Spatial joint models through Bayesian structured piecewise additive joint modelling for longitudinal and time-to-event data

Rappl, Anja; Kneib, Thomas; Lang, Stefan; Bergherr, Elisabeth

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Rappl, A., Kneib, T., Lang, S., & Bergherr, E. (2023). Spatial joint models through Bayesian structured piecewise additive joint modelling for longitudinal and time-to-event data. *Statistics and Computing*, *33*(6). https://doi.org/10.1007/s11222-023-10293-5

Statistics and Computing (2023) 33:135
https://doi.org/10.1007/s11222-023-10293-5

ORIGINAL PAPER

Spatial joint models through Bayesian structured piecewise additive joint modelling for longitudinal and time-to-event data

Anja Rappl[1] · Thomas Kneib[2] · Stefan Lang[3] · Elisabeth Bergherr[4]

Received: 9 February 2023 / Accepted: 30 August 2023

Abstract
Joint models for longitudinal and time-to-event data simultaneously model longitudinal and time-to-event information to avoid bias by combining usually a linear mixed model with a proportional hazards model. This model class has seen many developments in recent years, yet joint models including a spatial predictor are still rare and the traditional proportional hazards formulation of the time-to-event part of the model is accompanied by computational challenges. We propose a joint model with a piecewise exponential formulation of the hazard using the counting process representation of a hazard and structured additive predictors able to estimate (non-)linear, spatial and random effects. Its capabilities are assessed in a simulation study comparing our approach to an established one and highlighted by an example on physical functioning after cardiovascular events from the German Ageing Survey. The Structured Piecewise Additive Joint Model yielded good estimation performance, also and especially in spatial effects, while being double as fast as the chosen benchmark approach and performing stable in an imbalanced data setting with few events.

Keywords Bayesian statistics · Joint models · Piecewise additive mixed models · Piecewise exponential

1 Introduction

Biometrical studies often capture time-to-event and longitudinal data on the same topic simultaneously. And often these observation are endogenous in the sense that the longitudinal observations inform on the event and the event on the longitudinal information, which is ceased to be recorded with the occurrence of an event. Frequently used examples are the count of CD4 lymphocytes in HIV-positive patients and their time till onset of AIDS (Faucett and Thomas 1996; Rizopoulos 2011; Wulfsohn and Tsiatis 1997) or the level of serum bilirubin and other liver biomarkers in primary biliary cirrhosis patients and time to death (Crowther et al. 2013; Hickey et al. 2018). Other examples include PSA cancer marker and progression to recurrence of prostate cancer (Jacqmin-Gadda et al. 2010), autoantibody titers in children preceding the onset of Type 1 diabetes (Köhler et al. 2017) or physical functioning after a cardiovascular event and death (Rappl et al. 2022). Separate analysis of these longitudinal and time-to-event outcomes leads to biased estimates and to avoid this both should be modelled jointly. These joint models consist of two submodels: A longitudinal submodel and a survival submodel with both being linked through a subject specific predictor and an association parameter. This predictor may include a set of covariates relevant for both outcomes and becomes subject specific via random effects.

While Wulfsohn and Tsiatis (1997) and Henderson et al. (2000) proposed to maximize the likelihood of a joint model via an Expectation-Maximization (EM) algorithm, Faucett and Thomas (1996) used a Bayesian Gibbs-sampling approach. In recent years advances have been made into statistical boosting (Griesbach et al. 2021; Waldmann et al. 2017). Software is available for all three estimation approaches across various statistical computation platforms, of which especially `R` hosts a number of well-established packages such as `JM` (Rizopoulos 2010), `JMbayes` (Rizopoulos

✉ Anja Rappl
anja.rappl@fau.de

1 Institute of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

2 Chair of Statistics, Georg-August-Universität Göttingen, Göttingen, Germany

3 Department of Statistics, Universität Innsbruck, Innsbruck, Austria

4 Chair of Spatial Data Science and Statistical Learning, Georg-August-Universität Göttingen, Göttingen, Germany

Published online: 12 October 2023

2016), `joineRML` (Hickey et al. 2018) and `bamlss` (Umlauf et al. 2021). Comparisons of selections of available software can be found in Yuen and Mackinnon (2016) and Rappl et al. (2022).

Traditionally the longitudinal submodel of a joint model is a linear mixed model (LMM) and the time-to-event submodel is a proportional hazards (PH) model, though other variants are possible. Depending on the scaling of the longitudinal outcome a generalized linear mixed model (GLMM) (Faucett et al. 1998; Rizopoulos et al. 2008; Viviani et al. 2014) or quantile regression model (Huang and Chen 2016; Zhang et al. 2019) might be better suited. An alternative to the PH models in the time-to-event submodel are accelerated failure time models (Huang and Chen 2016; Tseng et al. 2005) and in certain data situations competing risks models are best suited (Andrinopoulou et al. 2014; Blanche et al. 2015; Huang et al. 2011). Also models with multivariate longitudinal outcomes are in use (Lin et al. 2002; Mauff et al. 2020; Rizopoulos and Ghosh 2011) as are location-scale models (Barrett et al. 2019). Köhler et al. (2018) expanded joint models to structured additive joint models with possibly smooth random effects and established non-linear association structures (2018) both via a Bayesian flexible tensor-product approach using Newton-Raphson procedures and derivative-based Metropolis-Hastings sampling. A good historic overview on joint models can be found in Tsiatis and Davidian (2004), while Alsefri et al. (2020) give a concise summary of recent developments in Bayesian joint models in particular.

From a perspective of longitudinal modelling joint models are known as shared parameter models and are used to account for "missing not at random" (MNAR) dropout. The idea was first presented by Wu and Carroll (1988) and Wu and Bailey (1988) with later expansions by Follmann and Wu (1995) and Hogan and Laird (1997). Roy (2003) proposes the use of latent classes, i.e. discrete random effects, while Tsonaka et al. (2009) leave the random effects unspecified altogether. More recent research on modelling longitudinal data including their dropout investigates the usage of hidden Markov models replacing the random effects (Bartolucci and Farcomeni 2015, 2019). Still joint models with a spatial component are rare. Martins et al. (2016, 2017) have described estimation of a Bayesian joint model with a spatial effect and a Weibull baseline hazard using OpenBUGS and WinBUGS respectively. The above mentioned Bayesian tensor-product approach by Köhler et al. (2017) implemented in the `R` package `bamlss` also has the capability of estimating spatial joint models. In terms of model formulation both methods have in common that they use a PH model for the survival submodel. However, assuming a parametric baseline hazard such as a Weibull hazard can be restrictive and derivative-based Metropolis-Hastings algorithms are computationally expensive as well as may prove sensitive towards data with few events.

Therefore, in this paper we propose a Bayesian joint model with a structured additive LMM for the longitudinal outcome, but exchange the time-to-event submodel for a piecewise additive mixed model (PAMM). The latter has been suggested by Bender et al. (2018) for modelling survival times based on the proportionality of a time-to-event process with a Poisson-distributed count process (Friedman 1982) thus expanding the available options for time-to-event models (e.g. accelerated failure times, competing risk). This formulation allows for estimation of the baseline hazard without any assumptions about its distributional form and is similarly flexible to the Köhler et al. (2018) model with respect to the inclusion of (non-)linear, spatial and random effects. At the same time, it reduces runtimes by about 50% (compared to an established method) and has proven stable in imbalanced data settings with few events.

The rest of the paper is structured as follows: In the next section, the methodology of piecewise additive joint models is described in more detail and our extension of the concept is explained. In section three the results of a simulation study comparing our approach to an established one to proof the feasibility of the model formulation, its ability to estimate spatial effects and its runtime performance. We then apply this method to an example of physical functioning from the German Aging Survey. Section five concludes with some final remarks and further technical details can be found in the Appendix.

2 Methods

2.1 Theoretical background

A joint model is applied for data collecting both longitudinal and time-to-event outcomes on the same topic, which inform each other in terms of drop-out. The model itself links a submodel for the longitudinal information to a submodel to the time-to-event information via an association factor and a subject specific predictor, which contains random effects to represent the subject specific variability and may contain a set of further explanatory covariates. Those covariates - and other covariates in the submodels - may follow (non-)linear, spatial or interaction effects.

In its original form the joint model assumes a linear mixed model (LMM) for the longitudinal outcome and a proportional hazards model (PH) for the time-to-event outcome (Faucett and Thomas 1996; Henderson et al. 2000; Wulfsohn and Tsiatis 1997).

Let $\boldsymbol{y}$ denote the vector of longitudinal outcomes across all individuals $i = \{1, \ldots, n\}$ and observations at time points t. Further, let $\boldsymbol{\lambda}(t)$ be the vector of individual specific risks

to experience an event at time t proportional to the baseline hazard $\lambda_0(t)$ and based on the observed event or censoring times $\boldsymbol{T}$ and event indicator $\boldsymbol{\delta}$. Then in its most generic variant the original joint model takes the form

$$\boldsymbol{y}(t) = \boldsymbol{\eta}_{\mathrm{l}}(t) + \boldsymbol{\eta}_{\mathrm{ls}}(t) + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, \sigma^2 \boldsymbol{I}) \tag{1}$$

$$\boldsymbol{\lambda}(t) = \lambda_0(t) \exp\{\boldsymbol{\eta}_{\mathrm{s}} + \alpha \boldsymbol{\eta}_{\mathrm{ls}}(t)\}, \tag{2}$$

where $\boldsymbol{\eta}_{\mathrm{s}}$ and $\boldsymbol{\eta}_{\mathrm{l}}(t)$ are survival and longitudinal submodel specific predictors respectively and $\boldsymbol{\eta}_{\mathrm{ls}}(t)$ is the shared predictor, via which both model parts are connected. The parameter α quantifies the association between the longitudinal and the time-to-event outcome. Also note that, while it is theoretically possible to estimate time-varying survival predictors $\boldsymbol{\eta}_{\mathrm{s}}(t)$, the time-varying covariates included in that predictor may be prone to measurement error and it is therefore in most cases better to model them jointly.

The predictors $\boldsymbol{\eta}_{.}$ are additive and may include (non-) linear, spatial or random effects of potentially time-varying covariates $\boldsymbol{x}_k(t)$, i.e. $\boldsymbol{\eta}_{.} = \sum_{k=1}^{p_{.}} f_k(\boldsymbol{x}_k(t))$, where f_k is a function representing the respective effect and $p_{.}$ denotes the predictor specific number of covariates. Restrictions apply to random effects, which need to be part of the shared predictor $\boldsymbol{\eta}_{\mathrm{ls}}(t)$, since it is the subject specific variability that mutually informs both the outcomes and spatial effects, of which there can only be one in the model for identifiability reasons.

Reformulating predictor $\boldsymbol{\eta}_{.}$ in matrix notation yields

$$\boldsymbol{\eta}_{.} = \boldsymbol{Z}_1 \boldsymbol{\gamma}_1 + \cdots + \boldsymbol{Z}_{p_{.}} \boldsymbol{\gamma}_{p_{.}}, \tag{3}$$

where $\boldsymbol{Z}_k$ is an effect appropriate design matrix and $\boldsymbol{\gamma}_k$ a vector of corresponding effect coefficients. For the Bayesian estimation of this model the generic prior for the coefficients $\boldsymbol{\gamma}_k$ is proportional to a normal distribution with zero mean, variance $\sigma^2_{\gamma_k}$ and penalty matrix $\boldsymbol{K}_k$

$$p\left(\boldsymbol{\gamma}_k \mid \sigma^2_{\gamma_k}\right) \propto \left(\sigma^2_{\gamma_k}\right)^{-\mathrm{rk}(\boldsymbol{K}_k)} \exp\left\{-\frac{1}{2\sigma^2_{\gamma_k}} \boldsymbol{\gamma}_k' \boldsymbol{K}_k \boldsymbol{\gamma}_k\right\}. \tag{4}$$

For non-linear and spatial effects the penalty matrix $\boldsymbol{K}_k$ is rank deficient and as a result prior (4) is partially improper.

2.1.1 Linear effects

For a vector $\boldsymbol{\gamma}_k = (\gamma_{k_1}, \ldots, \gamma_{k_{J_k}})'$ of J_k linear fixed effects the penalty matrix $\boldsymbol{K}_k$ is an $J_k \times J_k$ identity matrix $\boldsymbol{I}_{J_k}$ reducing (4) to a J_k-variate normal distribution. An alternative is to set $p(\gamma_{k_j} \mid \cdot) \propto const \quad \forall j = 1, \ldots, J_k$. The corresponding design matrix $\boldsymbol{Z}_k$ is a matrix of covariates of order $n \times J_k$, where n denotes the number of observations.

