



# A two-stage approach for Bayesian joint models: reducing complexity while maintaining accuracy

Danilo Alvares<sup>1</sup> · Valeria Leiva-Yamaguchi<sup>1</sup>

Received: 6 October 2022 / Accepted: 16 July 2023 / Published online: 31 July 2023  
© The Author(s) 2023

## Abstract

Several joint models for longitudinal and survival data have been proposed in recent years. In particular, many authors have preferred to employ the Bayesian approach to model more complex structures, make dynamic predictions, or use model averaging. However, Markov chain Monte Carlo methods are computationally very demanding and may suffer convergence problems, especially for complex models with random effects, which is the case for most joint models. These issues can be overcome by estimating the parameters of each submodel separately, leading to a natural reduction in the complexity of the joint modelling, but often producing biased estimates. Hence, we propose a novel two-stage approach that uses the estimations from the longitudinal submodel to specify an informative prior distribution for the random effects when estimating them within the survival submodel. In addition, as a bias correction mechanism, we incorporate the longitudinal likelihood function in the second stage, where its fixed effects are set according to the estimation using only the longitudinal submodel. Based on simulation studies and real applications, we empirically compare our proposal with joint specification and standard two-stage approaches considering different types of longitudinal responses (continuous, count and binary) that share information with a Weibull proportional hazard model. The results show that our estimator is more accurate than its two-stage competitor and as good as jointly estimating all parameters. Moreover, the novel two-stage approach significantly reduces the computational time compared to the joint specification.

**Keywords** Bayesian inference · Bias reduction · Longitudinal data · Time-to-event

**Mathematics Subject Classification** 62F15 · 62N02

## 1 Introduction

Joint modelling for longitudinal and survival data has become very popular medical applications (Ibrahim and Molenberghs 2009; Neuhaus et al. 2009; Wu et al. 2011). This class of models is a useful tool when it is necessary to study the association between repeated measurements and time until an event of interest (Papageorgiou et al. 2019). When considering both information together (simultaneously) into a single model, the estimates are less biased and there is an increase in statistical efficiency, since clinical hypotheses consider in

advance that longitudinal and survival data are connected in some way (Muthén et al. 2009; Ibrahim et al. 2010).

The complexity of the joint modelling often entails two critical issues. The first one is related to identification problems due to the large number of parameters (Wu 2009; Gould et al. 2015; Papageorgiou et al. 2019). On the other hand, the calculation of marginal likelihood and survival functions include integrals without closed-form, so numerical integration methods are required, making the inferential process more time-consuming (Lesaffre and Spiessens 2001; Pinheiro and Chao 2006; Rizopoulos et al. 2009; Wu et al. 2010; Barrett et al. 2015).

Recently, Alvares and Rubio (Alvares and Rubio 2021) proposed a new specification for joint models of longitudinal and survival data that is flexible and does not require numerical integrations. The proposal divides the longitudinal linear predictor into variables that depend or not on time to share them with the survival submodel. This strategy leads to

---

✉ Danilo Alvares  
danilo.alvares@mrc-bsu.cam.ac.uk  
Valeria Leiva-Yamaguchi  
valeria.leiva@mrc-bsu.cam.ac.uk

<sup>1</sup> MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

some limitations presented by the authors and which will not be discussed here. Another incipient and very promising proposal is the use of integrated nested Laplace approximation (INLA) for joint models (van Niekerk et al. 2020). However, this approach is not yet accessible to non-expert INLA users.

A more classic alternative is the two-stage approach (Self and Pawitan 1992; Tsiatis et al. 1995). In this case, the longitudinal submodel is fitted separately and then the shared parameters and random effects estimates are used into the survival submodel. Several authors have applied this strategy in order to circumvent the identifiability issues and reduce processing time (Murawska et al. 2012; Guler et al. 2017; Donnelly et al. 2018; Mauff et al. 2020). Nevertheless, the main disadvantage of this approach is that the survival model parameter estimates are often biased (Tsiatis and Davidian 2004; Rizopoulos 2012).

From the Bayesian perspective, some bias correction strategies have been proposed for the two-stage approach (Tsiatis et al. 1995; Wulfsohn and Tsiatis 1997; Viviani et al. 2014; Leiva-Yamaguchi and Alvares 2021). Still, Markov chain Monte Carlo (MCMC) methods are quite demanding and their use may be impractical (Ye et al. 2008; Andrinopoulou et al. 2018).

In this paper, we propose a novel two-stage approach that uses the estimations from the longitudinal submodel to specify an informative prior distribution for the random effects when estimating them within the survival submodel. In addition, as a bias correction mechanism, we incorporate longitudinal likelihood function in the second stage, where its fixed effects are set according to the estimation using only the longitudinal submodel.

The rest of the paper is organised as follows. Section 2 introduces a generic formulation for Bayesian joint models of longitudinal and survival data. Section 3 presents a standard two-stage approach to making the inference in joint modelling framework. Section 4 describes our two-stage proposal. Sections 5 and 6 compare the bias and computational time for the three inferential strategies presented using simulated and real data, respectively. Finally, Sect. 7 discusses the results obtained and potential extensions of our methodology.

## 2 Bayesian joint model formulation

Bayesian joint models for longitudinal and survival data usually assume a full joint probability distribution given by (Armero et al. 2018; Alvares et al. 2021):

$$f(\mathbf{y}, s, \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s, \boldsymbol{\theta}_b) = f(\mathbf{y}, s \mid \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) f(\mathbf{b} \mid \boldsymbol{\theta}_b) \pi(\boldsymbol{\theta}_y) \pi(\boldsymbol{\theta}_s) \pi(\boldsymbol{\theta}_b), \quad (1)$$

where  $\mathbf{y}$  and  $s$  represent the longitudinal and the survival process, respectively. Random effects are denoted by  $\mathbf{b}$  and

have  $\boldsymbol{\theta}_b$  as their hyperparameters;  $\boldsymbol{\theta}_y$  and  $\boldsymbol{\theta}_s$  are the parameters and hyperparameters of each process. The factors on the right hand side of (1) are the conditional joint distribution of the processes  $\mathbf{y}$  and  $s$  given  $\mathbf{b}$ ,  $\boldsymbol{\theta}_y$ , and  $\boldsymbol{\theta}_s$ ,  $f(\mathbf{y}, s \mid \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s)$ ; the conditional distribution of  $\mathbf{b}$  given  $\boldsymbol{\theta}_b$ ,  $f(\mathbf{b} \mid \boldsymbol{\theta}_b)$ ; and the (independent) prior distributions of  $\boldsymbol{\theta}_y$ ,  $\boldsymbol{\theta}_s$  and  $\boldsymbol{\theta}_b$ ,  $\pi(\boldsymbol{\theta}_y)$ ,  $\pi(\boldsymbol{\theta}_s)$  and  $\pi(\boldsymbol{\theta}_b)$ , respectively.

Typically, the conditional joint distribution of  $\mathbf{y}$  and  $s$  is factorised as (Wu and Carroll 1988):

$$f(\mathbf{y}, s \mid \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) = f(\mathbf{y} \mid \mathbf{b}, \boldsymbol{\theta}_y) f(s \mid \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s), \quad (2)$$

where both processes are considered as conditionally independent given  $\mathbf{b}$ ,  $\boldsymbol{\theta}_y$ , and  $\boldsymbol{\theta}_s$ . Note that if the longitudinal parameters and hyperparameters,  $\boldsymbol{\theta}_y$ , are not shared with the survival process, then  $f(s \mid \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) = f(s \mid \mathbf{b}, \boldsymbol{\theta}_s)$ . In the next subsections, we will give more details about components from (1) and (2).

