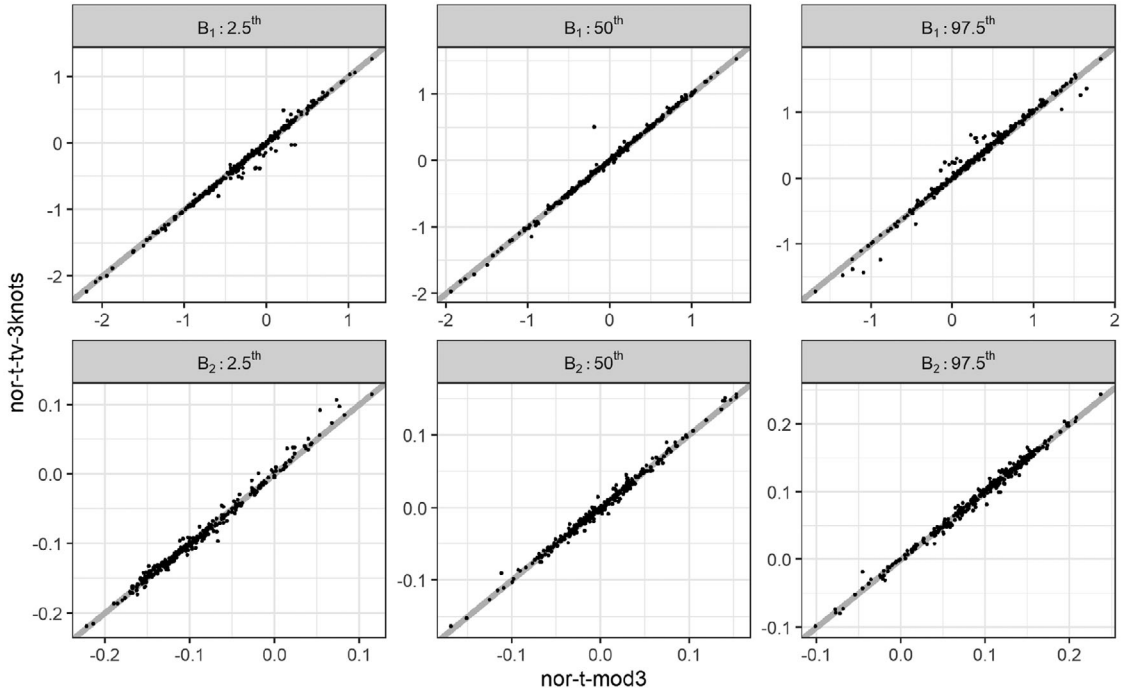


FIGURE 6 Posterior summaries (2.5%, 50% and 97.5% percentiles of the MCMC samples) for Approach 3 against the time-varying model with three knots from the PBC analysis for the predictions of the longitudinal random intercept (top row) and slope (bottom row)