2.1.2 Random effects

In the case of joint models random effects appear in the shared predictor exclusively. Thus let n be the number of individuals and n_i be the number of observations per individual i, so that the total number of observations amounts to $N = \sum_{i=1}^{n} n_i$. Further, let $\boldsymbol{u}_i$ be a vector of observations (or $\mathbf{1}_i$ for random intercepts) of length n_i specific to individual i. Then $\boldsymbol{Z}_k$ is a matrix of order $N \times n$ of the form $\boldsymbol{Z}_k = \mathrm{blockdiag}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ and $\boldsymbol{\gamma}_k$ is a vector of random effects b_i of length n, $\boldsymbol{\gamma}_k = (b_1, \ldots, b_n)'$. The penalty matrix $\boldsymbol{K}_k$ then is an $n \times n$ identity matrix $\boldsymbol{I}_n$.

2.1.3 Non-linear effects

Modelling non-linear effects follows the Bayesian P-spline approach with $\boldsymbol{Z}_k$ being a matrix of B-spline basis functions evaluated at observations $x_i(t)$. Then $\boldsymbol{\gamma}_k$ is a vector of corresponding basis coefficients. The common choice of prior for these basis coefficients is a first or second order random walk. This is achieved by setting the penalty matrix $\boldsymbol{K}_k$ equal to $\boldsymbol{D}'\boldsymbol{D}$, i.e. $\boldsymbol{K}_k = \boldsymbol{D}'\boldsymbol{D}$, where $\boldsymbol{D}$ is a matrix of first or second order differences.

2.1.4 Spatial effects

For spatial effects $\boldsymbol{Z}_k$ is assumed to be an $n \times S$ incidence matrix (potentially also $N \times S$, for spatio-temporal observations) with an entry of 1 if observation $i \ \forall \ i = 1, \ldots, n$ originates from location $s \ \forall \ s = 1, \ldots, S$ with S unique locations and 0 otherwise. The corresponding coefficients $\boldsymbol{\gamma}_k$ follow a Markov random field (MRF) prior achieved via the penalty matrix $\boldsymbol{K}_k$. $\boldsymbol{K}_k$ is an adjacency matrix of order $S \times S$ with entries as the number of neighbours $|n(s)|$ only when locations s and r are neighbours ($s \sim r$) of the form

$$\boldsymbol{K}_k[s, r] = \begin{cases} -1 & \text{if } s \neq r, s \sim r \\ 0 & \text{if } s \neq r, s \nsim r \\ |n(s)| & \text{if } s = r. \end{cases}$$

2.1.5 Interactions

This approach also allows for various interaction terms to be modelled. These include linear interactions $\boldsymbol{x}_1(t) \cdot \boldsymbol{x}_2(t) \cdot \beta$, varying coefficients $f(\boldsymbol{x}_1(t))\boldsymbol{x}_2(t)$, non-linear interactions $f(\boldsymbol{x}_1(t), \boldsymbol{x}_2(t))$ or spatio-temporal models $f(\boldsymbol{s}, \boldsymbol{t})$. The design matrix $\boldsymbol{Z}_k$ is in these cases a row tensor product ($\odot$) of the design matrices of the involved covariates, i.e. $\boldsymbol{Z}_k = \boldsymbol{Z}_1 \odot \boldsymbol{Z}_2$, with dimension $n \times q$ with $q = q_1 q_2$ and q_1 the number of columns in $\boldsymbol{Z}_1$ and q_2 the number of columns in $\boldsymbol{Z}_2$. The corresponding vector of coefficients is then also of length $q = q_1 q_2$, $\boldsymbol{\gamma} = (\gamma_{11}, \ldots, \gamma_{1,q_2}, \ldots, \gamma_{q_1,1}, \ldots, \gamma_{q_1,q_2})'$. The

Table 1 Illustration of data augmentation used for applying Poisson regression

Standard dataset for proportional hazards approach

i	δ_i	T_i	t_i	x_i
1	1	0.85	0	0.83
1	1	0.85	0.3	-0.28
1	1	0.85	0.6	-0.36
2	0	0.58	0	0.09
2	0	0.58	0.4	2.25

→

Augmented dataset for piecewise exponential approach

i	κ_{j-1}	κ_j	o_j	δ_j	x_j
1	0.0	0.30	-1.20	0	0.83
1	0.3	0.40	-2.30	0	-0.28
1	0.4	0.60	-1.61	0	-0.28
1	0.6	0.85	-1.39	1	-0.36
2	0.0	0.30	-1.20	0	0.09
2	0.3	0.40	-2.30	0	0.09
2	0.4	0.60	-1.71	0	2.25

Data augmentation in this toy example was carried out using `pammtools` (Bender and Scheipl 2018)

precision from (4) $\frac{1}{\sigma_k^2}\boldsymbol{K}_k$ then changes to $\frac{1}{\sigma_1^2}(\boldsymbol{K}_1 \otimes \boldsymbol{I}_{q_2}) + \frac{1}{\sigma_2^2}(\boldsymbol{I}_{q_1} \otimes \boldsymbol{K}_2)$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{I}$. an identity matrix of dimension as stated in the subscript. For more information on interaction terms Kneib et al. (2019) give a good overview in general and Köhler et al. (2017) specifically for joint models.

The variance parameters of the coefficient distributions $\sigma^2_{\gamma_k}$ as well as the model variance σ^2_ε will a priori follow inverse gamma distributions, in particular

$$\sigma^2_{\gamma_k} \sim \text{IG}(a, b) \quad \text{and } \sigma^2_\varepsilon \sim \text{IG}(a_0, b_0).$$

2.2 The piecewise exponential representation of the time-to-event submodel

The idea behind a PH model is that an individual's hazard at time t is determined by an individual specific deviation of an underlying baseline hazard $\lambda_0(t)$ at time t. In mathematical notation a generic PH model looks similar to (2) and takes the form

$$\boldsymbol{\lambda}(t) = \lambda_0(t)\exp\{\boldsymbol{\eta}(t)\},$$

where $\boldsymbol{\eta}(t)$ represents an unspecified, time-varying predictor. The aim of estimating such a model then is to quantify the coefficients governing $\boldsymbol{\eta}(t)$ and determining $\lambda_0(t)$ over time t given the times to event $\boldsymbol{T}$ and the events $\boldsymbol{\delta}$. Now this approach can be re-written as an equivalent log-linear Poisson-model. This is achieved by dividing the continuous observation time $t = (0, t_{max}]$ into J intervals and counting the events $\boldsymbol{\delta}_j$ in any given interval j. The intervals are specified by the boundaries $0 = \kappa_0 < \cdots < \kappa_J = t_{max}$ and assuming constant baseline hazards λ_j within each interval the generic PH model changes to a piecewise exponential model of the form

$$\boldsymbol{\lambda}(t) = \lambda_j \exp\{\boldsymbol{\eta}_j\}, \quad \forall\, t \in (\kappa_{j-1}, \kappa_j].$$

Then this formulation is proportional to a Poisson regression of the events $\boldsymbol{\delta}_j$ in intervals $j = 1, \ldots, J$ with expected value $\text{E}(\boldsymbol{\delta}_j)$ in the sense that

$$\boldsymbol{\lambda}(t) = \lambda_j \exp\{\boldsymbol{\eta}_j\} = \frac{\text{E}(\boldsymbol{\delta}_j)}{\exp\{\boldsymbol{o}_j\}}, \quad \text{where}$$
$$\text{E}(\boldsymbol{\delta}_j) = \exp\{\log\lambda_j + \boldsymbol{\eta}_j + \boldsymbol{o}_j\}$$

with transformed exposure times $\boldsymbol{o}_j = (o_{1j}, \ldots, o_{nj})'$ of each individual i in each interval j as offsets ($\exp\{o_{ij}\} = t_{ij}$) (Friedman 1982). This further generalises to a piecewise additive mixed model (PAMM), when the interval-specific log-baseline hazard $\log\lambda_j$ is represented as a smooth function of time $f_0(t_j)$ instead of a step-function and the predictor $\boldsymbol{\eta}_j$ contains (non-)linear, spatial, interaction and/or random effects (Bender et al. 2018).

This form of estimation requires the data to be structured differently than in the conventional way. Table 1 gives an example of this data augmentation and more details can be found in Bender et al. (2018).

2.3 Structured piecewise additive joint models (SPAJM)

Transferring this counting process representation to the context of joint models changes the notation thereof to

$$\boldsymbol{y}(t) = \boldsymbol{\eta}_1(t) + \boldsymbol{\eta}_{\text{ls}}(t) + \varepsilon, \quad \varepsilon \sim \text{N}(0, \sigma^2\boldsymbol{I}) \tag{5}$$
$$\boldsymbol{\lambda}(t) = \exp\left\{f_0(t_j) + \boldsymbol{\eta}_\text{s} + \alpha\boldsymbol{\eta}_{\text{ls}j}\right\}, \quad \forall\, t \in (\kappa_{j-1}, \kappa_j]. \tag{6}$$

The likelihoods then follow the distributions

$$\boldsymbol{y} \mid \boldsymbol{\eta}_1(t), \boldsymbol{\eta}_{\text{ls}}(t) \sim \text{N}(\boldsymbol{\eta}_1(t) + \boldsymbol{\eta}_{\text{ls}}(t), \sigma^2_\varepsilon\boldsymbol{I}) \quad \text{and}$$
$$\boldsymbol{\delta}_j \mid \boldsymbol{\eta}_\text{s}, \boldsymbol{\eta}_{\text{ls}j} \sim \text{Poi}(\exp\left\{f_0(t_j) + \boldsymbol{o}_j + \boldsymbol{\eta}_\text{s} + \alpha\boldsymbol{\eta}_{\text{ls}j}\right\})$$

$\forall\, t \in (\kappa_{j-1}, \kappa_j]$ for the longitudinal and the time-to-event submodel respectively.

2.4 Posterior estimation and implementation

Posterior estimation of this model is accomplished via a Markov Chain Monte Carlo (MCMC) sampler, which in short is a combination of Gibbs-sampling and a Metropolis-Hastings (MH)-algorithm with iteratively weighted least squares (IWLS) proposals. The steps of this sampler are outlined in Algorithm 1 and are implemented in the current version of the statistical software BayesX (Belitz et al. 2022).

Algorithm 1 Posterior estimation of Structured Piecewise Additive Jont Models (Part 1)

Require: $\boldsymbol{\theta}^{[0]} = (\boldsymbol{\theta}_{\mathrm{l}}^{[0]}; \boldsymbol{\theta}_{\mathrm{ls}}^{[0]}; \boldsymbol{\theta}_{\mathrm{s}}^{[0]})'$ with ▷ Initiate starting values
$\boldsymbol{\theta}_{\mathrm{l}}^{[0]} = (\boldsymbol{\gamma}_{\mathrm{l},1}^{[0]}, \ldots, \boldsymbol{\gamma}_{\mathrm{l},p_{\mathrm{l}}}^{[0]}, \sigma_{\varepsilon}^{2[0]}, \sigma_{\gamma_{\mathrm{l},1}}^{2[0]}, \ldots, \sigma_{\gamma_{\mathrm{l},p_{\mathrm{l}}}}^{2[0]})'$,
$\boldsymbol{\theta}_{\mathrm{ls}}^{[0]} = (\boldsymbol{\gamma}_{\mathrm{ls},1}^{[0]}, \ldots, \boldsymbol{\gamma}_{\mathrm{ls},p_{\mathrm{ls}}}^{[0]}, \sigma_{\gamma_{\mathrm{ls},1}}^{2[0]}, \ldots, \sigma_{\gamma_{\mathrm{ls},p_{\mathrm{ls}}}}^{2[0]})'$ and
$\boldsymbol{\theta}_{\mathrm{s}}^{[0]} = (\boldsymbol{\gamma}_{\mathrm{s},1}^{[0]}, \ldots, \boldsymbol{\gamma}_{\mathrm{s},p_{\mathrm{s}}}^{[0]}, \sigma_{\gamma_{\mathrm{s},1}}^{2[0]}, \ldots, \sigma_{\gamma_{\mathrm{s},p_{\mathrm{s}}}}^{2[0]}, \alpha^{[0]}, \sigma_{\alpha}^{2[0]}, \boldsymbol{\gamma}_{t}^{[0]}, \sigma_{\gamma_t}^{2[0]})'$.