### 2.1 Longitudinal submodel

Typically, the terms  $f(\mathbf{y} \mid \mathbf{b}, \boldsymbol{\theta}_y)$  in (2) and  $f(\mathbf{b} \mid \boldsymbol{\theta}_b)$  in (1) for an individual  $i$  at time  $t$  are expressed through a generalised linear mixed specification:

$$\begin{aligned} y_i(t \mid \mathbf{b}_i, \boldsymbol{\theta}_y) &\sim EF(\mu_i(t \mid \mathbf{b}_i, \boldsymbol{\theta}_y)), \\ (\mathbf{b}_i \mid \boldsymbol{\theta}_b) &\sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \quad (3)$$

where  $EF$  represents an exponential family distribution with mean  $\mu_i(t \mid \cdot)$ ;  $\boldsymbol{\theta}_y$  contains the parameters and hyperparameters of  $\mu_i(t \mid \cdot)$ ; and  $\boldsymbol{\theta}_b = \boldsymbol{\Sigma}$ . The mean trajectory function can be formulated in different ways, but the most common corresponds to the linear mixed model specification with  $\mu_i(t \mid \cdot) = \mathbf{x}_i^\top(t) \boldsymbol{\beta} + \mathbf{z}_i^\top(t) \mathbf{b}_i$ , where  $\mathbf{x}_i(t)$  and  $\mathbf{z}_i(t)$  are covariates of the individual  $i$  at time  $t$ ,  $\boldsymbol{\beta}$  is the vector of fixed effects, and  $\mathbf{b}_i$  represents a  $K$ -dimensional vector of individual random effects (Pinheiro and Bates 2000).

### 2.2 Survival submodel

We will introduce some key elements of survival analysis in order to characterise the term  $f(s \mid \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s)$  in (2) using a hazard function. Let  $T_i^*$  denote the event time for individual  $i$ ,  $C_i$  the censoring time,  $T_i = \min\{T_i^*, C_i\}$  the observed time, and  $\delta_i = \mathbb{I}(T_i^* \leq C_i)$  the event indicator. The proportional hazards specification is commonly used to model these problems (Kumar and Klefsjö 1994) and at the same time incorporate the longitudinal information from submodel (3). So, the hazard function of the survival time  $T_i$  of individual  $i$  is expressed by:

$$h_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) = h_0(t | \boldsymbol{\theta}_s) \exp \left\{ \mathbf{x}_i^\top \boldsymbol{\gamma} + \alpha g_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y) \right\}, \tag{4}$$

where  $h_0(t | \boldsymbol{\theta}_s)$  represents an arbitrary baseline hazard function at time  $t$ ,  $\mathbf{x}_i$  is a covariate vector with coefficients  $\boldsymbol{\gamma}$ , and  $g_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y)$  is a temporal function of terms that comes from the longitudinal submodel and can be specified in different ways, such as  $\mu_i(t | \cdot)$ ,  $d\mu_i(t | \cdot)/dt$ , and  $\int_0^t \mu_i(v | \cdot) dt$  (Rizopoulos 2012). The component  $g_i(t | \cdot)$  has the role of connecting both processes, while  $\alpha$  measures the strength of this association. Finally,  $\boldsymbol{\theta}_s$  includes the parameters of  $h_0(t | \cdot)$ ,  $\boldsymbol{\gamma}$  and  $\alpha$ .

### 2.3 Prior distributions

We assume independent and diffuse marginal prior distributions (Gelman et al. 2013). More specifically, all longitudinal and survival regression coefficients (including the association parameter) follow Normal distributions with mean at zero and large variance;  $\sigma$  follows a weakly-informative half-Cauchy(0, 5) (Gelman 2006); and  $\boldsymbol{\Sigma}$  follows an inverse-Wishart( $V, r$ ), where  $V$  is a  $K \times K$  identity matrix,  $r = K$  is the degrees-of-freedom parameter (Schuurman et al. 2016). Once the baseline hazard function  $h_0(t | \boldsymbol{\theta}_s)$  is defined, diffuse priors are also specified for its parameters.

### 3 Standard two-stage (STS) approach

As a potential competitor, we use Tsiatis, De Gruttola and Wulfsohn (Tsiatis et al. 1995) approach adapted to the Bayesian framework (Leiva-Yamaguchi and Alvares 2021). Specifically, in the first stage, we calculate the *maximum a posteriori* (MAP) of the random effects  $\mathbf{b}$  and parameters  $\boldsymbol{\theta}_y$  estimated from longitudinal submodel fit separately. Posterior mean or median can also be used instead. In the second stage, we incorporate the temporal function into the survival submodel considering  $g_i(t | \hat{\mathbf{b}}_i, \hat{\boldsymbol{\theta}}_y)$ , where  $\hat{\mathbf{b}}_i$  and  $\hat{\boldsymbol{\theta}}_y$  are the MAP of  $\mathbf{b}_i$  and  $\boldsymbol{\theta}_y$ , and then the posterior distribution of  $\boldsymbol{\theta}_s$  is calculated.

### 4 Novel two-stage (NTS) approach

Like all two-stage methods for joint models, the first stage is to fit the longitudinal submodel separately. So, we calculate the MAP of longitudinal and random effects parameters, which will be denoted by  $\hat{\boldsymbol{\theta}}_y$  and  $\hat{\boldsymbol{\theta}}_b$ , respectively. Note that it is not necessary to estimate the random effects  $\mathbf{b}$  and therefore, we can run a Bayesian estimation algorithm using the marginalised likelihood function (integrating out the random effects) of the longitudinal submodel (3).

In the second stage, we estimate the random effects  $\mathbf{b}$  and parameters  $\boldsymbol{\theta}_s$  replacing the full joint probability distribution (1) with:

$$f(\mathbf{y}, \mathbf{s}, \mathbf{b}, \boldsymbol{\theta}_s | \hat{\boldsymbol{\theta}}_y, \hat{\boldsymbol{\theta}}_b) = f(\mathbf{y}, \mathbf{s} | \mathbf{b}, \hat{\boldsymbol{\theta}}_y, \boldsymbol{\theta}_s) f(\mathbf{b} | \hat{\boldsymbol{\theta}}_b) \pi(\boldsymbol{\theta}_s) \tag{5}$$

$$\stackrel{(2)}{=} f(\mathbf{y} | \mathbf{b}, \hat{\boldsymbol{\theta}}_y) f(\mathbf{s} | \mathbf{b}, \hat{\boldsymbol{\theta}}_y, \boldsymbol{\theta}_s) f(\mathbf{b} | \hat{\boldsymbol{\theta}}_b) \pi(\boldsymbol{\theta}_s),$$

where the  $f(\cdot)$ 's and  $\pi(\cdot)$  have the same functional form as in (1) and (2), except that the parameters  $\boldsymbol{\theta}_y$  and  $\boldsymbol{\theta}_b$  are now assumed to be known and so it is not necessary to include their prior distributions in (5). Note that when using  $\boldsymbol{\theta}_b$  estimated from the longitudinal submodel (i.e.,  $\hat{\boldsymbol{\theta}}_b$ ), random effects  $\mathbf{b}$  become fixed effects, further reducing the complexity of (5) with respect to the full joint probability distribution (1).