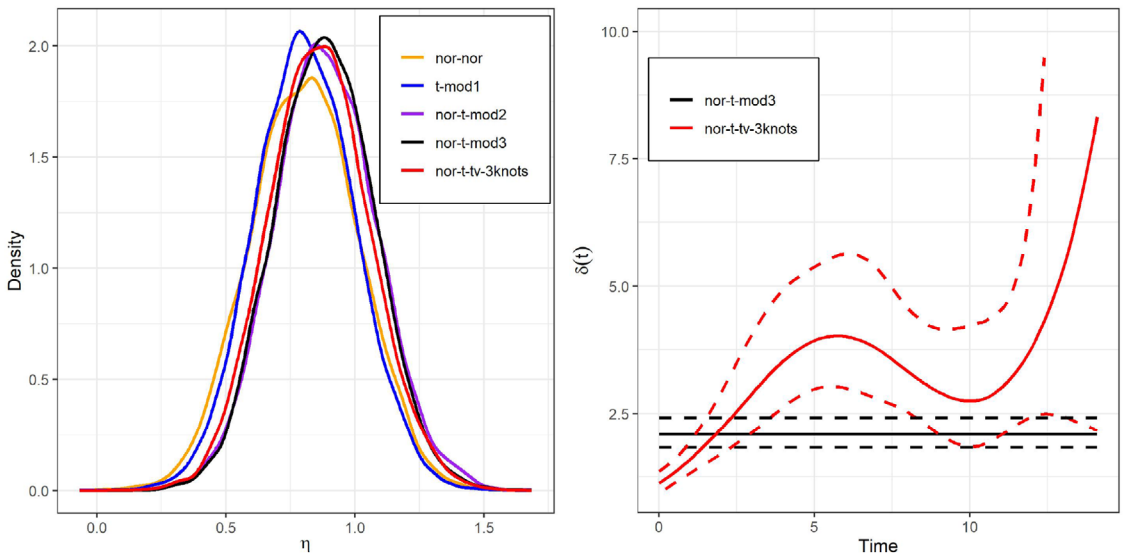


FIGURE 7 Left panel: smoothed MCMC samples of η under different models. Right panel: Medians of degrees-of-freedom parameters (mid-lines) and 2.5% and 97.5% percentiles of the MCMC samples under Model 3 and tv model with three knots. Note that the upper limit for the y-axis of the right panel was cut at 10 for better visualisation as towards the end the upper limit for the time-varying model is quite large

of a common degrees-of-freedom parameter between the random effects and residuals, an assumption still commonly used within the literature. In the case of the PBC analysis, removing the coupling of b- and e-outliers, such that $\mathbf{B}_i \perp Z_{ij}$ is achieved, is necessary as heavy tails were only indicated for the residuals. Forcing both $\mathbf{B}_i$ and Z to be heavy tailed (Approach 1) when this is not the case, changes the interpretation of the relationship and impact of the time-varying biomarker on the survival of PBC patients. The tighter credible intervals that are achieved through Approaches 3 and 4 have the potential to reduce the chance of a Type II error, incorrectly concluding that the association between the

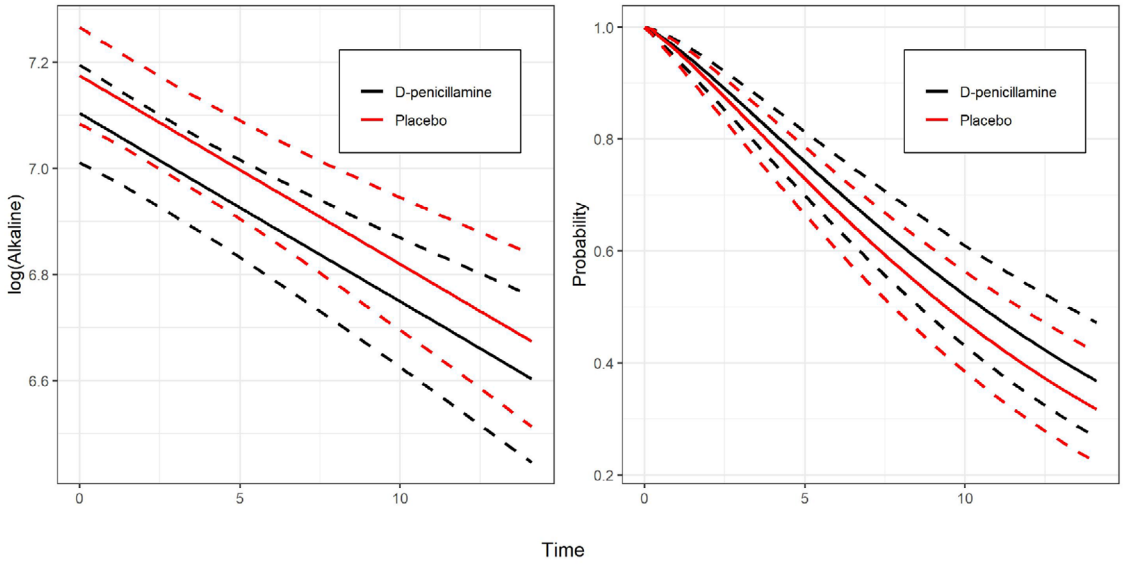


FIGURE 8 Average longitudinal profiles (left panel) and survival curves (right panel) based one Approach 3. Solid lines are the medians, whereas the accompanying dashed lines are the 2.5% and 97.5% quantiles. The plots are for a subject who was 50 year old and had $Y_i^*(t)$ of 7

longitudinal and survival processes is not of significance and should therefore be ignored. This may have the potential in certain circumstances of failing to detect the influence, and the true extent of the impact that a longitudinal biomarker can have on patients' survival. Such differences in the estimation of η between the approaches can be seen in the smoothed densities of the MCMC samples of η that are displayed on the left panel of Figure 7.