for $t = 1$ to T **do**

1. Longitudinal effects ▷ Gibbs-update

for $k = 1$ to p_{l} **do** draw $\boldsymbol{\gamma}_{\mathrm{l},k}^{[t]}$ from $\mathrm{N}\left(\mu_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{*}, \Sigma_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{*}\right)$ with

$$\Sigma_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{*} = \left(\frac{1}{\sigma_{\varepsilon}^{2}} \boldsymbol{Z}_{l,k}' \boldsymbol{Z}_{l,k} + \frac{1}{\sigma_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{2}} \boldsymbol{K}_{\mathrm{l},k}\right)^{-1} \quad \text{and}$$

$$\mu_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{*} = \Sigma_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{*} \left(\frac{1}{\sigma_{\varepsilon}^{2}} \left(\boldsymbol{Z}_{l,k}' (\boldsymbol{y} - \boldsymbol{\eta}_{\mathrm{l,\,-k}} - \boldsymbol{\eta}_{\mathrm{ls}})\right) + \frac{1}{\sigma_{\boldsymbol{\gamma}_{\mathrm{l},k}}^{2}} \boldsymbol{K}_{\mathrm{l},k}\right).$$

In $\mu_{\boldsymbol{\gamma}_{\mathrm{l},k}}$ and $\Sigma_{\boldsymbol{\gamma}_{\mathrm{l},k}}$ use $\sigma_{\varepsilon}^{2[t-1]}$, $\sigma_{\gamma_{\mathrm{l},k}}^{2[t-1]}$, $\boldsymbol{\eta}_{\mathrm{ls}}^{[t-1]}$ and $\boldsymbol{\eta}_{\mathrm{l},-k}^{[t-1]} = \boldsymbol{\eta}_{\mathrm{l}}^{[t-1]} - \boldsymbol{\eta}_{l,k}^{[t-1]}$.
end for

2. Survival effects ▷ IWLS-MH-update

for $k = 1$ to p_{s} **do** determine $\boldsymbol{\gamma}_{\mathrm{s},k}^{[t]}$ as follows:

Draw IWLS proposal $\boldsymbol{\gamma}_{\mathrm{s},k}^{*}$ from $q\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{*} \mid \boldsymbol{\gamma}_{\mathrm{s},k}^{[t-1]}\right) = N\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}_{\mathrm{s},k}}, \boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{s},k}}^{-1}\right)$ with

$$\boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{s},k}} = \boldsymbol{Z}_{\mathrm{s},k}' \boldsymbol{W}_{\mathrm{s}} \boldsymbol{Z}_{\mathrm{s},k} + \frac{1}{\sigma_{\boldsymbol{\gamma}_{\mathrm{s},k}}^{2}} \boldsymbol{K}_{\boldsymbol{\gamma}_{\mathrm{s},k}} \quad \text{and}$$

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}_{\mathrm{s},k}} = \left(\boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{s},k}}\right)^{-1} \boldsymbol{Z}_{\mathrm{s},k}' \boldsymbol{W}_{\mathrm{s}} \left(\tilde{\boldsymbol{y}}_{\mathrm{s}} - \boldsymbol{\eta}_{\mathrm{s},-k}\right).$$

In $\boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{s},k}}$ and $\boldsymbol{\mu}_{\boldsymbol{\gamma}_{\mathrm{s},k}}$ use $\sigma_{\boldsymbol{\gamma}_{\mathrm{s},k}}^{2[t-1]}$, $\boldsymbol{\eta}_{\mathrm{s},-k}^{[t-1]} = \boldsymbol{\eta}_{\mathrm{s}}^{[t-1]} - \boldsymbol{\eta}_{s,k}^{[t-1]}$, working weights $\boldsymbol{W}_{\mathrm{s}}$ and working observations $\tilde{\boldsymbol{y}}_{\mathrm{s}}$. The definition of working weights and observations is given in Appendix A.2.
Accept draw $\boldsymbol{\gamma}_{\mathrm{s},k}^{*}$ with probability

$$\alpha\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{*} \mid \boldsymbol{\gamma}_{\mathrm{s},k}^{[t]}\right) = \min\left\{\frac{L\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{*}\right) p\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{*}\right) q\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{[t]} \mid \boldsymbol{\gamma}_{\mathrm{s},k}^{*}\right)}{L\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{[t]}\right) p\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{[t]}\right) q\left(\boldsymbol{\gamma}_{\mathrm{s},k}^{*} \mid \boldsymbol{\gamma}_{\mathrm{s},k}^{[t]}\right)}, 1\right\}$$

with likelihood $L(\boldsymbol{\gamma}_{\mathrm{s},k}) = p(\boldsymbol{\delta} \mid \boldsymbol{\gamma}_{\mathrm{s},k}, \cdot)$.
end for

3. Shared effects ▷ IWLS-MH-update

for $k = 1$ to p_{ls} **do** determine $\boldsymbol{\gamma}_{\mathrm{ls},k}^{[t]}$ as follows:

Draw IWLS proposal $\boldsymbol{\gamma}_{\mathrm{ls},k}^{*}$ from $q\left(\boldsymbol{\gamma}_{\mathrm{ls},k}^{*} \mid \boldsymbol{\gamma}_{\mathrm{ls},k}^{[t-1]}\right) = N\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}_{\mathrm{ls},k}}, \boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{ls},k}}^{-1}\right)$ with

$$\boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{ls},k}} = \boldsymbol{Z}_{\mathrm{ls},k}' \boldsymbol{W}_{\mathrm{ls}} \boldsymbol{Z}_{\mathrm{ls},k} + \frac{1}{\sigma_{\boldsymbol{\gamma}_{\mathrm{ls},k}}^{2}} \boldsymbol{K}_{\boldsymbol{\gamma}_{\mathrm{ls},k}} \quad \text{and}$$

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}_{\mathrm{ls},k}} = \left(\boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{ls},k}}\right)^{-1} \boldsymbol{Z}_{\mathrm{ls},k}' \boldsymbol{W}_{\mathrm{ls}} \left(\tilde{\boldsymbol{y}}_{\mathrm{ls}} - \boldsymbol{\eta}_{\mathrm{ls},-k}\right).$$

In $\boldsymbol{P}_{\boldsymbol{\gamma}_{\mathrm{ls},k}}$ and $\boldsymbol{\mu}_{\boldsymbol{\gamma}_{\mathrm{ls},k}}$ use $\sigma_{\boldsymbol{\gamma}_{\mathrm{ls},k}}^{2[t-1]}$, $\boldsymbol{\eta}_{\mathrm{ls},-k}^{[t-1]} = \boldsymbol{\eta}_{\mathrm{ls}}^{[t-1]} - \boldsymbol{\eta}_{s,k}^{[t-1]}$, working weights $\boldsymbol{W}_{\mathrm{ls}}$ and working observations $\tilde{\boldsymbol{y}}_{\mathrm{ls}}$. The definition of working weights and observations is given in Appendix A.3.

Algorithm 1 continued Posterior estimation of Structured Piecewise Additive Jont Models (Part 2)

Accept draw $\boldsymbol{\gamma}^{*}_{\text{ls},k}$ with probability

$$\alpha\left(\boldsymbol{\gamma}^{*}_{\text{ls},k} \mid \boldsymbol{\gamma}^{[t]}_{\text{ls},k}\right) = \min\left\{\frac{L\left(\boldsymbol{\gamma}^{*}_{\text{ls},k}\right)\, p\left(\boldsymbol{\gamma}^{*}_{\text{ls},k}\right)\, q\left(\boldsymbol{\gamma}^{[t]}_{\text{ls},k} \mid \boldsymbol{\gamma}^{*}_{\text{ls},k}\right)}{L\left(\boldsymbol{\gamma}^{[t]}_{\text{ls},k}\right)\, p\left(\boldsymbol{\gamma}^{[t]}_{\text{ls},k}\right)\, q\left(\boldsymbol{\gamma}^{*}_{\text{ls},k} \mid \boldsymbol{\gamma}^{[t]}_{\text{ls},k}\right)}, 1\right\}$$

with likelihood $L(\boldsymbol{\gamma}_{\text{ls},k}) = p(\boldsymbol{y} \mid \boldsymbol{\gamma}_{\text{ls},k}, \cdot)\; p(\boldsymbol{\delta} \mid \boldsymbol{\gamma}_{\text{ls},k}, \cdot)$.
end for

4. Update variance parameters ▷ Gibbs-update

4.1 Model variance
Let $N = \sum_{i=1}^{n} n_i$ be the total number of longitudinal observations as the sum of all observations n_i per individual i across all individuals n.
Draw $\sigma_{\varepsilon}^{2[t]}$ from $\text{IG}(\tilde{a}_0, \tilde{b}_0)$ with

$$\tilde{a}_0 = a_0 + \frac{N}{2} \quad \text{and} \quad \tilde{b}_0 = b_0 + (\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}})'(\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}}).$$

In $\tilde{a}_0$ and $\tilde{b}_0$ use $\boldsymbol{\eta}_{\text{l}}^{[t]}$ and $\boldsymbol{\eta}_{\text{ls}}^{[t]}$.

4.2 Effect variance
for F **do**or $k_{\cdot} = 1$ to $p_{\cdot}$ draw $\sigma_{\boldsymbol{\gamma}_{k\cdot}}^{2[t]}$ from $\text{IG}(\tilde{a}_{k\cdot}, \tilde{b}_{k\cdot})$ with

$$\tilde{a}_{k\cdot} = a_{k\cdot} + \text{rk}(\boldsymbol{K}_{k\cdot}) \quad \text{and} \quad \tilde{b}_{k\cdot} = b_{k\cdot} + \frac{1}{2}\boldsymbol{\gamma}'_{\cdot,k\cdot}\boldsymbol{K}_{k\cdot}\boldsymbol{\gamma}_{\cdot,k\cdot}.$$

In $\tilde{a}_{k\cdot}$ and $\tilde{b}_{k\cdot}$ use $\boldsymbol{\gamma}_{\cdot,k\cdot}^{[t]}$.
end for

end for

3 Simulation study

With the following simulation study we want to (a) illustrate the flexibility of the SPAJM with regard to effect specification, (b) highlight its capability for estimating spatial effects and (c) confirm its computational advantage by comparing the performance of our approach to an already existing one. In order to meet intention (a) the simulated model will be maximally generic, i.e. include various types of effects in all possible predictors alongside a non-linear baseline hazard, and to meet intention (b) the model will comprise a spatial effect. Since the spatial effect can only be located in one of the predictors for identifiability reasons we will look at three settings to determine whether the quality of performance is location specific:

Setting 1 The spatial effect is located in the shared predictor η_{ls},
Setting 2 It is located in the survival predictor η_{s} and
Setting 3 In the longitudinal predictor η_{l}.

Lastly, to ascertain intention (c) the runtimes of our approach will be contrasted to an already existing one.

In terms of software we will use the BayesX implementation of the SPAJM and benchmark it against the similarly flexible joint model implementation of the tensor-product approach using Newton-Raphson procedures and derivative-based Metropolis-Hastings sampling by Köhler et al. (2018) in the `R` package `bamlss` (Umlauf et al. 2021). We will use the current developer version of BayesX (Belitz te al. 2022) as well as `bamlss` version 1.1-8 on `R`-4.1.2 (R Core Team 2022).

3.1 Setup

The simulation mimics a study setup of duration (0, 1) with a fixed number of planned visits per individual, of which not all have been observed due to an event or censoring/ drop-outs. The covariates in the data were assumed to be continuous with some being time-constant and some time-varying. The simulated effects underlying the data and determining the longitudinal outcome $\boldsymbol{y}$ as well as the event outcome $(\boldsymbol{T}, \boldsymbol{\delta})$ have been chosen to represent a maximally generic joint model. Therefore, each predictor $\boldsymbol{\eta}_{\cdot}$ consists of at least one linear and one non-linear effect. The shared predictor features the random intercepts and slopes, while the spatial effect is rotated through each predictor according to the above presented rationale. We generate longitudinal measurements $\boldsymbol{y}(t)$ for $n = 200$ individuals over $n_i = 6$ individual specific, original time points each in the range of $t \in (0, 1)$ according to the generic model given in (5) and (6) with $\alpha = -0.3$ and the following predictors

$$\begin{aligned}\boldsymbol{\eta}_{\text{l}} &= 0.5\ \boldsymbol{x}_{\text{l1}} + f_1(\boldsymbol{x}_{\text{l2}}),\\ \boldsymbol{\eta}_{\text{ls}} &= 0.9\ \boldsymbol{x}_{\text{ls1}} - 0.5\ f_2(\boldsymbol{x}_{\text{ls2}}) - 0.5\ \boldsymbol{x}_{\text{l3}}(t)\\ &\quad + 0.4\ \boldsymbol{t} + \boldsymbol{b}_0 + \boldsymbol{b}_1\ \boldsymbol{t} \quad \text{and}\\ \boldsymbol{\eta}_{\text{s}} &= 0.1\ \boldsymbol{x}_{\text{s1}} + 0.5\ f_2(\boldsymbol{x}_{\text{s2}})\end{aligned}$$

with the non-linear functions $f_1(x) = 0.5\ x + 15\ \phi(2(x - 0.2)) - \phi(x + 0.4)$ and $f_2(x) = \sin(x)$. The individual specific time points were determined as six general, equidistant time points with a randomly sampled deviation for each individual. The procedure mimics one visit per individual during the course of a year (random element) over n_i years (general element) with subsequent standardisation to limit the range to $t \in (0, 1)$. All covariates $\boldsymbol{x}_{\text{ls}\cdot}$ and $\boldsymbol{x}_{\text{s}\cdot}$ are simulated as time constant with the exception of $\boldsymbol{x}_{\text{ls3}}$, which is simulated time dependent just like covariates $\boldsymbol{x}_{\text{l}\cdot}$, with all $\boldsymbol{x}_{\cdot\cdot} \sim U(-1, 1)$. Further the model variance is set to $\sigma_\varepsilon^2 = 0.5$ and the variances of the random intercepts and slopes are set to $\sigma_{b_0}^2 = \sigma_{b_1}^2 = 2$.

