Flexible parametric joint modelling of longitudinal and survival data

Michael J. Crowther[1]*[†], Keith R. Abrams[1] and Paul C. Lambert[1,2]

The joint modelling of longitudinal and survival data is a highly active area of biostatistical research. The submodel for the longitudinal biomarker usually takes the form of a linear mixed effects model. We describe a flexible parametric approach for the survival submodel that models the log baseline cumulative hazard using restricted cubic splines. This approach overcomes limitations of standard parametric choices for the survival submodel which can lack the flexibility to effectively capture the shape of the underlying hazard function. Numerical integration techniques, such as Gauss-Hermite quadrature, are usually required to evaluate both the cumulative hazard and the overall joint likelihood; however, by using a flexible parametric model the cumulative hazard has an analytically tractable form, providing considerable computational benefits. An extensive simulation study is conducted to assess the proposed model, comparing it to a B-spline formulation, illustrating insensitivity of parameter estimates to the baseline cumulative hazard function specification. Furthermore, we compare non-adaptive and fully adaptive quadrature, showing the superiority of adaptive quadrature in evaluating the joint likelihood. A useful technique to simulate survival times from complex baseline hazard functions is described. The methods are illustrated using an example dataset investigating the effect of longitudinal prothrombin index on survival of patients with liver cirrhosis, showing greater flexibility and improved stability with fewer parameters under the proposed model compared to the B-spline approach. User-friendly Stata software is provided. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: joint modelling, flexible parametric survival models, mixed effects, restricted cubic splines, Gauss-Hermite quadrature.

1. Introduction

The joint modelling of longitudinal and time-to-event data has received remarkable attention in the literature over the past 15 years [1, 2], with the ability to investigate the inter-relationships between the joint processes being advocated in ever widening fields of study [3, 4]. Extensions to the now standard single longitudinal response and single time-to-event joint model include: incorporation of multiple longitudinal markers, both classically [5] and using a Bayesian approach [6], extension to the competing risks setting [7], investigation of a cure proportion [8], and a variety of time-to-event submodels [3, 9]. Extensive outlines of the field are described by Tsiatis and Davidian [10] and Yu et al. [11].

The form of joint model which has dominated the literature assumes that the association between the time-to-event and longitudinal marker is characterised by shared random effects, and it is this approach which we adopt. In

[1] *Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH.*
[2] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, S-171 77 Stockholm, Sweden.*
Correspondence to: Michael J. Crowther. Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH.
†*michael.crowther@le.ac.uk*

order to estimate such models, computationally intensive numerical integration techniques, such as Gauss-Hermite quadrature [12], are required to evaluate both the cumulative hazard function and the full joint likelihood.

We propose to use restricted cubic splines to model the log baseline cumulative hazard function, to provide a highly flexible framework to capture complex hazard functions. Royston and Parmar [13] first proposed this form of survival model by expanding log time into a restricted cubic spline basis. Rizopoulos et al. [14] developed a joint model using this form, but expanded log time by using B-splines. We extend their approach by alternatively using restricted cubic splines [15, 16], which impose the restriction that the fitted function is forced to be linear beyond the boundary knots. The number of parameters in a restricted cubic spline specification is equal to the number of internal knots plus 2 (one of which is an intercept). The number of parameters in a B-spline formulation is equal to the order plus the number of internal knots.

There are 2 primary motivations for our approach. Firstly, standard parametric survival models can often lack the flexibility to capture complex underlying hazard functions, for example, the Weibull assumes a monotonic shape, which will be insufficiently flexible to fully capture a hazard function with a turning point. Patient specific conditional survival predictions [17], a key tool of the joint model framework, rely on sufficiently capturing the baseline hazard function. Secondly, joint models can be considered computationally intensive, therefore by modelling on the log cumulative hazard scale we avoid numerically integrating the hazard function, required to evaluate the joint likelihood.

When applying this form of survival model, knot locations are often defined based on the distribution of event times. The linearity assumption is before the first knot and after the final knot which leads to stability in the estimated function at the extremes of the data. Furthermore, the linearity assumption beyond the final boundary knot is likely to be sensible in terms of extrapolation for conditional survival predictions. The parametric nature of the time-to-event submodel ensures that smooth continuous time predictions can be obtained and tailored at the individual level, allowing out of sample predictions to be made.