Essentially, our proposal simplifies/approximates the likelihood function from the joint approach by assuming that the shared parameters can be satisfactorily estimated from their MAP using the longitudinal submodel separately. In particular, as  $n_i$  grows, the longitudinal submodel  $f(\mathbf{y} | \mathbf{b}, \boldsymbol{\theta}_y)$  becomes the dominating part in the conditional joint likelihood function (see Eq. (2)), implying that the survival submodel is of minimal importance to obtain the MAP of  $\boldsymbol{\theta}_y$ . Then, the MAP of  $\boldsymbol{\theta}_s$  is estimated considering  $\boldsymbol{\theta}_y$  as a nuisance parameter (Murphy and van der Vaart 2000) and assuming that the contribution from random effects and priors in the posterior distribution is similar using either JS or NTS. See Appendix 1 for technical details.

### 5 Simulation studies

Simulation studies were performed to empirically corroborate that our methodological proposal (see Sect. 4) presents improvements to the dual problem “biased estimation versus high computational time” when compared to the joint specification (JS, see Sect. 2) and standard two-stage (STS, see Sect. 3) approaches. Algorithm 1 summarises a simulation scheme for joint models of longitudinal and survival data (Rizopoulos 2022; Austin 2012).

---

#### Algorithm 1 Simulation scheme for joint models.

---

- 1: **Initialisation:** Set  $\boldsymbol{\theta}_y, \boldsymbol{\theta}_s, \boldsymbol{\theta}_b, n, \Delta$ , and  $t_{\max}$ .
  - 2: **Simulating survival data:**
    - 3: – Simulate  $x_i \sim \text{Bernoulli}(0.5)$  and  $\mathbf{b}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}) \forall i$ .
    - 4: – Calculate  $T_i^* \forall i$  based on the survival submodel (4).
    - 5: – Simulate  $C_i \sim \text{Uniform}(0, t_{\max}) \forall i$ .
    - 6: – Set  $T_i = \min\{T_i^*, C_i\}$  and  $\delta_i = I(T_i^* \leq C_i) \forall i$ .
  - 7: **Simulating longitudinal data:**
    - 8: – Set  $0 = t_1, \dots, t_{n_i} \leq T_i \forall i$  such that  $t_{j+1} - t_j = \Delta$ .
    - 9: – Compute  $y_i(t_1), \dots, y_i(t_{n_i}) \forall i$  based on the longitudinal submodel (3).
-

In this algorithm,  $x_i$  denotes a binary covariate,  $\mathbf{b}_i$  the vector of random effects,  $T_i^*$  the event time (not always observed),  $C_i$  the censoring time,  $T_i$  the observed time, and  $\delta_i$  the event indicator. Additionally,  $n$ ,  $\Delta$ , and  $t_{\max}$  represent number of individuals, interval between longitudinal observations, and maximum observational time, respectively.

We analysed the joint formulation based on submodels (3) and (4) considering longitudinal continuous, count and binary responses, where the mean of the longitudinal submodel,  $\mu_i(t | \cdot)$ , was shared with the survival submodel.

### 5.1 Longitudinal continuous response

The longitudinal continuous process for an individual  $i$  at time  $t$  is written as a linear mixed-effects model:

$$\begin{aligned} y_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y) &\sim \text{Normal}(\mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y), \sigma^2), \\ \mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})t + \beta_2 x_i, \\ (\mathbf{b}_i | \boldsymbol{\theta}_b) &= (b_{0i}, b_{1i} | \boldsymbol{\theta}_b)^\top \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \quad (6)$$

where  $\boldsymbol{\theta}_y = (\beta_0, \beta_1, \beta_2, \sigma^2)^\top$  and  $\boldsymbol{\theta}_b = \boldsymbol{\Sigma}$ . The covariate  $x_i$  is a binary group indicator. The hazard function of the survival time  $T_i$  of individual  $i$  is specified as follows:

$$h_i(t | \mathbf{b}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) = \phi t^{\phi-1} \exp\{\gamma_0 + \gamma_1 x_i + \alpha \mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y)\}, \quad (7)$$

where  $h_0(t | \boldsymbol{\theta}_s)$  is a Weibull baseline hazard with  $\phi$  and  $\exp(\gamma_0)$  being shape and scale parameters, respectively, and  $\boldsymbol{\theta}_s = (\phi, \gamma_0, \gamma_1, \alpha)^\top$ .

### 5.2 Longitudinal count response

The longitudinal count process for an individual  $i$  at time  $t$  is written as a Poisson mixed-effects model:

$$\begin{aligned} y_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y) &\sim \text{Poisson}(\mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y)), \\ \log(\mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y)) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})t + \beta_2 x_i, \end{aligned} \quad (8)$$

where  $\boldsymbol{\theta}_y = (\beta_0, \beta_1, \beta_2)^\top$ , random effects are defined as in (6), and the survival submodel is specified as in (7).

### 5.3 Longitudinal binary response

Similar to the previous case, the longitudinal binary process for an individual  $i$  at time  $t$  is also written as a Bernoulli mixed-effects model:

$$\begin{aligned} y_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y) &\sim \text{Bernoulli}(\mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y)), \\ \text{logit}(\mu_i(t | \mathbf{b}_i, \boldsymbol{\theta}_y)) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})t + \beta_2 x_i, \end{aligned} \quad (9)$$

where  $\boldsymbol{\theta}_y = (\beta_0, \beta_1, \beta_2)^\top$ , random effects are defined as in (6), and the survival submodel is specified as in (7).

## 5.4 Simulation settings

For each scenario (Normal, Poisson, and Bernoulli), we simulated 300 datasets with  $n = 500$ ,  $\Delta = 1$ , and  $t_{\max} = 20$ . Table 1 presents the specified parameter values for each scenario.

Figure 1 shows the proportion of the number of longitudinal observations per individual from the simulated data using the parameter values from Table 1. The resulting censoring rates varied between 26% and 40% (Normal), 22% and 35% (Poisson), and 16% and 27% (Bernoulli).

## 5.5 Results

The longitudinal submodels for continuous, count and binary responses were set as in (6), (8) and (9), respectively, and the survival submodels as in (7). Prior distributions were specified as in Sect. 2.3. The MCMC configuration was defined as follows: 2000 iterations with warm-up of 1000 for the joint model using the joint specification (JS) and for the longitudinal submodel from the novel two-stage (NTS) and standard two-stage (STS) approaches. Additionally, 1000 iterations with warm-up of 500 were set to run the survival submodel from both two-stage approaches. All models were implemented using the `rstan` R-package version 2.21.7 (Stan Development Team 2022). Simulations were performed on a Dell laptop with 2.2 GHz Intel Core i7, 32 GB RAM, OS Windows.

Table 2 displays the running times of each estimation approach.

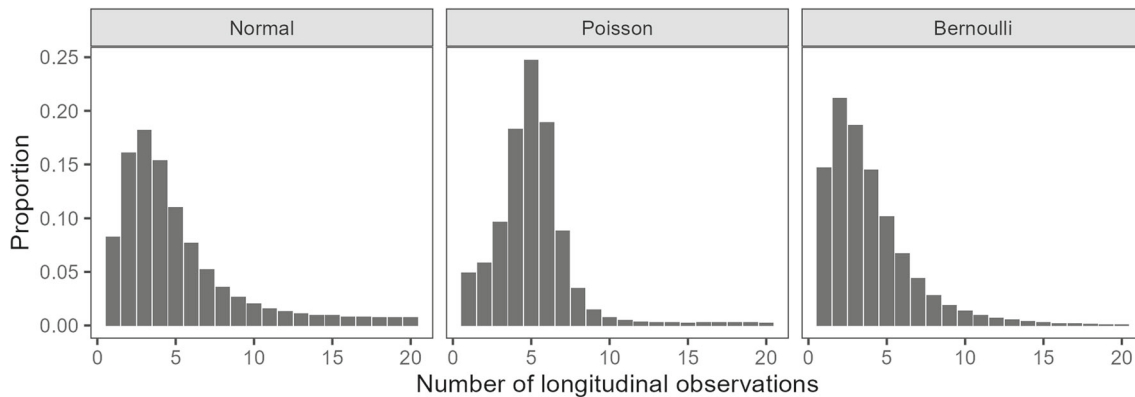
Unsurprisingly, STS had the lowest computational time, while the NTS required less processing time than the JS approach.