The impact of the choice of robust approaches is evident in the estimation of the residual variance, σ^2. Assuming normality for the residuals in the presence of *e*-outliers produces larger estimates of the residual variance compared to using the more robust approaches, with Approach 3, for example, providing an 80% decrease in the residual variance compared to the standard joint model in the analysis of the PBC data. Thus, not properly handling the *e*-outliers within the PBC data, possibly caused by the start of the new treatment, results in larger differences between the expected and true alkaline phosphatase responses. Only Approaches 3 and 4 that provide the most natural approaches to handling *e*-outliers seem to capture the true nature of the longitudinal biomarker, evidenced by the reduction in residual variance.

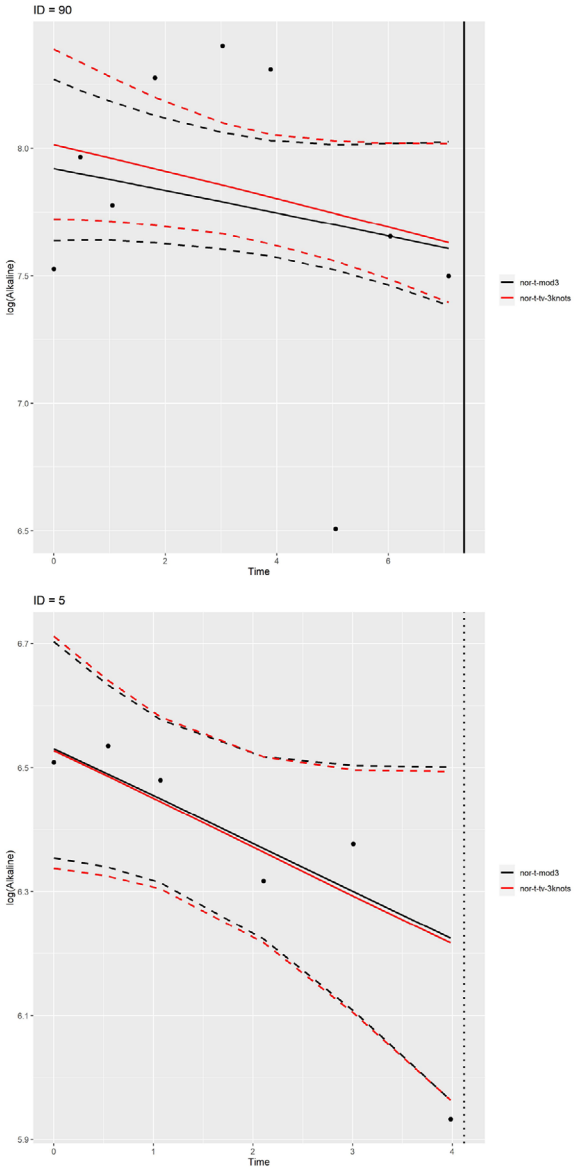
2.5%, 50% and 97.5% percentiles of the MCMC samples for a 50-year old subject were plotted with respect to the two treatment arms in Figure 8. The left panel presents the results for the average longitudinal evolution through time assuming that the random-effects terms being 0, whereas the right panel presents the survival curves based on the assumption that $Y_i^*(t) = 7$. Note that the average age was 50 and average log(Alkaline) was 7. Both panels indicate that there is no difference between the two treatment arms.