The estimation of joint models has almost exclusively been implemented using EM algorithms, where in the expectation step the unknown random effects are treated as missing values. Alternatively, estimation can be conducted via a direct maximisation of the observed data log-likelihood using standard maximisation techniques such as the Newton-Raphson algorithm. It is the second approach which we adopt to fit the models. As has been discussed in Rizopoulos et al. [14], the score equations can be evaluated analytically; however, as with the log-likelihood, numerical integration is required to compute them. Within a generalised linear mixed effects model context, Lessaffre and Spiessens [18] have shown that often the integrals required for such analytical derivatives are more poorly approximated by quadrature compared to the numerical estimates obtained using finite differences. However, an issue neglected in the joint model context is an evaluation of Gauss-Hermite quadrature to calculate the likelihood. We conduct a simulation study to not only evaluate the proposed joint model, but also evaluating non-adaptive quadrature with varying numbers of nodes, and fully adaptive quadrature.

The methods are illustrated using a dataset of 488 patients with liver cirrhosis [19]. A total of 251 patients were randomised to receive prednisone, with 237 assigned to a placebo. Prothrombin index was measured repeatedly at baseline, with subsequent scheduled measurements at 3, 6, 12 months and then annually; however, observed time of measurements varied substantially. A total of 2968 measurements were recorded. We investigate the effect of treatment after adjusting for the relationship between prothrombin index and time to death.

The remainder of the paper is organised as follow: Section 2 details the formulation of each submodel and the full joint model, Section 3 describes a simulation study evaluating the finite sample performance of the proposed joint model, including an assessment of Gauss-Hermite quadrature, and Section 4 applies the proposed model to the liver cirrhosis dataset. We conclude the paper in Section 5 with a discussion.

2. Defining the joint model

For the ith patient, we observe time-to-event data, longitudinal response data and covariate data. Let S_i be the survival time of patient $i = 1, \ldots, n$, and $T_i = \min(S_i, C_i)$ the observed survival time, with C_i the censoring time. Define an event indicator d_i, which takes the value of 1 if $S_i \leq C_i$ and 0 otherwise. Let $y_{ij} = \{y_i(t_{ij}), j = 1, \ldots, m_i\}$ denote the jth longitudinal response measurement of a continuous biomarker for the ith patient taken at times t_{ij}. Furthermore, we define shared random effects, b_i, which characterise the survival and longitudinal processes. Each submodel can be dependent on a set of baseline covariates, U_i, which can potentially differ between submodels. Both censoring and time of measurements are assumed to be non-informative.

2.1. Longitudinal submodel

We assume a linear mixed effects model for the continuous longitudinal process. Therefore, we observe:

$$y_i(t_{ij}) = W_i(t_{ij}) + e_{ij}, \qquad e_{ij} \sim N(0, \sigma_e^2) \tag{1}$$

$$W_i(t_{ij}) = X_i'(t_{ij})\beta + Z_i'(t_{ij})b_i + u_i\delta \tag{2}$$

with design matrices X_i and Z_i for the fixed (β) and random (b_i) effects, respectively, consisting of time variables. Furthermore we also have a vector of time invariant baseline covariates, $u_i \in U_i$, and corresonding regression coefficients, δ. We assume the error is independent from the random effects, and that $\text{cov}(e_{ij}, e_{ik}) = 0$ (where $j \neq k$). Flexibility in the longitudinal submodel can be incoporated through the use of fractional polynomials of time, for example, which will often be sufficient to capture the longitudinal trajectory [20].

2.2. Survival submodel

We define the proportional cumulative hazards time-to-event submodel:

$$\log\{H(t|\boldsymbol{b}_i, u_{bs,i})\} = \log\{H_0(t)\} + \alpha W_i(t_{ij}) + \phi v_i \tag{3}$$

where $H_0(t)$ is the cumulative baseline hazard function, α denotes the association parameter and ϕ is a set of regression coefficients associated with a set of baseline covariates, v_i, again a subset of U_i. In this formulation we assume the association is based on the current value of the longitudinal response. A useful discussion regarding the choice of association measure can be found in Rizopoulos and Ghosh [6].