From an inferential perspective, we analysed the posterior mean for the group ( $\gamma_1$ ) and association ( $\alpha$ ) parameters in order to study the (potential) bias produced by two-stage approaches by ignoring the joint nature between the longitudinal and survival processes. Figure 2 shows the bias (posterior mean - true parameter value) and standard deviation for the group parameter using the JS, NTS and STS estimation approaches in each simulated scenario.

The group parameter was robustly estimated by the three approaches in all scenarios. However, as expected, the standard deviation from the STS estimators is smaller than the competitors. This happens because sharing the MAP of random effects without correction causes losses in the variability of such components, leading to an artificial and wrong reduction in the uncertainty of survival parameter estimators. Furthermore, it is important to note that NTS and JS produced nearly equivalent standard deviations. Figure 3 shows the bias and standard deviation for the association parameter using the JS, NTS and STS estimation approaches in each simulated scenario.

**Table 1** Setting true parameter values for each scenario

Scenario	$\beta_0$	$\beta_1$	$\beta_2$	$\sigma$	$\Sigma_{11}$	$\Sigma_{22}$	$\gamma_0$	$\gamma_1$	$\phi$	$\alpha$
Normal	-1.00	-0.10	-0.30	0.30	0.10	0.01	-1.00	-0.50	2.50	2.00
Poisson	-1.00	0.03	0.50	-	0.05	0.01	-7.00	-0.50	5.00	-2.00
Bernoulli	-3.00	2.00	-5.00	-	10.00	2.00	-3.00	-1.00	1.50	3.00



**Fig. 1** Proportion of the number of longitudinal observations per individual from 300 simulated datasets for each scenario

**Table 2** Average computational time (in minutes) from each estimation approach by simulated scenario

Scenario	JS	NTS	STS
Normal	5.65	2.70	1.65
Poisson	7.68	5.90	2.10
Bernoulli	8.45	4.25	1.38

The association parameter was unbiased using JS and NTS, while the STS approach underestimated the positive associations ( $\alpha = 2$  for Normal and  $\alpha = 3$  for Bernoulli) and overestimated the negative one ( $\alpha = -2$  for Poisson). Also, our proposal presented standard deviations for the association parameter with insignificant differences in relation to those from the JS estimator, while STS again reduced such standard deviations.

The reduction of complexity of the inferential process by using two-stage strategies was quite clear when analysing computational times. However, only our proposal corrected the estimation bias without affecting the estimator variance, producing results equivalent to the JS approach.

Appendix 1 shows complementary results using the same scenarios as in Table 1, but increasing the number of repeated measurements per individual ( $30 \leq n_i \leq 50$ ). Once again, our two-stage strategy corrected the estimation bias as well as kept the standard deviation of the estimator equivalent to that from the JS approach.

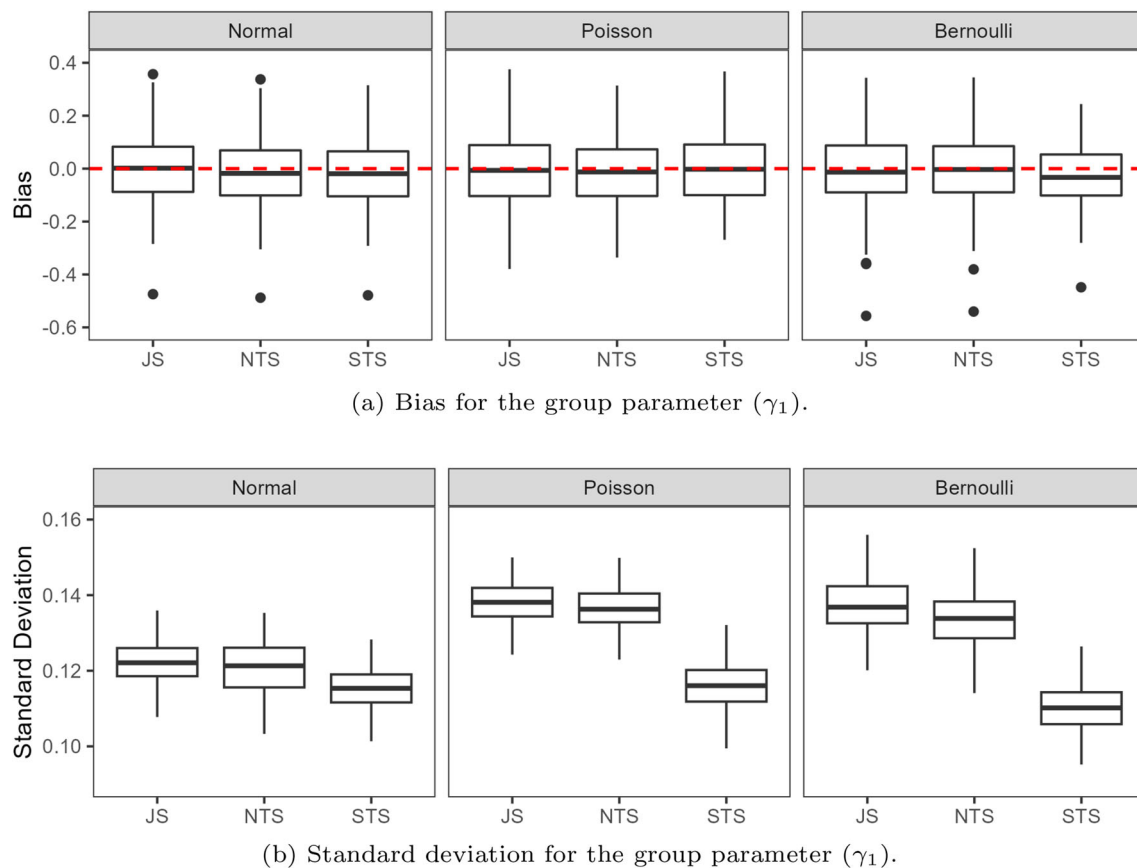
## 6 Applications

We used three real datasets, which are available in the JM R-package version 1.5-2 (Rizopoulos 2022), in order to

illustrate the performance of the novel two-stage (NTS, see Sect. 4) against the joint specification (JS, see Sect. 2) and the standard two-stage (STS, see Sect. 3) approach. A brief description of each dataset is presented below:

- **Acquired immunodeficiency syndrome (aids):** This dataset includes information about 467 patients with advanced human immunodeficiency virus infection during antiretroviral treatment who had failed or were intolerant to zidovudine therapy (Goldman et al. 1996). The longitudinal variable “CD4” was transformed by applying the square root; and “drug” was set as a group variable (zalcitabine = 0 and didanosine = 1).
- **Liver cirrhosis (prothro):** This dataset includes 488 patients with histologically verified liver cirrhosis, with 237 patients randomised to treatment with prednisone and the remaining received placebo (Andersen et al. 1993). The longitudinal variable “pro” was divided by 10 and truncated to the nearest integer; and “treat” was set as a group variable (placebo = 0 and prednisone = 1).
- **Primary biliary cirrhosis (pbc):** This dataset includes 312 patients with primary biliary cirrhosis, a rare autoimmune liver disease, at Mayo Clinic, where 158 patients were randomised to D-penicillamine and 154 to placebo (Murtaugh et al. 1994). The longitudinal variable “ser-Bilir” was dichotomised into  $\leq 2$  (0, control group) or  $> 2$  (1, late stage of PBC); and “drug” was set as a group variable (placebo = 0 and D-penicil = 1).