5.4 | Outlier analysis

As mentioned in Section 3.3, we could use posterior distributions of V_i^Z for outlier analysis. In Figure 9, we present posterior summaries (2.5%, 50% and 97.5% percentiles) of V_i^Z for two subjects, based on Approaches 3 and 4 with three knots. For the first subject (ID = 90), fifth, sixth and seventh repeats were identified as outliers, whereas none of the repeats for the second (ID = 5) was identified as an outlier. This observation is based on the credibility intervals excluding 1.

5.5 | Posterior predictive checks

We conducted posterior predictive checks to see how the replicated data under the fitted models replicate the observed data. For this, we simulated data from the $f(Y_i, T_i, E_i|\boldsymbol{B}_i, V_i, W_i, \theta)$, where θ indicates the collection of all the model



Model	Quantile	t = 0	t = 0.5	t = 1	t = 1.8	t = 3	t = 3.9	t = 5.1	t = 6	t = 7.1
	2.5%	0.61	0.27	0.27	0.77	1.6	1.25	6.99	0.25	0.27
nor-t-mod3	50%	3.86	1.23	1.26	4.34	7.82	6.58	29.06	1.08	1.3
	97.5%	50.03	14.63	15.42	47.57	87.58	76.95	295.18	13.02	19.81
	2.5%	0.9	0.27	0.39	0.55	0.99	0.9	5.35	0.31	0.34
nor-t-tv-3knots	50%	7.27	1.16	1.35	2.02	3.62	3.42	18.32	1.07	1.26
	97.5%	186.75	12.45	9.46	11.88	22.99	22.19	122.63	8.23	10.76

Model	Quantile	t = 0	t = 0.5	t = 1.1	t = 2.1	t = 3	t = 4
	2.5%	0.24	0.24	0.24	0.24	0.27	0.42
nor-t-mod3	50%	0.97	0.97	0.96	0.99	1.11	2.59
	97.5%	10.64	10.54	9.29	11.29	10.39	34.67
	2.5%	0.18	0.26	0.32	0.35	0.34	0.47
nor-t-tv-3knots	50%	0.96	0.99	0.97	0.99	1.08	1.82
	97.5%	19.8	8.39	5.14	4.79	5.86	12.44

FIGURE 9 Individual fits (left panel) and posterior summaries of V_i^Z (right panel) for two subjects, based on Approaches 3 and 4 with three knots

parameters, for each element of the MCMC samples of size 12,000 under each model. For each simulated repeated measures data, we calculated mean absolute error (MAE) and root mean squared error (RMSE) by comparing the simulated repeated measures datasets with the repeated measures from the PBC. For each simulated survival data, we computed Kaplan–Meier survival probability estimates and plotted them together with the Kaplan–Meier curve of the observed survival data from the PBC.

RMSE and MAE values are displayed in Figure 10, whereas the Kaplan–Meier plots are displayed in Figure 11. Note that we used only the first 1000 of the Kaplan–Meier estimates in the plots as these were able to cover the survival curve of the PBC dataset. In terms of MAE, Approaches 3 and 4 seem to outperform the other models, and Approaches 1 and 2 outperform the standard joint model. In terms of RMSE, Approach 3 and especially Approach 4 seem to produce RMSE in wider intervals, yet overall performances of all the models are similar, for example, in terms of the medians. Note that the more robust a model is, the less weight is put on minimising the RMSE. This explains why the Normal model has the best performance in terms of RMSE.