True survival times T_i^* are determined based on a Weibull baseline hazard function $\lambda_0(t) = pqt^{q-1}$ with scale $p = 0.4$ and shape $q = 1.5$. The event times are then set to $T_i = \min(T_i^*, 1)$ with event indicator $\delta_i = 1$ if $T_i^* \leq 1$ and $\delta_i = 0$ otherwise. All censored individuals, i.e. those with event indicator $\delta_i = 0$, thus receive an event time of $T_i = 1$. However, to achieve a more realistic censoring scenario, we apply in addition uniform censoring to 50% of the censored individuals by randomly sampling u_i from $U(0, 1)$ and setting the censoring time of those individuals to $T_i = \min(u_i, 1)$.

The spatial effect is based on the map of counties in western Germany available from the `R` package `BayesX` and calculated as $f_{geo} = \sin(\boldsymbol{c}_x) \cdot \cos(0.5\ \boldsymbol{c}_y)$ with $\boldsymbol{c}_x$ and $\boldsymbol{c}_y$ being the scaled x- and y-coordinates respectively of the centroids of each region. The regions are then randomly distributed across the individuals.

For each setting we use $R = 100$ replications. Convergence is achieved in BayesX by using 70,000 iterations per run with a burn-in of 10,000 and a thinning factor of 60 and in `bamlss` by using 44,000 iterations (54,000 with f_{geo} in η_{l}) with a burn-in of 4000 and a thinning factor of 40 (50). In order to compare the results of both implementations we calculate the mean squared error (MSE), bias and coverage of the 95%-high density interval (HDI) of the posterior distribution of each parameter and compare runtimes between BayesX and `bamlss`.

3.2 Results

The outcome of the estimation performance of the simulation study can be found in Fig. 1 and the computational performance is illustrated in Fig. 2. The summarized results in Fig. 1 already make it clear that a joint model with a piecewise additive formulation of the survival submodel is is equal in terms of effect estimation to its established PH counterpart given the small MSE and bias values as well as the high coverage rates. Detailed results of the individual effects can be found in the Appendix in Fig. 4, which confirm this high level impression. Both methods exhibit the largest deviation from the true data in the shared predictors η_{ls}. In terms of estimation any effect in this predictor belongs to the most demanding to estimate, as the corresponding likelihood features both model parts. Thus the larger bias here is to be expected. Furthermore, it quickly becomes clear that BayesX outperforms `bamlss` in the estimation results of the shared predictors η_{ls} and the survival predictors η_{s}. The reason for the performance of `bamlss` in the shared predictors η_{ls} is due to the random effects, which can be seen from the more detailed Fig. 4 in the Appendix. Their estimates remain rather small, which is why their high density intervals do not cover the simulated (true) random effects $\boldsymbol{b}_0$ and $\boldsymbol{b}_1$, which in turn affects the overall results for the shared predictor η_{ls}. Similarly the survival predictors η_{s} perform rather weakly with `bamlss`, which is mainly due to the rather large bias in the association α (see Fig. 4). The estimation procedure implemented in `bamlss` is in fact tailored to identify advanced association structures in joint models, which is why the bias in α is highly likely a result of the underestimation of the random effects. Only in the estimation of the longitudinal predictor η_{l} did `bamlss` surpass BayesX, which is interesting, since the formulation implemented in `bamlss` does not extend to longitudinal-only-predictors. It assumes η_{l} to be a part of η_{ls}, but since the data of η_{l} is simulated such that it is not associated with the survival part of the model, the results of η_{l} under `bamlss` are more precise than those of η_{ls}. Figure 1 further demonstrates the capability of both methods to estimate spatial effects, while it also shows the indifference to the position (in η_{ls}, η_{s} or η_{l}) of the spatial effect within the model. First of all, the figure indicates a stable performance of the spatial effect f_{geo} in both implementations independent of the predictor it belongs to. Secondly, also the other predictors remain very stable in their performance regardless of the simulation setting. If the position of the spatial effect f_{geo} mattered, it would not just show in the estimation accuracy of the effect itself, but it would also affect the effects in other parts of the model, which is not the case here. Again the reason for the `bamlss` results are similar to before. The estimation results for f_{geo} exhibit the same behaviour as for the random effects: They remain surprisingly small, resulting in a larger bias and thus only achieving a rather low coverage.

Lastly, in terms of computational cost the piecewise additive approach in BayesX has an advantage over the PH approach in `bamlss` with lower runtimes (see Fig. 2). With both methods Setting 1 with the spatial effect f_{geo} in the shared predictor η_{ls} is the most time consuming. But this is also the most complex setting in terms of estimation, therefore, the increased runtime is not surprising. Setting 2 with

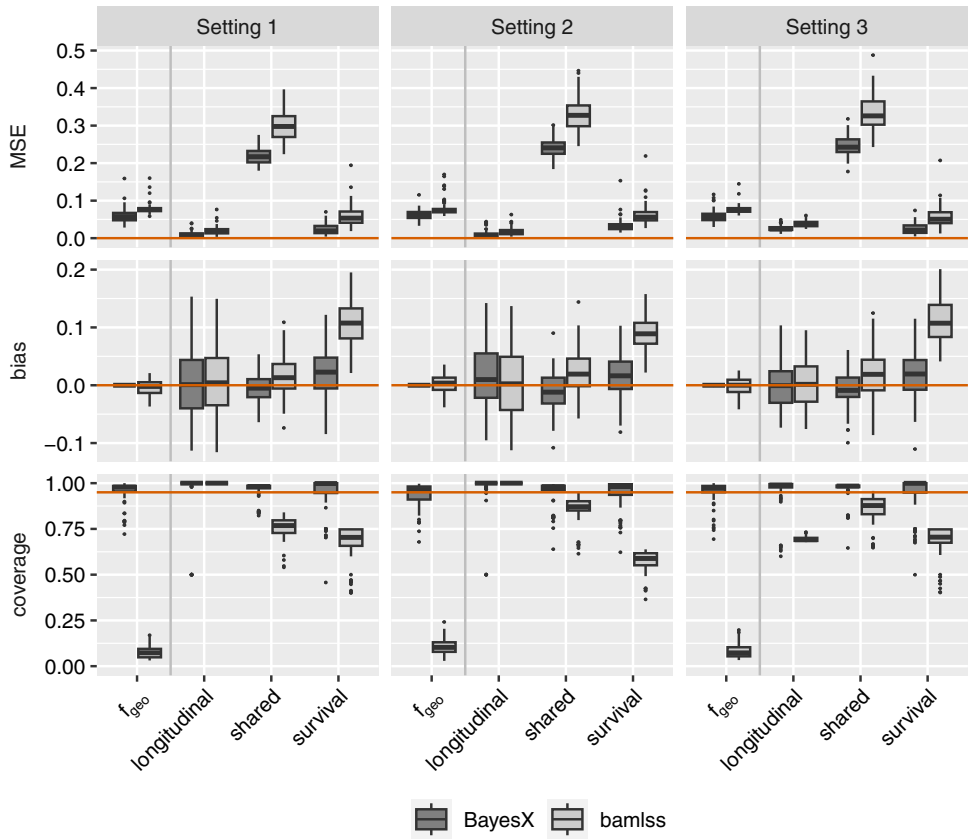

Fig. 1 Boxplots of mean squared error (MSE), bias and 95%-coverage for the extcolorredspatial effect f_{geo} as well as the predictors per method and simulation setting (Setting 1—f_{geo} in η_{ls}, Setting 2—f_{geo} in η_s, Setting 3—f_{geo} in η_l). The orange horizontal line marks the reference value of each statistic

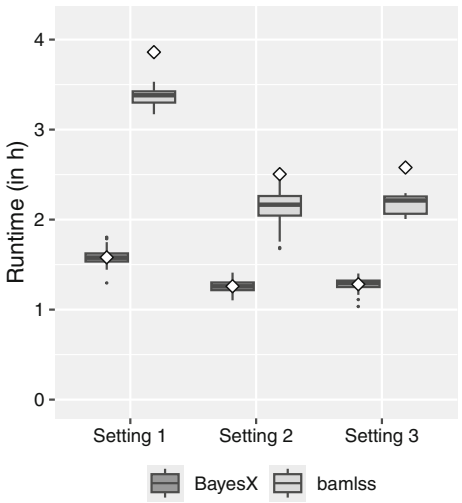

Fig. 2 Boxplots of runtimes in hours per method and simulation setting. For readability reasons plots are clipped, therefore extreme outliers are excluded. Diamonds represent respective means. BayesX has visibly lower runtimes than `bamlss`. With both methods Setting 1 (f_{geo} in η_{ls}) is the most time consuming, while Setting 2 (f_{geo} in η_s) and 3 (f_{geo} in η_l) are faster and take equally long

f_{geo} in the survival predictor η_s and Setting 3 with f_{geo} in the longitudinal predictor η_l are less complex from an estimation perspective, which is also evident in the short runtimes. More detailed descriptive statistics on the runtimes can be found in the Appendix in Table 3.

4 Physical functioning after a caesura

"Caesurae" describe certain physiological events that act like a turning point in the course of an individuals health, among them heart attacks, strokes or diagnoses of cancer. In 2015 the World Health Organsiation (WHO) concluded in their "World Report on Ageing and Health" that the physical capacity dimension of "Healthy Ageing" still suffers from a lack of understanding. Physical capacity can be measured as functional health (aka physical functioning), which decreases naturally over time until death. A caesura, however, has the power to alter the trajectory of an individual's functional health both in a negative and positive way (WHO, 2015). While the longitudinal modelling of these trajectories is already of interest, the trajectories themselves influence an individual's survival time. Therefore, a joint model is appropriate to capture both these aspects of the data.

To examine the development of physical functioning after a caesura in Germany we will resort to the German Ageing Survey (DEAS), which aims at studying the second half of life with people between 40 and 85 years old and living in Germany being eligible for study participation. The DEAS has collected information on physical functioning from a SF-36 survey, health conditions qualifying as caesurae, terminal dates and a multitude of other variables, which might help explain the development of physical functioning after a

Table 2 BayesX estimates of linear effects of physical functioning after a caesura

	Posterior mean	95%-HDI
β_0	0.848	[0.827, 0.871]
Gender	−0.091	[− 0.123, − 0.059]
t	−0.247	[− 0.277, − 0.219]
α	−3.381	[− 4.329, − 2.432]

caesura, over the course of seven waves (1996, 2002, 2008, 2011, 2014, 2017, 2021) (Engstler et al. 2014; Klaus and Engstler 2017).

Our analysis will focus on data from waves 2008 to 2021 with originally 6622 participants, of which 750 suffered from a heart attack or stroke i.e. a cardiovascular caesura, during their panel participation. Single observations, cases with missing data and caesurae with onset prior to the participant's entry into the panel were excluded from the analysis. For the remaining 636 the time of onset of the caesura was set to coincide with the interview date, in which the caesura was first reported, since the exact onset date of the caesurae is not collected. Out of 636 participants 79 (12.4%) died.

As explanatory variables for the trajectory of functional health (`sf36`) we consider time ($\boldsymbol{t}$), gender (`gender`), the age of onset (`aoo`) of the caesura as well as living location of the participant on the level of European Nomenclature of Territorial Units for Statistics (NUTS) 2. In order to avoid re-identification of participants few regions had to be combined leaving now 33 regions of the original 36. The continuous and strictly positive variables SF-36 `sf36`, age of onset `aoo` and time $\boldsymbol{t}$ are scaled to the domain (0, 1). We then consider the model

$$\begin{aligned}
\texttt{sf36}_i(t) &= \beta_0 + \eta_{\text{ls}i} + \varepsilon_i(t)\\
\lambda(t)x &= \exp\{f_0(t) + \alpha\eta_{\text{ls}i}\}\\
\eta_{\text{ls}i} &= \beta_1\texttt{gender}_i + f(\texttt{aoo}_i) + f_{geo}(\texttt{NUTS2}_i)\\
&\quad + \beta_t\boldsymbol{t} + b_{0i} + b_{1i}\boldsymbol{t}
\end{aligned}$$

and estimate it with `BayesX` and `bamlss`.