Figure 4 shows the frequency of the number of longitudinal observations per individual from the three datasets.



**Fig. 2** Simulation results from 300 datasets comparing the joint specification (JS), the novel two-stage (NTS), and the standard two-stage (STS) for longitudinal continuous, count, and binary responses. The panels show **a** bias and **b** standard deviation for the group parameter

Censoring rates are 59.74% (aids), 40.16% (prothro), and 55.13% (pbc).

Based on the nature of the longitudinal response type of each dataset, the joint models are specified as follows: (6) and (7) for aids dataset; (8) and (7) for prothro dataset; and (9) and (7) for pbc dataset. Prior distributions and MCMC configuration were defined as in Sect. 5.5. We provide the code for reproducibility at [www.github.com/daniloalvares/BJM-NTS2](http://www.github.com/daniloalvares/BJM-NTS2).

Table 3 shows the posterior summary of the group  $(\gamma_1)$  and association  $(\alpha)$  parameters estimated from JS, NTS and STS approaches for each dataset as well as the average computational time.

In the three applications, the group parameters  $(\gamma_1)$  are quite equivalent, except for small differences in the posterior mean using the STS approach, but their 95% credible intervals are similar to those of the other approaches. On the other hand, the biased estimation using the STS approach is more notorious for the association parameter  $(\alpha)$ . When using the Normal submodel (aids dataset), JS and NTS are almost identical, while STS has a slightly higher point estimate and a narrower 95% credible interval. Equivalence between JS and NTS is maintained for the joint models for count (prothro)

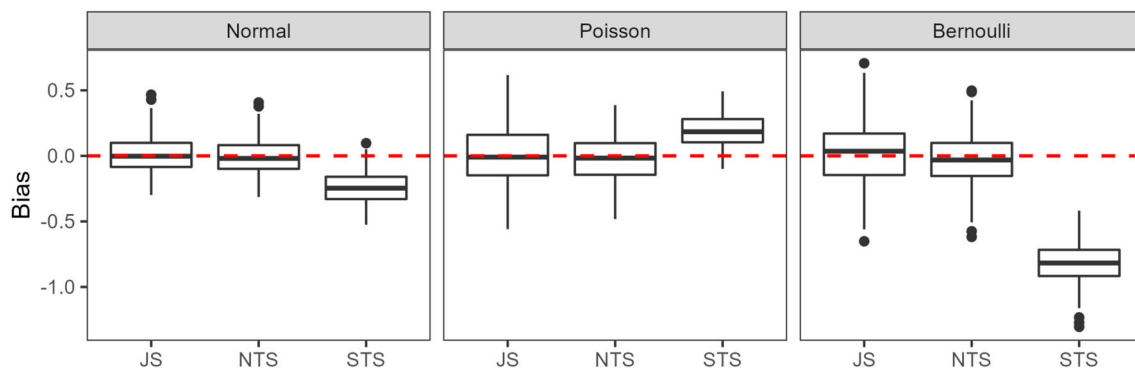
and binary (pbc) data, and in both cases the estimation bias using STS is more prominent, in which the credible intervals do not include the posterior mean of  $\alpha$  calculated by the JS approach (reference estimation). Again, STS remains the fastest option, while NTS has reasonable average times and significantly lower than those of the JS approach.

## 7 Discussion

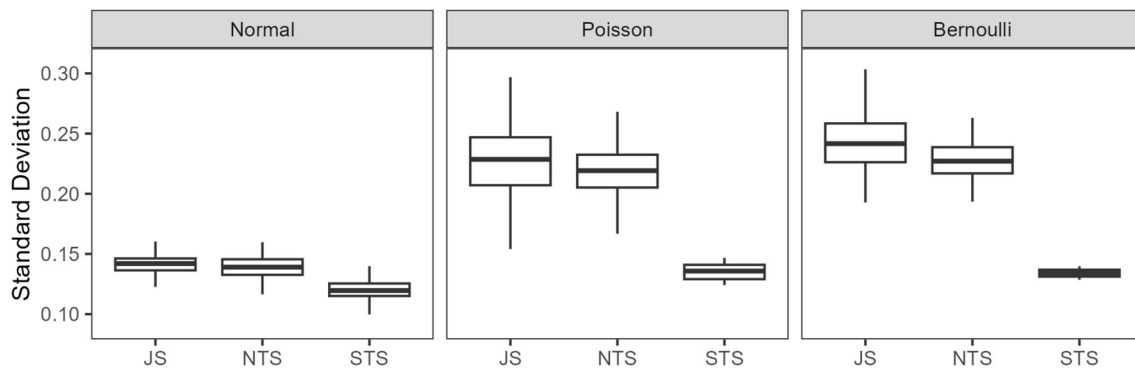
We have proposed a novel two-stage (NTS) approach that can be used in a Bayesian joint analysis of longitudinal and survival data. We have simplified the complex nature of the joint specification (JS) by using a two-stage strategy that at the same time corrects the well-known bias from a standard two-stage (STS) approach.

NTS was shown to be successful, producing satisfactory results both in simulated scenarios and in real applications considering different types of longitudinal responses (continuous, count and binary) that share information with a Weibull proportional hazard model.

In practical terms, the group parameter  $(\gamma_1)$  has not been significantly affected by the use of two-stage approaches.

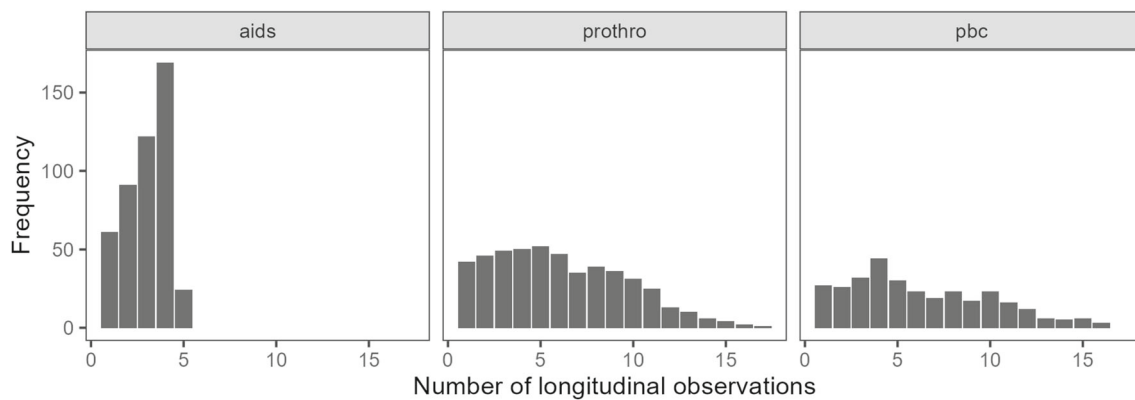


(a) Bias for the association parameter ( $\alpha$ ).