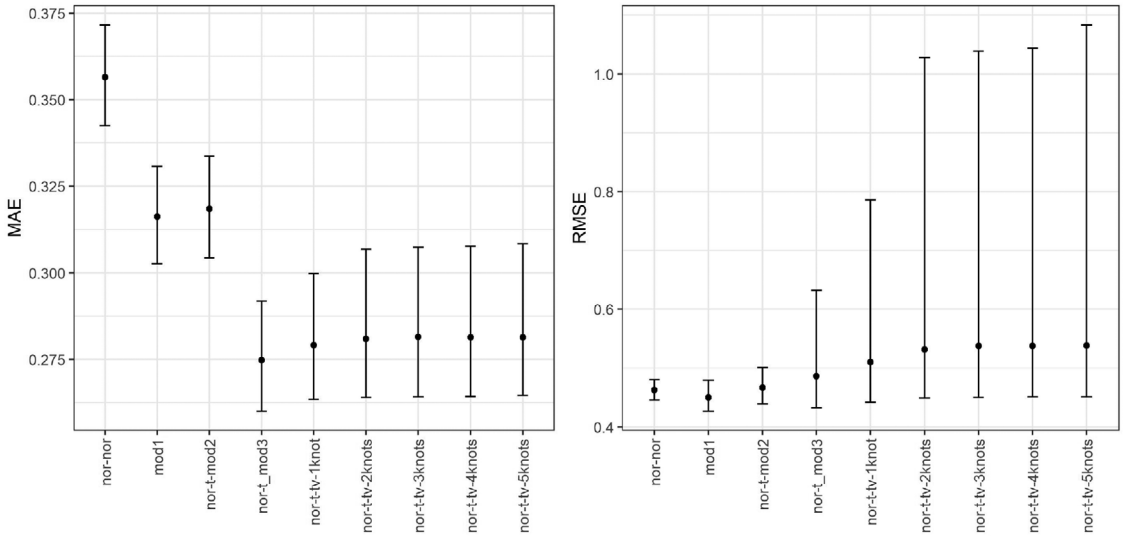


FIGURE 10 Posterior predictive checks: MAE (left panel) and RMSE (right panel)

6 | DISCUSSION

This work introduces a new flexible approach to model longitudinal outliers in a joint modelling framework, where the degrees-of-freedom parameter for the residuals is assumed to vary over time. This scenario replicates the common situation when patients take time to adjust to new treatments, resulting in more outlying measurements (*e*-outliers) at different times across the period of observation, as illustrated in the motivating example that analysed data from PBC patients. In addition, this paper contrasts the new time-varying approach with three alternative time-invariant formulations of robust joint models currently found in the literature, but which have never previously been compared via a simulation study, and the standard joint model with Gaussian assumptions for the random terms. Each of the approaches presented can be fitted in the accompanying `robjm` software package.

Throughout the results presented within the simulation study and the real-life data analysis, the need to properly account for longitudinal outliers is evident, a practice not widely adopted in the current joint modelling literature. This paper highlights the need to investigate the presence of outliers and provides a flexible approach to downweighing their negative impact, contrasting the effect of various assumptions made by the different approaches to robust joint modelling. When comparing the standard and robust joint models, it is evident that bias and inefficiency result from ignorance of longitudinal outliers and the assumption of Normality for the random terms. Previous research on time-invariant robust joint models noted that failing to account for outliers in a joint modelling context can affect both the calibration and discrimination of dynamic predictions and thus may have a major impact on the treatment plans and prognosis of patients (Asar et al., 2021). The extent of the negative impact from Normality assumptions in the presence of outliers is highlighted through the simulation study we have undertaken in this paper.

This research not only stresses the need to account for longitudinal outliers but contrasts various robust approaches, using both time-invariant and time-varying methods, to properly estimate the degrees-of-freedom parameter. To the knowledge of the authors, there is no work in the literature that has contrasted Approaches 1–3 under any scenario in a simulation study. Although only a limited number of papers so far utilise robust joint modelling approaches, the majority of these focus on Approaches 1 and 2, which impose restrictive assumptions on the relationship between parameters. Approach 1 makes the strong assumptions that *e*- and *b*-outliers are coupled together whilst also coupling the *e*-outliers within the individual ($\mathbf{B}_i \not\perp Z_{ij}$ and $Z_{ij} \not\perp Z_{ij'}$). Approach 2 removes the coupling of *e*- and *b*-outliers; however, *e*-outliers within the individual are still coupled ($\mathbf{B}_i \perp Z_{ij}$ and $Z_{ij} \not\perp Z_{ij'}$). In Approaches 3 and 4, all of the outliers are decoupled ($\mathbf{B}_i \perp Z_{ij}$ and $Z_{ij} \perp Z_{ij'}$). Note that in certain applications or under certain conditions different models may be more suitable, for instance, there could be a situation where a subject having the characteristic of *b*-outlier which creates coupled outliers. However, it is of great importance that practitioners understand the assumptions of the different approaches and the corresponding implications of these assumptions and thus chooses the most appropriate model for the situation under analysis.