With 79 events out of 636 individuals the survival data is unbalanced and presents a situation that would already prove difficult to estimate in a standard survival analysis setting. We only present BayesX results here because `bamlss` proved to react sensitive to this imbalance. The reason for this lies in the combination of a lack of information due to few events and the likelihood of a PH model involving an integral, which can only be approximated numerically, whereas the piece-wise additive approach relies on discretizing time and thus avoiding this integral. More detail on this can be found in Appendix D. The linear effects can be found in Table 2 and the non-linear effects in Fig. 3.

None of the linear effects includes zero in their HDI, thus they are significantly different from zero. The association α of both model parts is negative meaning that a lower level of the modelled trajectory of physical functioning `sf36` translates to a higher probability of experiencing an event.

The intercept can be interpreted as a male individual at scaled age of 0.402 (i.e. an unscaled age of 40.2) years old at onset of the caesura can be expected to have an average scaled SF-36 level `sf36` of 0.848 [0.827, 0.871].

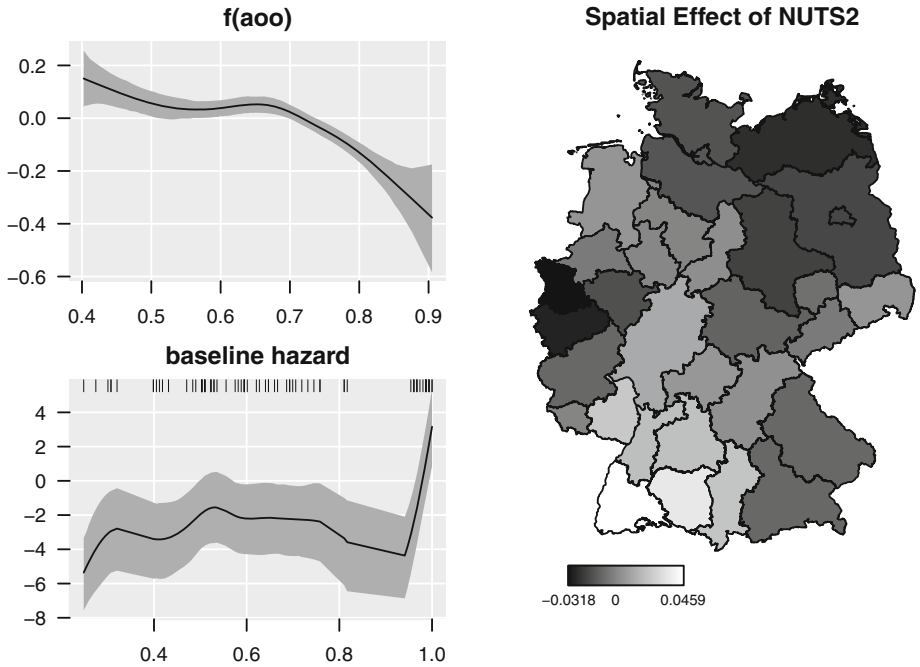

Fig. 3 BayesX estimates of the smooth effect of the age of onset `aoo` and the geographical location on physical functioning as well as the estimated baseline hazard of the model

For women this reduces on average by − 0.091 [− 0.123, − 0.059]. Every scaled month after the caesura further reduces the level of `sf36` by − 0.247 [− 0.277, − 0.219]. The age of onset `aoo` has in general a decreasing effect on `sf36` (upper left panel Fig. 3). Though it needs to be pointed out that before the scaled age of 0.55 (55 years old) the effect is positive, i.e. it increases the level of `sf36` thus slowing down the natural decline of physical functioning, while for an `aoo` between roughly 0.55 and 0.7 (55–70 years old) the effect is constant around zero, i.e. it is negligible, and a caesura after an `aoo` of 0.7 (70 years old) has a negative effect on `sf36` translating to an accelerated decline of physical functioning. In terms of living location there is a South-West against North-East (and Mid-West) divide (right panel Fig. 3). People in the North-Eastern part of Germany especially in the area of Mecklenburg-Pommerania, Brandenburg and Saxony-Anhalt as well as those of the Western parts in the Dusseldorf and Cologne regions see a negative effect on their level of `sf36`. Those living in the South-Western part especially in South-West Baden-Wurttemberg (Black Forest region) see an increasing effect on their `sf36` level. Given that the association is negative this means that the probability for an event is decreased most for people from the South West of Germany and increased most for those in the North-East and Mid-West. What these two areas have in common is that they comprise the most and least densely populated areas in Germany. This might be a starting point for further research to investigate what exactly triggers the effect to take this particular shape, since the living location in this example can be interpreted as a proxy for other variables that have not been included in the model.

The baseline hazard is almost linear over time (lower left panel Fig. 3), thus the risk of experiencing an event is roughly the same at all times throughout the study.

5 Conclusion and discussion

The focus of this article has been on proposing a piecewise additive joint model for longitudinal and time-to-event data allowing for spatial, (non-)-linear and random effects to be included as well as estimation of the baseline hazard without any assumptions about its distributional form. In a simulation study comprising (non-)linear as well as a spatial effect it became evident that the piecewise additive approach yields results similar or better to the equally flexibly `bamlss`-methodology for joint models in `R` and that this performance is high independent of the position of the spatial effect. This method was illustrated by an example of the development of physical functioning after a caesura in people in their second half of life.

The concept of piecewise additive joint models has not just proven its accuracy in estimating complex effects, but also its ability in handling unbalanced data in terms of availability of event observations.

Applying the piecewise additive approach requires augmenting data, which is part of the time-to-event process. This augmentation artificially increases the size of the dataset and when the original data is large, it can lead to longer runtimes. In our experience this is, however, seldom the case. Furthermore, this method could also be combined with other models in the longitudinal part of the model such as quantile regression, a location-scale model or multiple longitudinal outcomes in a multivariate joint model. Also, Bayesian variable or effect selection in this type of joint model could be investigated since very few methods for variable selection in joint models exist yet.

Acknowledgements Elisabeth Bergherr gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant WA 4249/2-1.

Author Contributions All authors contributed to the study conception and design. Data preparation and analysis were performed by AR under the supervision of EB and TK. SL and TK conceptualized and implemented the software. The first draft of the manuscript was written by AR and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Elisabeth Bergherr gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant WA 4249/2-1.

Data availabilty Data of the Germany Ageing Survey can be obtained after application from https://www.dza.de/forschung/fdz/deutscher-alterssurvey/deas-datennutzung.

Code Availability Samples of the code used for the simulation study can be found on https://github.com/arappl/SPAJM.git.

Declarations

Conflict of interest Other than the above funding the authors have no relevant financial or non-financial interests to disclose.

Ethics approval Ethics approval does not apply to this article.

A Derivation of the full conditional and IWLS-proposal distributions used for posterior estimation

A.1 Longitudinal effects

Let $\boldsymbol{\gamma}_{1,k}$ be the coefficients of one of $k = 1, \ldots, p_1$ effects in the longitudinal predictor with a prior as given in (4) and let further denote $\boldsymbol{\eta}_{1,\text{-k}} = \boldsymbol{\eta}_1 - \boldsymbol{\eta}_{1,\text{k}}$, i.e. the longitudinal predictor without the k^{th} element. Then the derivation of the full conditionals for this effect follows as:

$$
\begin{aligned}
p(\boldsymbol{\gamma}_{1,k} \mid \cdot) &\propto p(\boldsymbol{\gamma}_1 \mid \sigma^2_\varepsilon, \sigma^2_{\boldsymbol{\gamma}_1})\ p(\boldsymbol{y} \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_{\text{ls}}, \cdot) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2_{\boldsymbol{\gamma}_1}} \boldsymbol{\gamma}'_{1,k} \boldsymbol{K}_{1,k} \boldsymbol{\gamma}_{1,k}\right\} \\
&\quad \exp\Bigg\{-\frac{1}{2\sigma^2_\varepsilon}(\boldsymbol{Z}_{1,k}\boldsymbol{\gamma}_{1,k} - (\boldsymbol{y} - \boldsymbol{\eta}_{1,-k} - \boldsymbol{\eta}_{\text{ls}}))' \\
&\qquad (\boldsymbol{Z}_{1,k}\boldsymbol{\gamma}_{1,k} - (\boldsymbol{y} - \boldsymbol{\eta}_{1,-k} - \boldsymbol{\eta}_{\text{ls}}))\Bigg\} \\
\boldsymbol{\gamma}_{1,k} \mid \cdot &\sim \text{N}(\mu^*_{\boldsymbol{\gamma}_{1,k}}, \Sigma^*_{\boldsymbol{\gamma}_{1,k}}) \\
\Sigma^*_{\boldsymbol{\gamma}_{1,k}} &= \left(\frac{1}{\sigma^2_\varepsilon}\boldsymbol{Z}'_{l,k}\boldsymbol{Z}_{l,k} + \frac{1}{\sigma^2_{\boldsymbol{\gamma}_{1,k}}}\boldsymbol{K}_{1,k}\right)^{-1} \\
\mu^*_{\boldsymbol{\gamma}_{1,k}} &= \Sigma^*_{\boldsymbol{\gamma}_{1,k}} \\
&\quad \left(\frac{1}{\sigma^2_\varepsilon}\left(\boldsymbol{Z}'_{l,k}(\boldsymbol{y} - \boldsymbol{\eta}_{1,\text{-k}} - \boldsymbol{\eta}_{\text{ls}})\right) + \frac{1}{\sigma^2_{\boldsymbol{\gamma}_{1,k}}}\boldsymbol{K}_{1,k}\right)
\end{aligned}
$$

A.2 Survival effects

Since the full conditional distribution of the k^{th} survival specific coefficients $p(\boldsymbol{\gamma}_{\text{s},k} \mid \boldsymbol{\delta}, \sigma^2_{\boldsymbol{\gamma}_{\text{s},k}}, \cdot)$ out of $k = 1, \ldots, p_\text{s}$ survival specific effects are analytically intractable, we use MH-steps with IWLS proposals, which approximate the true log-full conditionals. Consider the (standard) full conditional

$$
p\left(\boldsymbol{\gamma}_{\text{s},k} \mid \boldsymbol{\delta}, \sigma^2_{\boldsymbol{\gamma}_{\text{s},k}}, \cdot\right) \propto p(\boldsymbol{\gamma}_{\text{s},k} \mid \sigma^2_{\boldsymbol{\gamma}_{\text{s},k}})\ p\left(\boldsymbol{\delta} \mid \boldsymbol{\eta}_\text{s}, \boldsymbol{\eta}_{\text{ls}}, \cdot\right).
$$

Let $\boldsymbol{Z}_{\text{s},k}$ be the corresponding design matrix of effect $\boldsymbol{\gamma}_{\text{s},k}$ and $\frac{1}{\sigma^2_{\boldsymbol{\gamma}_{\text{s},k}}}\boldsymbol{K}$ the variance statement of prior $p(\boldsymbol{\gamma}_{\text{s},k} \mid \sigma^2_{\boldsymbol{\gamma}_{\text{s},k}})$ (compare prior given in 4). Then draw IWLS proposal $\boldsymbol{\gamma}^*_{\text{s},k}$ from a normal distribution density $q(\boldsymbol{\gamma}^*_{\text{s},k} \mid \boldsymbol{\gamma}^{[t]}_{\text{s},k})$ with $\boldsymbol{\gamma}^{[t]}_{\text{s},k}$ being the value of $\boldsymbol{\gamma}_{\text{s},k}$ at iteration t of the MCMC algorithm. More specifically

$$
\boldsymbol{\gamma}^*_{\text{s},k} \sim N\left(\boldsymbol{\mu}^{[t]}_{\boldsymbol{\gamma}_{\text{s},k}}, \boldsymbol{P}^{[t]-1}_{\boldsymbol{\gamma}_{\text{s},k}}\right)
$$

$$
\text{with } \boldsymbol{P}^{[t]}_{\boldsymbol{\gamma}_{\text{s},k}} = \boldsymbol{Z}'_{\text{s},k}\boldsymbol{W}^{[t]}_\text{s}\boldsymbol{Z}_{\text{s},k} + \frac{1}{\sigma^{2[t]}_{\boldsymbol{\gamma}_{\text{s},k}}}\boldsymbol{K}_{\boldsymbol{\gamma}_{\text{s},k}}
$$

$$
\text{and } \boldsymbol{\mu}^{[t]}_{\boldsymbol{\gamma}_{\text{s},k}} = \left(\boldsymbol{P}^{[t]}_{\boldsymbol{\gamma}_{\text{s},k}}\right)^{-1}\boldsymbol{Z}'_{\text{s},k}\boldsymbol{W}^{[t]}_\text{s}\left(\tilde{\boldsymbol{y}}^{[t]}_\text{s} - \boldsymbol{\eta}^{[t]}_{\text{s},-k}\right).
$$

Here $\boldsymbol{\eta}^{[t]}_{\text{s},-k} = \boldsymbol{\eta}^{[t]}_\text{s} - \boldsymbol{\eta}^{[t]}_{\text{s},k}$, and $\boldsymbol{W}^{[t]}_\text{s}$ denotes the working weights and $\tilde{\boldsymbol{y}}^{[t]}_\text{s}$ the working observations all evaluated at the current state t of the MCMC chain. The definition of working weights and observations is given further below.