The spline basis for this specification is derived from the log cumulative hazard function of a Weibull proportional hazards model. The linear relationship with log time is relaxed through the use of restricted cubic splines. Further details can be found in Royston and Parmar [13] and Lambert and Royston [16]. We can therefore write a restricted cubic spline function of $\log(t)$, with knots $\mathbf{k}_0$, as $s\{\log(t)|\gamma, \mathbf{k}_0\}$. For example, with K knots and letting $x = \log(t)$, a restricted cubic spline function can be expressed as:

$$s(x) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \cdots + \gamma_{K-1} z_{K-1} \tag{4}$$

and

$$z_1 = x \tag{5}$$

$$z_p = (x - k_p)_+^3 - \kappa_p(x - k_p)_+^3 - (1 - \kappa_p)(x - k_K)_+^3 \qquad p = 2, \ldots, K-1 \tag{6}$$

This is now substituted for the log cumulative baseline hazard in equation (3).

$$\log\{H(t|\boldsymbol{b}_i, u_{bs,i})\} = \eta_i = s\{\log(t)|\gamma, \mathbf{k}_0\} + \alpha W_i(t_{ij}) + \phi u_{bs,i} \tag{7}$$

We can now transform to the hazard and survival scales

$$h(t|\boldsymbol{b}_i, u_{bs,i}) = \left\{\frac{1}{t}\frac{ds\{\log(t)|\gamma, \mathbf{k}_0\}}{d\log(t)} + \alpha\frac{dW(t)}{dt}\right\}\exp(\eta_i), \qquad S(t|\boldsymbol{b}_i, u_{bs,i}) = \exp\{-\exp(\eta_i)\} \tag{8}$$

Given the fully parametric nature of the model specification, the derivatives of the spline function required in the definition of the hazard function are easily calculable. These functions are of course specific to using the current value as the measure of association and can be adjusted according to the form of association that is being investigated.

2.3. Full joint likelihood

We can now construct the full likelihood for the joint model:

$$\prod_{i=1}^{n}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{m_i} f(y_i(t_{ij})|b_i, \theta)\right) f(b_i|\theta) f(T_i, d_i|b_i, \theta)\, db_i\right] \tag{9}$$

where

$$f(y_i(t_{ij})|b_i, \theta) = (2\pi\sigma_e^2)^{-1/2}\exp\left\{-\frac{y_i(t_{ij}) - W_i(t_{ij})}{2\sigma_e^2}\right\}, \tag{10}$$

$$f(b_i|\theta) = (2\pi|V|)^{-1/2}\exp\left\{-\frac{b_i'V^{-1}b_i}{2}\right\}, \tag{11}$$

and

$$f(T_i, d_i|b_i, \theta) = \left[\left\{\frac{1}{T_i}\frac{\mathrm{d}s\{\log(T_i)|\gamma, \mathbf{k}_0\}}{\mathrm{d}\log(T_i)} + \alpha\frac{\mathrm{d}W(T_i)}{\mathrm{d}T_i}\right\}\exp(\eta_i)\right]^{d_i}\exp\left\{-\exp(\eta_i)\right\} \tag{12}$$

with parameter vector θ, and V the variance-covariance matrix of the random effects. Equation (9) can be evaluated using m-point non-adaptive or fully adaptive Gauss-Hermite quadrature [12, 21]. Fully adaptive quadrature iteratively places the nodes at the optimum positions for each patient, resulting in a much reduced number of nodes required to obtain reliable estimates, providing substantial computational benefits.

3. Simulation study

A simulation study was undertaken to assess the performance of the proposed joint model for finite sample sizes, comparing to the model of Rizopoulos et al. (2009) [14]. Under each scenario we apply both the proposed joint model with 5 degrees of freedom, plus an intercept, resulting in 6 parameters to capture the baseline cumulative hazard function, and a B-spline function of degree 3 (cubic splines for consistency) and 2 internal knots, again resulting in 6 parameters to capture the baseline hazard, providing a fair comparison. We also apply the proposed joint model with 1 degree of freedom (equivalent to a Weibull based joint model). Our proposed model is implemented in the stjm command in Stata [22]. We use the JM package [23] version 0.4.0 in R to fit the model of Rizopoulos et al. (2009). Both Stata and R packages use a tolerance of 1E-04 for the parameter estimates, and Stata uses 1E-07 for the log-likelihood, with R using 1.5E-8.

Furthermore, what has often been neglected in the literature, is assessment on the number of quadrature nodes required to obtain consistent parameter estimates of effect and sufficient coverage probabilities. To each scenario we use non-adaptive Gauss-Hermite quadrature to evaluate the joint likelihood of both the proposed model and the B-spline model, with 5 and 15 nodes for each random effect dimension to assess performance. We also implement 5-point fully adaptive quadrature for the proposed model for comparison.