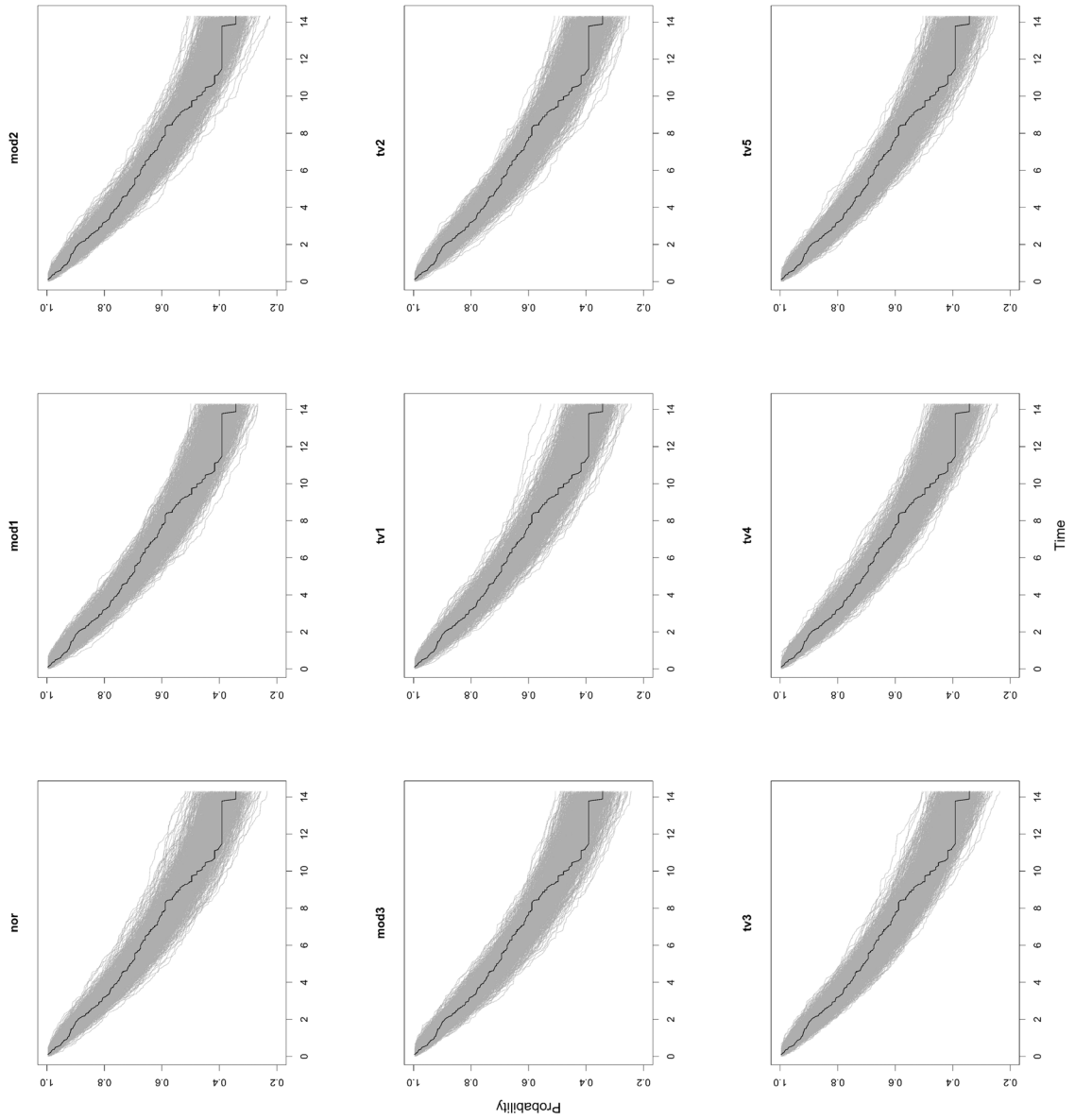


FIGURE 11 Posterior predictive checks: Kaplan–Meier survival probability estimates. Black lines are the survival curves for the PBC data, whereas grey lines are for the simulated datasets