The acceptance probability of the IWLS proposal $\boldsymbol{\gamma}^*$ is then

$$
\alpha\left(\boldsymbol{\gamma}^*_{\text{s},k} \mid \boldsymbol{\gamma}^{[t]}_{\text{s},k}\right) = \min\left\{\frac{L\left(\boldsymbol{\gamma}^*_{\text{s},k}\right)\ p\left(\boldsymbol{\gamma}^*_{\text{s},k}\right)\ q\left(\boldsymbol{\gamma}^{[t]}_{\text{s},k} \mid \boldsymbol{\gamma}^*_{\text{s},k}\right)}{L\left(\boldsymbol{\gamma}^{[t]}_{\text{s},k}\right)\ p\left(\boldsymbol{\gamma}^{[t]}_{\text{s},k}\right)\ q\left(\boldsymbol{\gamma}^*_{\text{s},k} \mid \boldsymbol{\gamma}^{[t]}_{\text{s},k}\right)}, 1\right\}
$$

with $L(\boldsymbol{\gamma}_{\text{s},k}) = p(\boldsymbol{\delta} \mid \boldsymbol{\gamma}_{\text{s},k}, \cdot)$ being the likelihood evaluated at the proposal $\boldsymbol{\gamma}^*_{\text{s},k}$ as well as the current state $\boldsymbol{\gamma}^{[t]}$ of the effect.

If the proposal then is accepted it becomes the new state $\boldsymbol{\gamma}^{[t+1]}_{\text{s},k} = \boldsymbol{\gamma}^*_{\text{s},k}$, otherwise the current state remains $\boldsymbol{\gamma}^{[t+1]}_{\text{s},k} = \boldsymbol{\gamma}^{[t]}_{\text{s},k}$. Acceptance is established via random draws from a uniform distribution following the logic:

1. Draw $u \sim \text{Unif}(0, 1)$
2. If $u \leq \alpha$
   then $\boldsymbol{\gamma}^{[t+1]} = \boldsymbol{\gamma}^*$
   else $\boldsymbol{\gamma}^{[t+1]} = \boldsymbol{\gamma}^{[t]}$.

For the definition of the working weights and observations consider the log-full conditional

$$
\log(p(\boldsymbol{\gamma}_{\text{s},k} \mid \boldsymbol{\delta}, \sigma^2_{\boldsymbol{\gamma}_{\text{s},k}}, \cdot)) \propto -\frac{1}{2\sigma^2_{\boldsymbol{\gamma}_{\text{s},k}}}\boldsymbol{\gamma}'_{\text{s},k}\boldsymbol{K}_{\boldsymbol{\gamma}_{\text{s},k}}\boldsymbol{\gamma}_{\text{s},k} + \ell(\boldsymbol{\eta}_\text{s}),
$$

where $\ell(\boldsymbol{\eta}_\text{s})$ denotes the log-likelihood depending on predictor $\boldsymbol{\eta}_\text{s} = \sum_{k=1}^{p_\text{s}} \boldsymbol{Z}_{\text{s},k}\boldsymbol{\gamma}_{\text{s},k}$ (compare 3) thus including $\boldsymbol{\gamma}_{\text{s},k}$.

Further, define the score vector $\boldsymbol{v}_\text{s}$ as

$$
\boldsymbol{v}^{[t]}_\text{s} = \frac{\partial \ell(\boldsymbol{\eta}^{[t]}_\text{s})}{\partial \boldsymbol{\eta}^{[t]}_\text{s}},
$$

i.e. the vector of first derivatives of $\ell(\boldsymbol{\eta}_\text{s})$ with respect to the predictor $\boldsymbol{\eta}_\text{s}$ evaluated at the current iteration t and the working weights also evaluated at iteration t as

$$
\boldsymbol{W}^{[t]}_\text{s} = \text{diag}\left(w_1\left(\eta^{[t]}_{\text{s},1}\right), \ldots, w_{n_a}\left(\eta^{[t]}_{\text{s},n_a}\right)\right)
$$

with n_a the number of observations in the augmented dataset (compare Sect. 2.2 Table 1) and

$$
w_i\left(\eta^{[t]}_{\text{s},i}\right) = -\text{E}\left(\frac{\partial \ell(\eta^{[t]}_{\text{s},i})}{\partial^2 \eta^{[t]}_{\text{s},i}}\right) = -\text{E}\left(\frac{\partial v^{[t]}_{\text{s},i}}{\partial \eta^{[t]}_{\text{s},i}}\right).
$$

The vector of working observations $\tilde{\boldsymbol{y}}^{[t]}_\text{s} = \left(\tilde{y}_{\text{s}1}\left(\eta^{[t]}_1\right), \ldots, \tilde{y}_{\text{s}n_a}\left(\eta^{[t]}_{n_a}\right)\right)'$ is then determined by

$$
\tilde{\boldsymbol{y}}^{[t]}_\text{s} = \boldsymbol{\eta}^{[t]}_\text{s} + \left(\boldsymbol{W}^{[t]}_\text{s}\right)^{-1}\boldsymbol{v}^{[t]}_\text{s}.
$$

A.3 Shared effects

The coefficients $p(\boldsymbol{\gamma}_{\text{ls},k} \mid \boldsymbol{y}, \boldsymbol{\delta}, \sigma^2_{\boldsymbol{\gamma}_{\text{ls},k}}, \cdot)$ of the shared effects neither have tractable full conditionals. Therefore, we also apply an MH-step with IWLS-proposal here. The procedure is similar to the survival specific coefficients but needs to consider the joint likelihood of both model parts. First consider the full conditional

$$p\left(\boldsymbol{\gamma}_{\text{ls, k}} \mid \cdot\right) \propto$$
$$p(\boldsymbol{\gamma}_{\text{ls, k}} \mid \sigma^2_{\boldsymbol{\gamma}_{\text{ls}}})\; p(\boldsymbol{y} \mid \boldsymbol{\eta}_{\text{l}}, \boldsymbol{\eta}_{\text{ls}}, \cdot)\; p(\boldsymbol{\delta} \mid \boldsymbol{\eta}_{\text{s}}, \boldsymbol{\eta}_{\text{ls}}, \cdot).$$

Let $\boldsymbol{Z}_{\text{ls},k}$ be the corresponding design matrix of effect $\boldsymbol{\gamma}_{\text{ls},k}$ and $\frac{1}{\sigma^2_{\boldsymbol{\gamma}_{\text{ls},k}}}\boldsymbol{K}$ the variance statement of prior $p(\boldsymbol{\gamma}_{\text{ls},k} \mid \sigma^2_{\boldsymbol{\gamma}_{\text{ls},k}})$ (compare prior given in 4). Now to approximate the full conditional we draw IWLS proposal $\boldsymbol{\gamma}^*_{\text{ls},k}$ from a normal distribution density $q(\boldsymbol{\gamma}^*_{\text{ls},k} \mid \boldsymbol{\gamma}^{[t]}_{\text{ls},k})$ with $\boldsymbol{\gamma}^{[t]}_{\text{ls},k}$ being the value of $\boldsymbol{\gamma}_{\text{ls},k}$ at iteration t of the MCMC algorithm. More specifically

$$\boldsymbol{\gamma}^*_{\text{ls},k} \sim N\left(\boldsymbol{\mu}^{[t]}_{\boldsymbol{\gamma}_{\text{ls},k}}, \left(\boldsymbol{P}^{[t]}_{\boldsymbol{\gamma}_{\text{ls},k}}\right)^{-1}\right)$$

$$\text{with} \quad \boldsymbol{P}^{[t]}_{\boldsymbol{\gamma}_{\text{ls},k}} = \boldsymbol{Z}'_{\text{ls},k}\boldsymbol{W}^{[t]}_{\text{ls}}\boldsymbol{Z}_{\text{ls},k} + \frac{1}{\sigma^{2[t]}_{\boldsymbol{\gamma}_{\text{ls},k}}}\boldsymbol{K}_{\boldsymbol{\gamma}_{\text{ls},k}}$$

$$\text{and} \quad \boldsymbol{\mu}^{[t]}_{\boldsymbol{\gamma}_{\text{ls},k}} = \left(\boldsymbol{P}^{[t]}_{\boldsymbol{\gamma}_{\text{ls},k}}\right)^{-1}\boldsymbol{Z}'_{\text{ls},k}\boldsymbol{W}^{[t]}_{\text{ls}}\left(\tilde{\boldsymbol{y}}^{[t]}_{\text{ls}} - \boldsymbol{\eta}^{[t]}_{\text{ls},-k}\right).$$

The rest of the algorithm is analogous to the survival effects.

To see how the working weights and observations build for the coefficients in the shared predictor, consider first the log-full conditional

$$\log(p(\boldsymbol{\gamma}_{\text{ls},k} \mid \boldsymbol{y}, \boldsymbol{\delta}, \sigma^2_{\boldsymbol{\gamma}_{\text{ls},k}}, \cdot)) \propto$$
$$-\frac{1}{2\sigma^2_{\boldsymbol{\gamma}_{\text{ls},k}}}\boldsymbol{\gamma}'_{\text{ls},k}\boldsymbol{K}_{\boldsymbol{\gamma}_{\text{ls},k}}\boldsymbol{\gamma}_{\text{ls},k} + \ell_y(\boldsymbol{\eta}_{\text{ls}}) + \ell_\delta(\boldsymbol{\eta}_{\text{ls}}),$$

where $\ell_y(\boldsymbol{\eta}_{\text{ls}})$ denotes the longitudinal part of the log-likelihood and $\ell_\delta(\boldsymbol{\eta}_{\text{ls}})$ the survival/ poisson part of the log-likelihood depending on predictor $\boldsymbol{\eta}_{\text{ls}} = \sum_{k=1}^{p_{\text{ls}}} \boldsymbol{Z}_{\text{ls},k}\boldsymbol{\gamma}_{\text{ls},k}$ (compare 3) thus including $\boldsymbol{\gamma}_{\text{ls},k}$.

The vector of scores, i.e.~first derivatives of the log-likelihoods with respect to the $\boldsymbol{\eta}_{\text{ls}}$ evaluated at iteration t, is

$$\boldsymbol{v}^{[t]}_{\text{ls}} = \boldsymbol{v}^{[t]}_{y,\text{ls}} + \alpha\boldsymbol{v}^{[t]}_{\delta,\text{ls}} = \frac{\partial\ell_y(\boldsymbol{\eta}^{[t]}_{\text{ls}})}{\partial\boldsymbol{\eta}^{[t]}_{\text{ls}}} + \frac{\partial\ell_\delta(\boldsymbol{\eta}^{[t]}_{\text{ls}})}{\partial\boldsymbol{\eta}^{[t]}_{\text{ls}}}.$$

The working weights evaluated at iteration t can then be derived as

$$\boldsymbol{W}^{[t]}_{\text{ls}} = \text{diag}\left(w_1\left(\eta^{[t]}_{\text{ls},1}\right), \ldots, w_{n_a}\left(\eta^{[t]}_{\text{ls},n_a}\right)\right)$$

with n_a the number of observations in the augmented dataset and

$$\begin{aligned} w_i(\eta^{[t]}_{\text{ls},i}) &= -\text{E}\left(\frac{\partial v^{[t]}_{\text{ls},i}}{\partial\eta^{[t]}_{\text{ls},i}}\right) = -\text{E}\left(\frac{\partial \boldsymbol{v}^{[t]}_{y,\text{ls},i} + \alpha\boldsymbol{v}^{[t]}_{\delta,\text{ls},i}}{\partial\eta^{[t]}_{\text{ls},i}}\right) \\ &= -\text{E}\left(\frac{\partial\ell_y(\boldsymbol{\eta}^{[t]}_{\text{ls},i})}{\partial^2\boldsymbol{\eta}^{[t]}_{\text{ls},i}}\right) - \text{E}\left(\frac{\partial\ell_\delta(\boldsymbol{\eta}^{[t]}_{\text{ls},i})}{\partial^2\boldsymbol{\eta}^{[t]}_{\text{ls},i}}\right) \\ &= w_{y,i}(\eta^{[t]}_{\text{ls},i}) + \alpha^2 w_{\delta,i}(\eta^{[t]}_{\text{ls},i}). \end{aligned}$$