(b) Standard deviation for the association parameter ( $\alpha$ ).

**Fig. 3** Simulation results from 300 datasets comparing the joint specification (JS), the novel two-stage (NTS), and the standard two-stage (STS) for longitudinal continuous, count, and binary responses. The panels show **a** bias and **b** standard deviation for the association parameter



**Fig. 4** Frequency of the number of longitudinal observations per individual for each dataset

However, we have once again observed that the association parameter ( $\alpha$ ) is biased when the STS is used. This issue is mitigated by replacing STS with NTS. Furthermore, we highlight that the strength of the association does not affect the NTS estimation, even with few longitudinal observations per individual. In particular, the strong association presented in the simulation studies and also observed in the pbc dataset modelling did not produce bias, as typically happens with

regression calibration approaches (Nevo et al. 2020). Moreover, our proposal had a much lower computational time than the JS approach in all scenarios.

Theoretically, the correction of shared information should be enough so that all other characteristics of the survival model (baseline hazard and coefficients of baseline variables) do not suffer estimation bias compared to the JS. However, in this paper, we have not investigated different baseline hazard

**Table 3** Posterior summary (posterior mean and 95% credible interval) and average computational time (in minutes) from each estimation approach by dataset (and longitudinal response type)

Dataset	Longitudinal	Approach	$\gamma_1$ (Group)	$\alpha$ (Association)	Time
Aids	Normal	JS	0.34(0.04, 0.63)	-0.29(-0.36, -0.22)	4.20
		NTS	0.34(0.07, 0.67)	-0.28(-0.35, -0.21)	3.24
		STS	0.33(-0.01, 0.66)	-0.24(-0.30, -0.19)	1.37
Prothro	Poisson	JS	-0.26(-0.53, 0.03)	-0.50(-0.61, -0.39)	7.11
		NTS	-0.23(-0.47, 0.05)	-0.45(-0.57, -0.34)	3.92
		STS	-0.16(-0.41, 0.09)	-0.35(-0.42, -0.27)	2.15
pbc	Bernoulli	JS	-0.03(-0.39, 0.34)	3.67(2.89, 4.60)	5.30
		NTS	-0.03(-0.37, 0.28)	3.65(2.91, 4.45)	4.14
		STS	-0.01(-0.33, 0.32)	3.04(2.46, 3.60)	1.31

specifications to corroborate such behaviour, but we hope to do so in future works.

Another advantage of the proposed two-stage approach is that it can be easily generalised to more complex longitudinal (e.g., skewed or multiple longitudinal data) and survival (e.g., competing-risks or multistate data) submodels.

**Acknowledgements** Danilo Alvares was supported by the Chilean National Fund for Scientific and Technological Development (FONDECYT), Grant No. 11190018, and by the UKRI Medical Research Council, Grant No. MC\_UU\_00002/5.

**Author Contributions** DA and VLY contributed equally to the writing of the manuscript and implementation of the methodology. All authors reviewed the manuscript.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A

Let  $f(y_i | \mathbf{b}_i, \boldsymbol{\theta}_y)$  and  $f(s_i | \mathbf{b}_i, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s)$  be the longitudinal and survival likelihood functions for individual  $i$ , respectively, where  $\mathbf{y}_i = \{y_{it_{ij}} : t_{ij} = 1, \dots, n_i\}$ . We will analyse the behaviour of the *maximum a posteriori* (MAP) of  $\boldsymbol{\theta}_y$  from the conditional joint log-likelihood function of individual  $i$ , as  $n_i$  grows:

$$\begin{aligned}
 l_{n_i}(\boldsymbol{\theta}_y) &= \log \left( \prod_{j=1}^{n_i} f(y_{it_{ij}} | \mathbf{b}_i, \boldsymbol{\theta}_y) \right) \\
 &\quad + \log \left( f(s_i | \mathbf{b}_i, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) \right) \\
 &= \sum_{j=1}^{n_i} \log \left( f(y_{it_{ij}} | \mathbf{b}_i, \boldsymbol{\theta}_y) \right) \\
 &\quad + \log \left( f(s_i | \mathbf{b}_i, \boldsymbol{\theta}_y, \boldsymbol{\theta}_s) \right). \tag{A1}
 \end{aligned}$$

Calculating  $\lim_{n_i \rightarrow \infty} [l_{n_i}(\boldsymbol{\theta}_y)/n_i]$ , we note that the longitudinal term is  $\text{Op}(1)$ , while the survival one is  $\text{op}(1)$ . Extending this result to the joint likelihood for all individuals is straightforward. In summary, as  $n_i$  grows, the MAP of  $\boldsymbol{\theta}_y$  (let's say  $\hat{\boldsymbol{\theta}}_y$ ) depends only on the longitudinal submodel. In this sense, we can calculate  $\hat{\boldsymbol{\theta}}_y$  both from the JS and from the NTS first stage equivalently. It is noteworthy that this result only applies to  $\hat{\boldsymbol{\theta}}_y$ , i.e., there are no guarantees of equivalence for the MAP of  $\mathbf{b}_i$  (let's say  $\hat{\mathbf{b}}_i$ ) from the first stage. This is a key difference between our proposal and STS. While NTS needs the MAP of  $\boldsymbol{\theta}_y$  and  $\boldsymbol{\theta}_s$  from the first stage to then correct the random effects in the second one, STS uses  $\hat{\boldsymbol{\theta}}_y$  and  $\hat{\mathbf{b}}_i$  (both calculated from the longitudinal submodel) in the second stage. Therefore, the bias produced by STS when estimating  $\boldsymbol{\theta}_s$  is mainly due to uncorrected random effects.

Taking the MAP of  $\boldsymbol{\theta}_y$  from the NTS approach, our next goal is to calculate the MAP of  $\boldsymbol{\theta}_s$ . Hence, similar to a profiled likelihood strategy (Murphy and van der Vaart 2000), we can consider  $\hat{\boldsymbol{\theta}}_y$  as a vector of nuisance parameters and then estimate the MAP of  $\boldsymbol{\theta}_s$  from the conditional joint log-likelihood function:

$$l(\boldsymbol{\theta}_s) = \log \left( f(\mathbf{y} | \mathbf{b}, \hat{\boldsymbol{\theta}}_y) \right) + \log \left( f(s | \mathbf{b}, \hat{\boldsymbol{\theta}}_y, \boldsymbol{\theta}_s) \right). \tag{A2}$$

The MAP of  $(\boldsymbol{\theta}_y, \boldsymbol{\theta}_s)$  jointly estimated from the conditional joint log-likelihood function is equivalent to calculating the MAP of  $\boldsymbol{\theta}_y$  from  $l_{n_i}(\boldsymbol{\theta}_y)$  (assuming that  $n_i \rightarrow \infty$ ) and then obtaining the MAP of  $\boldsymbol{\theta}_s$  from  $l(\boldsymbol{\theta}_s)$ . Note that Eq. (A2)