In fact, unexpectedly, despite a true underlying assumption of time-varying degrees-of-freedom within the simulation study, Approach 3 was found to be in good agreement with the time-varying model (Approach 4). Within the simulation study and in the real-life example, the degrees of freedom were assumed or estimated to be consistently low and varying over a limited range throughout the period of observation. Under such scenarios, Approach 3 appears to provide a time-invariant average of the time-varying degrees of freedom for the residuals over the period of observation, as demonstrated within Figure 4 based on the simulation results. This work highlights the robustness provided by Approach 3 in the estimation of the degrees-of-freedom, providing similar estimates as Approach 4 in spite of the presence of time-varying degrees-of-freedom. It signifies that in cases where the degrees-of-freedom parameter does not vary widely, Approach 3 is recommended when longitudinal outliers are present.

We did not consider model selection methods to select the number of knots, mainly because the number of knots does not seem to change the results considerably. In the time-varying degrees-of-freedom model, we assumed the scale parameter, σ, being time constant. This parameter might also be assumed to be time varying. Another approach would also be to specify the random effects with time-varying degrees-of-freedom. For this, the random effects shall be time varying. Inclusion of a stochastic process as done in Asar et al. (2020) could be opted for this purpose.

ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council [Reference: EP/P026028/1].

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Özgür Asar https://orcid.org/0000-0003-0603-1409

REFERENCES

Asar, Ö, Bolin, D., Diggle, P. J., & Wallin, J. (2020). Linear mixed-effects models for non-Gaussian continuous repeated measurement data (with discussion). *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *69*(5), 1015–1065.

Asar, Ö., Fournier, M.-C., & Dantan, E. (2021). Dynamic predictions of kidney graft survival in the presence of longitudinal outliers. *Statistical Methods in Medical Research*, *30*(1), 185–203. https://doi.org/10.1177/0962280220945352.

Asar, Ö., Ritchie, J., Kalra, P. A., & Diggle, P. J. (2015). Joint modelling of repeated measurement and time-to-event data: An introductory tutorial. *International Journal of Epidemiology*, *44*(1), 334–344.

Baghfalaki, T., Ganjali, M., & Berridge, D. (2013). Robust joint modeling of longitudinal measurements and time to event data using normal/independent distributions: A Bayesian approach. *Biometrical Journal*, *34*(2), 187–220.

Baghfalaki, T., Ganjali, M., & Hashemi, R. (2014). Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. *Journal of Biopharmaceutical Statistics*, *24*(4), 834–855.

Brooks, S. P., & Gelman, A. (1997). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *1*, 1–32.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B (Methodological)*, *34*(2), 187–220.

Dey, D. K., Chen, M. H., & Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics*, *53*, 1239–1252.

Diggle, P. J., Sousa, I., & Chetwynd, A. G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: The fourth Armitage lecture. *Statistics in Medicine*, *27*(16), 2981–2998.

Elashoff, R. M. (2017). *Joint modeling of longitudinal and time-to-event data*. CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC Press.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 9971016.

Hickey, G. L., Philipson, P., Jorgensen, A., & Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, *16*, 117.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Huang, X., Li, G., & Elashoff, R. (2010). A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and Its Interface*, *3*(2), 185–195.

Kravitz, R. L., Duan, N., & Braslow, J. (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly*, *82*(4), 661–687.

Li, N., Elashoff, R. M., & Li, G. (2009). Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biometrical Journal*, *51*(1), 19–30.

McCrink, L. M. (2014). Outlier effects on robust joint modelling of longitudinal and survival data (Ph.D. Thesis). Queen's University Belfast.

McCrink, L. M., Marshall, A. H., & Cairns, K. J. (2013). Advances in joint modelling: A review of recent developments in software and techniques using the modelling of survival of end stage renal disease patients as a working example. *International Statistical Review*, *81*, 249–269.