The working observations then follow analogously as

$$\tilde{\boldsymbol{y}}^{[t]}_{\text{ls}} = \boldsymbol{\eta}^{[t]}_{\text{ls}} + \left(\boldsymbol{W}^{[t]}_{\text{ls}}\right)^{-1}\boldsymbol{v}^{[t]}_{\text{ls}}.$$

A.4 Variances

A.4.1 Model variance

Let $N = \sum_{i=1}^{n} n_i$ be the total number of longitudinal observations as the sum of all observations n_i per individual i across all individuals n. Then the full conditional of the model variance follows as

$$\begin{aligned} p(\sigma^2_\varepsilon \mid \cdot) &\propto p(\sigma^2_\varepsilon)\; p(\boldsymbol{y} \mid \boldsymbol{\eta}_{\text{l}}, \boldsymbol{\eta}_{\text{ls}}, \sigma^2_\varepsilon) \\ &\propto (\sigma^2_\varepsilon)^{-a_0-1}\; \exp\left\{-\tfrac{b_0}{\sigma^2_\varepsilon}\right\} \\ &\quad \left(\sigma^2_\varepsilon\right)^{-\frac{N}{2}} \\ &\quad \exp\left\{-\tfrac{1}{2\sigma^2_\varepsilon}(\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}})'(\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}})\right\} \\ &\propto \left(\sigma^2_\varepsilon\right)^{-(a_0+\frac{N}{2})-1} \\ &\quad \exp\left\{-\tfrac{1}{\sigma^2_\varepsilon}(b_0 + \tfrac{1}{2}(\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}})'(\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}}))\right\} \\ \sigma^2_\varepsilon \mid \cdot &\sim \text{IG}\left(a_0 + \tfrac{N}{2}, b_0 + (\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}})'(\boldsymbol{y} - \boldsymbol{\eta}_{\text{l}} - \boldsymbol{\eta}_{\text{ls}})\right) \end{aligned}$$

A.4.2 Variance of coefficients

With $p_.$ being the number of covariates in each predictor (longitudinal, shared, survival) the full conditional of the $k_. = 1, \ldots, p_.$ variances of the corresponding effects is analogous

$$\begin{aligned} p(\sigma^2_{\boldsymbol{\gamma}_{k.}} \mid \cdot) &\propto p(\sigma^2_{\boldsymbol{\gamma}_{k.}})\; p(\boldsymbol{\gamma}_{k.} \mid \sigma^2_{\boldsymbol{\gamma}_{k.}}) \\ &\propto (\sigma^2_{\boldsymbol{\gamma}_{k.}})^{-a_{k.}-1}\; \exp\left\{-\tfrac{b_{k.}}{\sigma^2_{\boldsymbol{\gamma}_{k.}}}\right\} \\ &\quad \left(\sigma^2_{\boldsymbol{\gamma}_{k.}}\right)^{-\text{rk}\boldsymbol{K}_{k.}}\; \exp\left\{-\tfrac{1}{2\sigma^2_{\boldsymbol{\gamma}_{k.}}}\boldsymbol{\gamma}'_{\cdot,k.}\boldsymbol{K}_{k.}\boldsymbol{\gamma}_{\cdot,k.}\right\} \\ &\propto \left(\sigma^2_{\boldsymbol{\gamma}_{k.}}\right)^{-(a_{k.}+\text{rk}\boldsymbol{K}_{k.})-1} \\ &\quad \exp\left\{-\tfrac{1}{\sigma^2_{\boldsymbol{\gamma}_{k.}}}(b_{k.} + \boldsymbol{\gamma}'_{\cdot,k.}\boldsymbol{K}_{k.}\boldsymbol{\gamma}_{\cdot,k.})\right\} \\ \sigma^2_{\boldsymbol{\gamma}_{k.}} \mid \cdot &\sim \text{IG}\left(a_{k.} + \text{rk}\boldsymbol{K}_{k.}, b_{k.} + \boldsymbol{\gamma}'_{\cdot,k.}\boldsymbol{K}_{k.}\boldsymbol{\gamma}_{\cdot,k.}\right) \end{aligned}$$

B Detailed results of simulations study

Figure 4 displays the results of the simulation study detailed by individual effect.

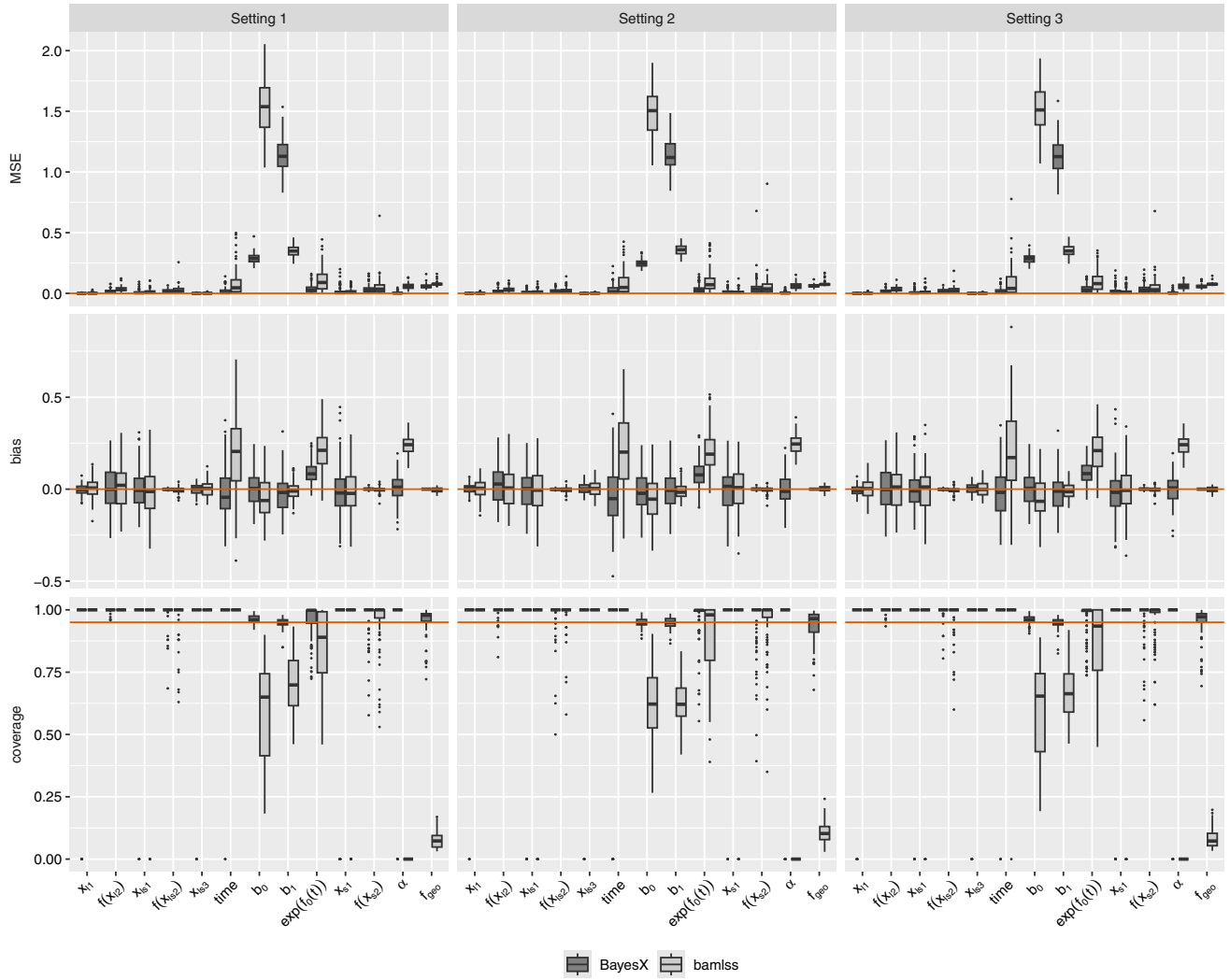

Fig. 4 Boxplots of mean squared error (MSE), bias and 95%-coverage per method by effect and simulation setting (Setting 1 - f_{geo} in η_{ls}, Setting 2 - f_{geo} in η_s, Setting 3 - f_{geo} in η_l). The orange horizontal line marks the reference value of each statistic

C Statistical overview of run times

Table 3 details descriptive measures of the run times.

Table 3 Statistical overview of run times of 100 replications per setting and estimation method

Method	Min.	1st.Qu.	Median	Mean	3rd.Qu	Max
Setting 1						
BayesX	1.30	1.53	1.58	1.58	1.62	1.80
bamlss	3.17	3.30	3.38	3.86	3.42	10.62
Setting 2						
BayesX	1.10	1.22	1.26	1.26	1.30	1.41
bamlss	1.68	2.04	2.17	2.50	2.26	9.62
Setting 3						
BayesX	1.03	1.25	1.30	1.28	1.32	1.40
bamlss	2.01	2.06	2.21	2.58	2.26	12.98

D Unbalanced datasets, the PH likelihood and algorithmic stability

In a joint modelling setting using the PH formulation the survival submodel is denoted as

$$\begin{aligned}\boldsymbol{\lambda}(t) &= \lambda_0(t)\exp\left\{\boldsymbol{\eta}_{\mathrm{s}} + \alpha\boldsymbol{\eta}_{\mathrm{ls}}(t)\right\} \\ &= \exp\left\{\log\lambda_0(t) + \boldsymbol{\eta}_{\mathrm{s}} + \alpha\boldsymbol{\eta}_{\mathrm{ls}}(t)\right\} \end{aligned} \tag{7}$$

(note its similarity to 6) and the likelihood takes the form

$$L(\boldsymbol{T},\boldsymbol{\delta}) = \prod_{i=1}^{n}\lambda_i(T_i)^{\delta_i}\ \exp\left\{-\int_0^{T_i}\lambda_i(u)\,du\right\}.$$

Here T_i is the event or censoring time of individual i, δ_i is the indicator whether i is censored ($\delta_i = 0$) or not ($\delta_i = 1$).

The respective log-likelihood the follows as

$$l(\boldsymbol{T},\boldsymbol{\delta}) = \sum_{i=1}^{n}\left\{\delta_i\lambda_i(T_i) - \int_0^{T_i}\lambda_i(u)\,du\right\} \tag{8}$$

and shows rather distinctly that the contribution of individual i to the (log-)likelihood consists of the risk of experiencing an event at time T_i, i.e. $\lambda_i(T_i)$ if i experienced an event or 0 if i was censored, minus the (logarithmic) probability of survival of i up to point T_i, i.e. $\log S(T_i) = -\int_0^{T_i}\lambda_i(u)\,du$.

If there are few events observed in the data, there are few observations that contribute to the likelihood with information on the risk of an event at time T_i, i.e. the distribution of actual events T_i^*, and thus to the estimation of the baseline hazard $\lambda_0(t)$, which remains very small as a consequence and the probability for survival is very high, $S(t) \approx 1$.

This can be seen when looking at the traditional Cox-PH model, where all covariates are assumed constant (Cox 1972). Then (7) simplifies to $\boldsymbol{\lambda}(t) = \exp\{\log\lambda_0(t) + \boldsymbol{\eta}\}$ with $\boldsymbol{\eta} = \boldsymbol{\eta}_{\mathrm{s}} + \alpha\boldsymbol{\eta}_{\mathrm{ls}}$. For illustration let $\boldsymbol{\eta}_{\mathrm{s}} = \boldsymbol{Z}_{\mathrm{s}}\gamma_{\mathrm{s}}$, i.e. a linear predictor, then the parameter vector corresponding to $\boldsymbol{\eta}$ follows as $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\mathrm{s}}, \alpha)'$ and the design matrix of the data as $\boldsymbol{Z} = (\boldsymbol{Z}_{\mathrm{s}}, \boldsymbol{\eta}_{\mathrm{ls}})$. The hazard ratio of subject i to subject j then is

$$\frac{\lambda_i(t \mid \boldsymbol{\eta}_i)}{\lambda_j(t \mid \boldsymbol{\eta}_j)} = \exp\{\boldsymbol{\gamma}'(\boldsymbol{z}_i - \boldsymbol{z}_j)\},$$

which results in a partial likelihood for estimation of the form (Cox 1975)

$$pl(\boldsymbol{\gamma}) = \sum_{i=1}^{n}\delta_i\left\{\eta_i - \log\sum_{T_j\geq T_i}\exp\eta_j\right\}.$$

From a numerical perspective another problem emerges: The integral in the likelihood (8) has an analytical solution only when the baseline hazard is constant and covariates with the exception of time itself are constant. In all other cases it is analytically intractable. Expectation-Maximization algorithms for joint models usually use Gauss-Kronrod quadrature, `bamlss` uses the trapezoid rule with 25 fixed integration points (Köhler et al. 2018).