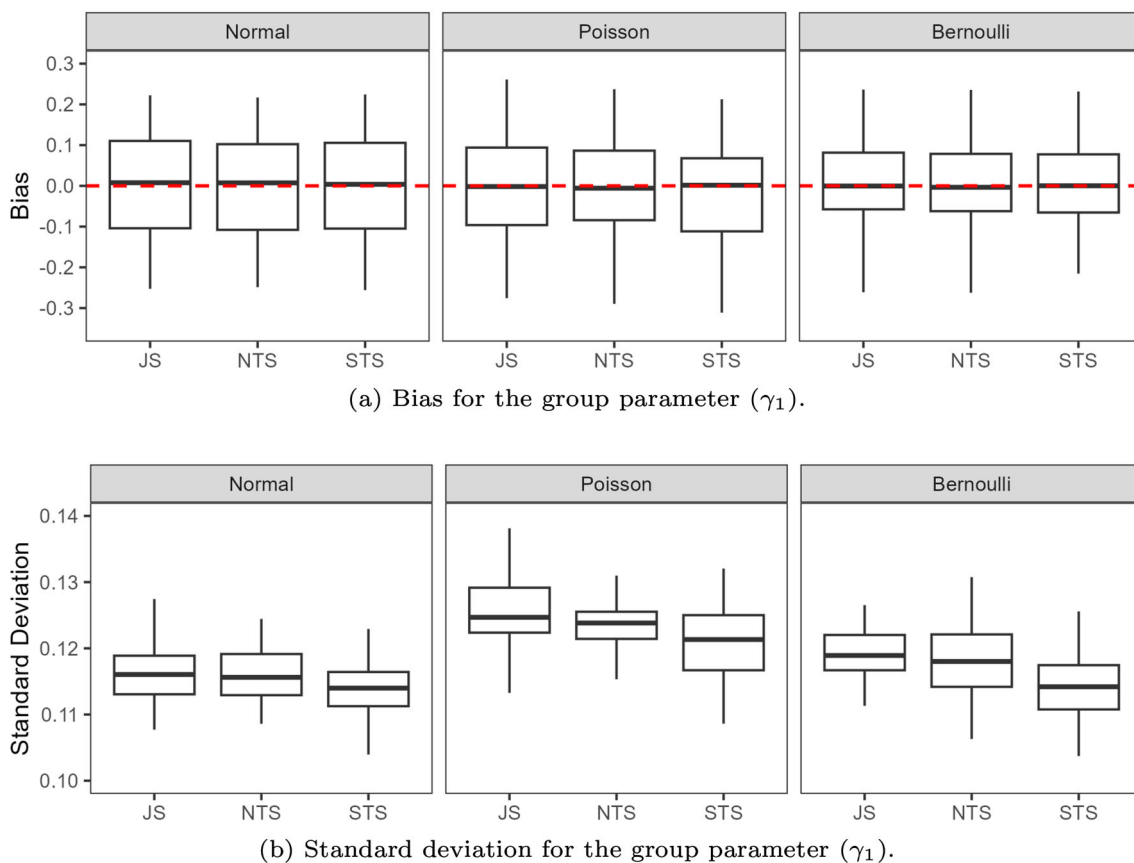
is specified as the log-likelihood function in NTS second stage.

So far we have omitted the specification of random effects and priors for the estimation of the MAP of  $\theta_y$  and  $\theta_s$ . We assume that the distribution of  $\mathbf{b}$ ,  $f(\mathbf{b} | \theta_b)$ , from the JS approach is the same as the prior distribution of fixed effects  $\mathbf{b}$ ,  $\pi(\mathbf{b} | \hat{\theta}_b)$ , from the NTS approach. So, the contribution of random effects to calculate the MAP of  $\theta_y$  and  $\theta_s$  from both approaches is the same. Additionally, even if the random effects distributions are misspecified, Rizopoulos, Verbeke and Molenberghs' Theorem 1 (Rizopoulos et al. 2008) demonstrates that, as  $n_i$  grows, the estimates of  $\theta_y$  and  $\theta_s$  will be minimally affected. On the other hand, we assume that the prior distributions are proper and weakly-informative, so

they must have an irrelevant role in calculating the MAP of  $\theta_y$  and  $\theta_s$ . Furthermore, as the sample size increases, the choice of prior distributions have a minimal impact.

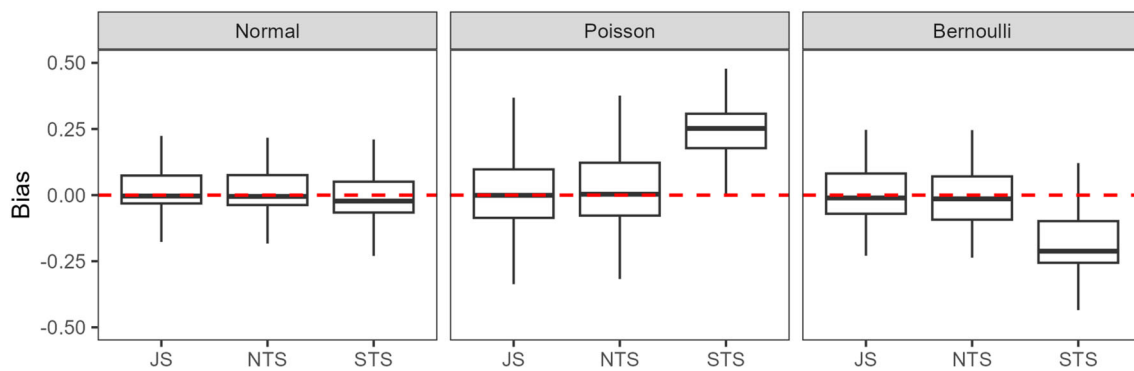
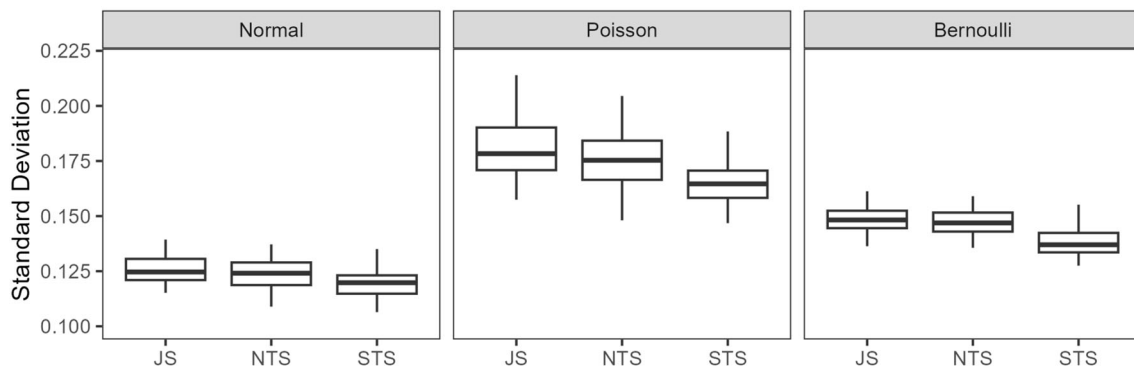
### Appendix B

See Figs. 5 and 6



**Fig. 5** Simulation results from 300 datasets comparing the joint specification (JS), the novel two-stage (NTS), and the standard two-stage (STS) for longitudinal continuous, count, and binary responses. The

panels show **a** bias and **b** standard deviation for the group parameter. Number  $n_i$  of repeated measurements per individual varying between 30 and 50

(a) Bias for the association parameter ( $\alpha$ ).(b) Standard deviation for the association parameter ( $\alpha$ ).