Murtaugh, P., Dickson, E., Van Dam, G., Malincho, M., Grambsch, P., Langworthy, A., & Gips, C. (1994) Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. *Hepatology*, *20*, 126–134.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, & X. L Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 113–162). Chapman & Hall/CRC Press.

Pinheiro, J. C., Liu, C., & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate *t* distribution. *Journal of Computational and Graphical Statistics*, *10*, 249–276.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. Chapman & Hall/CRC Press.

Song, H., Peng, Y., & Tu, D. (2012). A new approach for joint modelling of longitudinal measurements and survival times with a cure fraction. *Canadian Journal of Statistics*, *40*(2), 207–224.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.17.3. http://mc-stan.org/.

Stan Development Team (2021). Stan user's guide, Version 2.27. https://mc-stan.org/docs/2_27/stan-users-guide-2_27.pdf

Taylor, J. M. G., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., & Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, *69*(1), 206–213.

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

Watanebe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: McFetridge, L. M., Asar, Ö., Wallin, J. Robust joint modelling of longitudinal and survival data: Incorporating a time-varying degrees-of-freedom parameter. *Biometrical Journal*. 2021;*63*:1587–1606. https://doi.org/10.1002/bimj.202000253

APPENDIX

To describe the WAIC, we will use the notation $E^{MC}_{\theta,\boldsymbol{B}_i}$ and $V^{MC}_{\theta,\boldsymbol{B}_i}$ to denote the mean and variance of the MCMC runs, that is the empirical mean and variance from the MCMC samples. WAIC is calculated as

$$\text{WAIC} = -2\left[\underbrace{\sum_{i=1}^{n}\log\left\{E^{MC}_{\theta,\boldsymbol{B}_i}[L_i(\boldsymbol{Y}_i,S_i,E_i\mid\theta,\boldsymbol{B}_i)]\right\}}_{lppd} - \underbrace{V^{MC}_{\theta,\boldsymbol{B}_i}[\log\{L_i(\boldsymbol{Y}_i,S_i,E_i\mid\theta,\boldsymbol{B}_i)\}]}_{p}\right],$$

where

$$L_i(\boldsymbol{Y}_i,S_i,E_i\mid\theta,\boldsymbol{B}_i) = \left[\prod_{j=1}^{m_i} f(\boldsymbol{Y}_{ij}\mid\theta,\boldsymbol{B}_i)f(S_i,E_i\mid\theta,\boldsymbol{B}_i)\right] f(S_i,E_i\mid\theta,\boldsymbol{B}_i),$$

with

$$f(\boldsymbol{Y}_{ij}\mid\theta,\boldsymbol{B}_i) = N(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\alpha}+\boldsymbol{d}_{ij}^{\top},\sigma^2W_{ij})$$

$$f(S_i,E_i\mid\theta,\boldsymbol{B}_i) = \left\{\lambda\nu S_i^{\nu-1}\exp(\boldsymbol{c}_i\boldsymbol{\omega}+\eta Y_i^*(S_i))\right\}^{E_i}\left\{\exp\left(-\int_0^{S_i}\lambda\nu u^{\nu-1}\exp(\boldsymbol{c}_i\boldsymbol{\omega}+\eta Y_i^*(u)du\right)\right\}.$$

The integral in the survival function is approximated by Gauss-Kronrod rule such that

$$\int_0^{S_i} h_i u du \approx 0.5S_i\sum_{q=1}^{Q} wt_q h_i(0.5S_i(1+pt_q)),$$

where wt_q are the weights, the pt_q are the abscissa and Q is the number of quadratures. lppd stands for log posterior predictive density, and ρ indicates the complexity term.

LPML is calculated as

$$\text{LPML} = \sum_{i=1}^{n} \log(\widehat{\text{CPO}_i}),$$

where

$$\widehat{\text{CPO}_i} = \left\{ \frac{1}{M} \sum_{k=1}^{M} \frac{1}{L_i\left(\boldsymbol{Y}_i, S_i, E_i \mid \theta^k, \boldsymbol{B}_i^k\right)} \right\}^{-1},$$

with θ^k being the kth element of the MCMC sample.