Since the latter employs a derivative-based MH-algorithm, an update on the parameters $\boldsymbol{\gamma}_{\cdot}$ governing predictors $\boldsymbol{\eta}_{\mathrm{s}}$ and $\boldsymbol{\eta}_{\mathrm{ls}}(t)$ ensues from

$$\boldsymbol{\gamma}_{\cdot}^{[t]} = \boldsymbol{\gamma}_{\cdot}^{[t-1]} + \nu^{[t-1]}\boldsymbol{H}(\boldsymbol{\gamma}_{\cdot})^{[t-1]}\boldsymbol{s}(\boldsymbol{\gamma}_{\cdot})^{[t-1]},$$

where $\boldsymbol{H}(\boldsymbol{\gamma}_{\cdot})$ is the Hessian matrix of the parameters, $\boldsymbol{s}(\boldsymbol{\gamma}_{\cdot})$ the score vector and ν a parameter specific step length, which is optimized over (0, 1] in each iteration. Apart from this procedure being computationally expensive, the Hessian matrix and the score vector involving the time-to-event likelihood need a numerical approximation as well. For more information see Köhler et al. (2017).

Both factors—few events and numerical instability—especially when combined create a challenging situation for such an algorithm.

In contrast the SPAJM avoids the integral in the time-to-event likelihood via the re-formulation of the risk $\boldsymbol{\lambda}(t)$ as a counting process. The log-likelihood of the ensuing Poisson distribution takes the form

$$\begin{aligned} l(\boldsymbol{\delta} \mid \boldsymbol{\eta}_{\mathrm{s}}, \boldsymbol{\eta}_{\mathrm{ls}}(t)) \ = &\sum_{i=1}^{n}\sum_{j=1}^{J_i}\delta_j\left(f_0(t_j) + o_{ij} + \eta_{\mathrm{s},i} + \alpha\eta_{\mathrm{ls},ij}\right) \\ &- \exp\left\{f_0(t_j) + o_{ij} + \eta_{\mathrm{s},i} + \alpha\eta_{\mathrm{ls},ij}\right\}, \end{aligned}$$

where J_i is the last interval individual i has been observed in. So the contribution of individual i to the (log-)likelihood still consists of the risk of experiencing an event at time $T_i = t_{J_i}$, i.e. $\delta_j \left(f_0(t_j) + o_{ij} + \eta_{\mathrm{s},i} + \alpha \eta_{\mathrm{ls},ij} \right)$, which is 0 if i is censored or $f_0(t_{J_i}) + o_{iJ_i} + \eta_{\mathrm{s},i} + \alpha \eta_{\mathrm{ls},iJ_i}$ if i experienced an event and the probability of survival up to this point, $\exp\left\{ f_0(t_j) + o_{ij} + \eta_{\mathrm{s},i} + \alpha \eta_{\mathrm{ls},ij} \right\}$.

Since the baseline hazard is assumed constant in the intervals defined by the actual event times T_i^* (and the observation times, at which the longitudinal observation happened), the survival probabilities are discretised and the integral can be replaced by a sum. This way numerical integration can be avoided. In addition, few events still impede estimation of the baseline hazard - the resulting intervals are fewer and the boundaries further apart assuming the baseline hazard to be constant over longer periods of time, but without the algorithm having to simultaneously perform numerical integration.

References

Alsefri, M., Sudell, M., García-Fiñana, M., Kolamunnage-Dona, R.: Bayesian joint modelling of longitudinal and time to event data: a methodological review. BMC Med. Res. Methodol. **20**(1), 1–17 (2020)

Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J.J.M., Lesaffre, E.: Joint modeling of two longitudinal outcomes and competing risk data. Stat. Med. **33**(18), 3167–3178 (2014). https://doi.org/10.1002/sim.6158

Barrett, J.K., Huille, R., Parker, R., Yano, Y., Griswold, M.: Estimating the association between blood pressure variability and cardiovascular disease: an application using the ARIC study. Stat. Med. **38**(10), 1855–1868 (2019)

Bartolucci, F., Farcomeni, A.: A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. Biometrics **71**(1), 80–89 (2015). https://doi.org/10.1111/biom.12224

Bartolucci, F., Farcomeni, A.: A shared-parameter continuous-time hidden Markov and survival model for longitudinal data with informative dropout. Stat. Med. **38**(6), 1056–1073 (2019). https://doi.org/10.1002/sim.7994

Belitz, C., Brezger, A., Kneib, T., Lang, S., Umlauf, N.: BayesX: software for Bayesian inference in structured additive regression models (2022). https://www.uni-goettingen.de/de/bayesx/550513.html

Bender, A., Groll, A., Scheipl, F.: A generalized additive model approach to time-to-event analysis. Stat. Model. **18**, 299–321 (2018)

Bender, A., Scheipl, F.: Pammtools: piece-wise exponential Additive Mixed Modeling tools (2018). arXiv:1806.01042 [Stat]

Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J.-F., Jacqmin-Gadda, H.: Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. Biometrics **71**(1), 102–113 (2015). https://doi.org/10.1111/biom.12232

Cox, D.R.: Regression models and life-tables. J. R. Stat. Soc. Ser. B (Methodology) **34**(2), 187–220 (1972)

Cox, D.R.: Partial likelihood. Biometrika **62**(2), 269 (1975). https://doi.org/10.2307/2335362

Crowther, M.J., Abrams, K.R., Lambert, P.C.: Joint modeling of longitudinal and survival data. Stata J. **13**(1), 165–184 (2013). https://doi.org/10.1177/1536867X1301300112

Engstler, H., Hameister, N., Schrader, S.: User manual DEAS SUF 2014. DZA German Centre of Gerontology (2014)

Faucett, C.L., Schenker, N., Elashoff, R.M.: Analysis of censored survival data with intermittently observed time-dependent binary covariates. J. Am. Stat. Assoc. **93**(442), 427–437 (1998). https://doi.org/10.1080/01621459.1998.10473692

Faucett, C.L., Thomas, D.C.: Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat. Med. **15**(15), 1663–1685 (1996)

Follmann, D., Wu, M.: An approximate generalized linear model with random effects for informative missing data. Biometrics **51**(1), 151–168 (1995)

Friedman, M.: Piecewise exponential models for survival data with covariates. Ann. Stat. **10**(1), 101–113 (1982). https://doi.org/10.1214/aos/1176345693

Griesbach, C., Groll, A., Bergherr, E.: Joint modelling approaches to survival analysis via Likelihood-Based boosting techniques. Comput. Math. Methods Med. **2021**, 4384035 (2021)

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

Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R.: JoineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. BMC Med. Res. Methodol. **18**(1), 50 (2018). https://doi.org/10.1186/s12874-018-0502-1

Hogan, J.W., Laird, N.M.: Mixture models for the joint distribution of repeated measures and event times. Stat. Med. **16**(3), 239–257 (1997)

Huang, X., Li, G., Elashoff, R.M., Pan, J.: A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. Lifetime Data Anal. **17**(1), 80–100 (2011)

Huang, Y., Chen, J.: Bayesian quantile regression-based nonlinear mixed-effects joint models for time-to-event and longitudinal data with multiple features. Stat. Med. **35**(30), 5666–5685 (2016). https://doi.org/10.1002/sim.7092

Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J.M.G., Commenges, D.: Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. Biometrics **66**(1), 11–19 (2010). https://doi.org/10.1111/j.1541-0420.2009.01234.x

Klaus, D., Engstler, H.: Daten und methoden des deutschen alterssurveys. In: Mahne, K., Wolff, J.K., Simonson, J., Tesch-Römer, C. (eds.) Altern im Wandel, pp. 29–45. Springer Fachmedien, Wiesbaden (2017). https://doi.org/10.1007/978-3-658-12502-8/_2

Kneib, T., Klein, N., Lang, S., Umlauf, N.: Modular regression-a lego system for building structured additive distributional regression models with tensor product interactions. Test **28**, 1–39 (2019)

Köhler, M., Beyerlein, A., Vehik, K., Greven, S., Umlauf, N., Lernmark, Å., Hagopian, W.A., Rewers, M., She, J.-X., Toppari, J., et al.: Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes: results from the TEDDY study. Acta Diabetol. **54**(11), 1009–1017 (2017)

Köhler, M., Umlauf, N., Beyerlein, A., Winkler, C., Ziegler, A.-G., Greven, S.: Flexible Bayesian additive joint models with an application to type 1 diabetes research. Biom. J. **59**(6), 1144–1165 (2017). https://doi.org/10.1002/bimj.201600224

Köhler, M., Umlauf, N., Greven, S.: Nonlinear association structures in flexible Bayesian additive joint models. Stat. Med. **37**(30), 4771–4788 (2018). https://doi.org/10.1002/sim.7967

Lin, H., McCulloch, C.E., Mayne, S.T.: Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. Stat. Med. **21**(16), 2369–2382 (2002)

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

Martins, R., Silva, G.L., Andreozzi, V.: Joint analysis of longitudinal and survival AIDS data with a spatial fraction of long-term survivors: a Bayesian approach. Biom. J. **59**(6), 1166–1183 (2017). https://doi.org/10.1002/bimj.201600159

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)

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/ (2022)

Rappl, A., Mayr, A., Waldmann, E.: More than one way: exploring the capabilities of different estimation approaches to joint models for longitudinal and time-to-event outcomes. Int. J. Biostat. **18**(1), 127–149 (2022). https://doi.org/10.1515/ijb-2020-0067

Rizopoulos, D.: JM: an R package for the joint modelling of longitudinal and time-to-event data. J. Stat. Softw. **35**(9), 1–33 (2010). https://doi.org/10.18637/jss.v035.i09

Rizopoulos, D.: Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics **67**(3), 819–829 (2011). https://doi.org/10.1111/j.1541-0420.2010.01546.x

Rizopoulos, D.: The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. J. Stat. Softw. (2016). https://doi.org/10.18637/jss.v072.i07

Rizopoulos, D., Ghosh, P.: A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat. Med. **30**(12), 1366–1380 (2011). https://doi.org/10.1002/sim.4205

Rizopoulos, D., Verbeke, G., Lesaffre, E., Vanrenterghem, Y.: A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. Biometrics **64**(2), 611–619 (2008). https://doi.org/10.1111/j.1541-0420.2007.00894.x

Roy, J.: Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. Biometrics **59**(4), 829–836 (2003). https://doi.org/10.1111/j.0006-341X.2003.00097.x

Tseng, Y.-K., Hsieh, F., Wang, J.-L.: Joint modelling of accelerated failure time and longitudinal data. Biometrika **92**(3), 587–603 (2005). https://doi.org/10.1093/biomet/92.3.587

Tsiatis, A.A., Davidian, M.: Joint modeling of longitudinal and time-to-event data: an overview. Stat. Sin. **14**(3), 809–834 (2004)

Tsonaka, R., Verbeke, G., Lesaffre, E.: A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. Biometrics **65**(1), 81–87 (2009). https://doi.org/10.1111/j.1541-0420.2008.01021.x

Umlauf, N., Klein, N., Simon, T., Zeileis, A.: bamlss: a Lego toolbox for flexible Bayesian regression (and beyond). J. Stat. Softw. **100**(4), 1–53 (2021). https://doi.org/10.18637/jss.v100.i04

Viviani, S., Alfó, M., Rizopoulos, D.: Generalized linear mixed joint model for longitudinal and survival outcomes. Stat. Comput. **24**(3), 417–427 (2014)

Waldmann, E., Taylor-Robinson, D., Klein, N., Kneib, T., Pressler, T., Schmid, M., Mayr, A.: Boosting joint models for longitudinal and time-to-event data. Biom. J. **59**(6), 1104–1121 (2017). https://doi.org/10.1002/bimj.201600158

WHO: World report on ageing and health. World Health Organisation; Geneva (2015)

Wu, M.C., Bailey, K.: Analysing changes in the presence of informative right censoring caused by death and withdrawal. Stat. Med. **7**(1–2), 337–346 (1988). https://doi.org/10.1002/sim.4780070134

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)

Wulfsohn, M.S., Tsiatis, A.A.: A joint model for survival and longitudinal data measured with error. Biometrics **53**(1), 330 (1997). https://doi.org/10.2307/2533118

Yuen, H.P., Mackinnon, A.: Performance of joint modelling of time-to-event data with time-dependent predictors: an assessment based on transition to psychosis data. PeerJ **4**, e2582 (2016). https://doi.org/10.7717/peerj.2582

Zhang, H., Huang, Y., Wang, W., Chen, H., Langland-Orban, B.: Bayesian quantile regression-based partially linear mixed-effects joint models for longitudinal data with multiple features. Stat. Methods Med. Res. **28**(2), 569–588 (2019). https://doi.org/10.1177/0962280217730852