**Fig. 6** Simulation results from 300 datasets comparing the joint specification (JS), the novel two-stage (NTS), and the standard two-stage (STS) for longitudinal continuous, count, and binary responses. The

panels show **a** bias and **b** standard deviation for the association parameter. Number  $n_i$  of repeated measurements per individual varying between 30 and 50

## References

- Alvares, D., Rubio, F.J.: A tractable Bayesian joint model for longitudinal and survival data. *Stat. Med.* **40**(19), 4213–4229 (2021)
- Alvares, D., Armero, C., Forte, A., Chopin, N.: Sequential Monte Carlo methods in Bayesian joint models for longitudinal and time-to-event data. *Stat. Model.* **21**(1–2), 161–181 (2021)
- Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N.: *Statistical models based on counting processes*, 1st edn. Springer, New York (1993)
- Andrinopoulou, E.R., Eilers, P.H.C., Takkenberg, J., Rizopoulos, D.: Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics* **74**(2), 685–693 (2018)
- Armero, C., Forte, A., Perpiñán, H., Sanahuja, M.J., Agustí, S.: Bayesian joint modeling for assessing the progression of chronic kidney disease in children. *Stat. Methods Med. Res.* **27**(1), 298–311 (2018)
- Austin, P.C.: Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Stat. Med.* **31**(29), 3946–3958 (2012)
- Barrett, J., Diggle, P., Henderson, R., Taylor-Robinson, D.: Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **77**(1), 131–148 (2015)
- Donnelly, C., McFetridge, L.M., Marshall, A.H., Mitchell, H.J.: A two-stage approach to the joint analysis of longitudinal and survival data utilising the Coxian phase-type distribution. *Stat. Methods Med. Res.* **27**(12), 3577–3594 (2018)
- Gelman, A.: Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **1**(3), 515–534 (2006)
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: *Bayesian data analysis*, 3rd edn. Chapman & Hall, New York (2013)
- Goldman, A.I., Carlin, B., Crane, L.R., Launer, C., Korvick, J.A., Deyton, L., Abrams, D.I.: Response of CD4 lymphocytes and clinical consequences of treatment using ddI or ddC in patients with advanced HIV infection. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **11**(2), 161–169 (1996)
- Gould, A.L., Boye, M.E., Crowther, M.J., Ibrahim, J.G., Quartey, G., Micallef, S., Bois, F.Y.: Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Stat. Med.* **34**(14), 2181–2195 (2015)
- Guler, I., Faes, C., Cadarso-Suárez, C., Teixeira, L., Rodrigues, A., Mendonça, D.: Two-stage model for multivariate longitudinal and survival data with application to nephrology research. *Biom. J.* **59**(6), 1204–1220 (2017)
- Ibrahim, J.G., Molenberghs, G.: Missing data methods in longitudinal studies: a review. *TEST* **18**(1), 1–43 (2009)
- Ibrahim, J.G., Chu, H., Chen, L.M.: Basic concepts and methods for joint models of longitudinal and survival data. *J. Clin. Oncol.* **28**(16), 2796–2801 (2010)

- Kumar, D., Klefsjö, B.: Proportional hazards model: a review. *Reliab. Eng. Syst. Safety* **44**, 177–188 (1994)
- Leiva-Yamaguchi, V., Alvares, D.: A two-stage approach for Bayesian joint models of longitudinal and survival data: correcting bias with informative prior. *Entropy* **23**, 1–10 (2021)
- Lesaffre, E., Spiessens, B.: On the effect of the number of quadrature points in a logistic random effects model: an example. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)* **50**(3), 325–335 (2001)
- Mauff, K., Steyerberg, E., Kardys, I., Boersma, E., Rizopoulos, D.: Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. *Stat. Comput.* **30**, 999–1014 (2020)
- Murawska, M., Rizopoulos, D., Lesaffre, E.: A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *J. Probab. Stat.* **2012**, 1–18 (2012)
- Murphy, S.A., van der Vaart, A.W.: On profile likelihood. *J. Am. Stat. Assoc.* **95**(450), 449–465 (2000)
- Murtaugh, P.A., Dickson, E.R., van Dam, G.M., Malinchoc, M., Grambsch, P.M., Langworthy, A.L., Gips, C.H.: Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology* **20**(1), 126–134 (1994)
- Muthén, B., Asparouhov, T., Boye, M.E., Hackshaw, M., Naegeli, A.: Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus. Muthén & Muthén, Technical report (2009)
- Neuhaus, A., Augustin, T., Heumann, C., Daumer, M.: A review on joint models in biometrical research. *J. Stat. Theory Practice* **3**(4), 855–868 (2009)
- Nevo, D., Hamada, T., Ogino, S., Wang, M.: A novel calibration framework for survival analysis when a binary covariate is measured at sparse time points. *Biostatistics* **21**(2), 148–163 (2020)
- Papageorgiou, G., Mauff, K., Tomer, A., Rizopoulos, D.: An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Rev. Stat. Appl.* **6**(1), 223–240 (2019)
- Pinheiro, J.C., Bates, D.M.: Linear mixed-effects models: basic concepts and examples. In: *Mixed-effects models in S and S-Plus*, pp. 3–56. Springer, New York (2000)
- Pinheiro, J.C., Chao, E.C.: Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *J. Comput. Graph. Stat.* **15**(1), 58–81 (2006)
- Rizopoulos, D.: JM: joint modeling of longitudinal and survival data. R Package Version 1.5-2. <https://cran.r-project.org/package=JM> (2022)
- Rizopoulos, D.: Joint models for longitudinal and time-to-event data: with applications in R, 1st edn. Chapman & Hall/CRC, Boca Raton (2012)
- Rizopoulos, D., Verbeke, G., Molenberghs, G.: Shared parameter models under random effects misspecification. *Biometrika* **95**(1), 63–74 (2008)
- Rizopoulos, D., Verbeke, G., Lesaffre, E.: Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *J. Royal Stat. Soci.: Ser. B (Stat. Methodol.)* **71**(3), 637–654 (2009)
- Schuurman, N.K., Grasman, R.P.P.P., Hamaker, E.L.: A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivar. Behav. Res.* **51**(2–3), 185–206 (2016)
- Self, S., Pawitan, Y.: Modeling a marker of disease progression and onset of disease. In: *AIDS Epidemiology: methodological Issues*, pp. 231–255. Springer, New York (1992)
- Stan Development Team: rstan: R Interface to Stan. R Package Version 2.21.7. <http://mc-stan.org> (2022)
- Tsiatis, A.A., Davidian, M.: Joint modeling of longitudinal and time-to-event data: an overview. *Stat. Sin.* **14**(3), 809–834 (2004)
- Tsiatis, A.A., De Gruttola, V., Wulfsohn, M.S.: Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J. Am. Stat. Assoc.* **90**(429), 27–37 (1995)
- van Niekerk, J., Bakka, H., Rue, H.: Competing risks joint models using R-INLA. *Stat. Model.* **21**(1–2), 56–71 (2020)
- Viviani, S., Alfó, M., Rizopoulos, D.: Generalized linear mixed joint model for longitudinal and survival outcomes. *Stat. Comput.* **24**(3), 417–427 (2014)
- Wu, L.: Mixed effects models for complex data, 1st edn. Chapman & Hall/CRC, New York (2009)
- Wu, M.C., Carroll, R.J.: Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**(1), 175–188 (1988)
- Wu, L., Liu, W., Hu, X.J.: Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. *Biometrics* **66**(2), 327–335 (2010)
- Wu, L., Wei, L., Yi, G.Y., Huang, Y.: Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *J. Probab. Stat.* **2012**, 1–17 (2011)
- Wulfsohn, M.S., Tsiatis, A.A.: A joint model for survival and longitudinal data measured with error. *Biometrics* **53**(1), 330–339 (1997)
- Ye, W., Lin, X., Taylor, J.M.G.: Semiparametric modeling of longitudinal measurements and time-to-event data - a two-stage regression calibration approach. *Biometrics* **64**(4), 1238–1246 (2008)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.