For each scenario, 300 patients were included in each of 500 repetitions. The true longitudinal profile was generated from $W_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \delta u_i$, with $\beta_{0i} \sim N(0,1)$, $\beta_{1i} \sim N(0, 0.25^2)$ and correlation between (β_{0i}, β_{1i}) of 0.25. The observed longitudinal measurements were then generated from $Y_{ij} \sim N(W_{ij}, 0.5^2)$. Time of measurements were fixed at $(0, 1, 2, 3, 4)$. Survival times were generated from $\log(H(t|b_i)) = \log(H_0(t)) + \alpha W_i(t) + \phi u_i$, where $H_0(t)$ is detailed below. Censoring was applied at 5 years. A binary treatment group variable was generated from $u_i \sim Bin(1, 0.5)$. The direct treatment effect on the longitudinal response, δ, is fixed at -0.25, the direct treatment effect on the time-to-event, ϕ, is fixed at 0.25, and the association paramter, α, is varied between $\{-0.25, 0.25\}$.

3.1. Generating survival times

Often simulation studies will generate survival times from an exponential distribution, which assumes a constant baseline hazard function. In many situations this may lack biological plausibility. For example, the method of Rizopoulos et al. (2009) was evaluated in a simulation study with survival times generated from an exponential distribution (however, the primary motivation of the simulation study was to evaluate the Laplacian estimation method, not the survival submodel).

Under standard survival models, Bender et al. [24] have described an efficient algorithm to generate survival times with a variety of parametric choices for the baseline hazard function; however, when incorporating a time varying biomarker, this produces an equation which cannot be directly solved for T, where T is the generated survival time. Furthermore, an assumption of a constant baseline hazard could be considered too simplistic to fully assess the performance of a model. To fully assess our approach in capturing complex baseline hazard functions, with turning points, we generate survival times from a 2-component mixture Weibull distribution [25], with:

$$S_0(t) = p\exp\{-\lambda_1 t^{\gamma_1}\} + (1-p)\exp\{-\lambda_2 t^{\gamma_2}\} \tag{13}$$

and

$$H_0(t) = -\log(p\exp\{-\lambda_1 t^{\gamma_1}\} + (1-p)\exp\{-\lambda_2 t^{\gamma_2}\}) \tag{14}$$

We now add the linear predictor for the association and time independent covariates, on the log cumulative hazard scale:

$$\log(H(t)) = \log(-\log(p\exp\{-\lambda_1 t^{\gamma_1}\} + (1-p)\exp\{-\lambda_2 t^{\gamma_2}\})) + \alpha W_i(t) + \phi u_i \tag{15}$$

Following the formulation of Bender et al. [24]:

$$S(t) = 1 - F(t), \quad \text{where} \quad F \sim U(0,1) \tag{16}$$

and using equation (15), we can generate survival times from:

$$1 - F = \exp\left(-H(t)\right) \tag{17}$$
$$= [\log(p\exp\{-\lambda_1 t^{\gamma_1}\} + (1-p)\exp\{-\lambda_2 t^{\gamma_2}\})]\exp(\alpha W_i(t) + \phi u_i) \tag{18}$$

Equation (18) is analytically intractible and so cannot be directly re-arranged to find t; however, methods to overcome this include Newton-Raphson iterations or non-linear least squares. This approach could be used in a variety of settings to better assess survival models. Hence, although we do not fit the true model from which we simulate, we use a sufficiently complex underlying shape to truly assess the proposed model specification.

Three scenarios of baseline parameters were chosen; a standard Weibull with increasing hazard function, $\{\lambda_1 = 0.1, \ \gamma_1 = 1.5, \ \text{and} \ p = 1\}$, a mixture Weibull with a single turning point in the baseline hazard function, $\{\lambda_1 = 0.1, \ \gamma_1 = 1.5, \ \lambda_2 = 0.1, \ \gamma_2 = 0.5, \ \text{and} \ p = 0.9\}$, and finally a Weibull distribution with $\{\lambda_1 = 1E - 05, \ \gamma_1 = 6.1, \ \text{and} \ p = 1\}$. The final scenario is to assess the validity of our approach when the hazard is essentially zero for a portion of the follow-up time.

3.2. Results

Tables 1, 2 and 3 present bias and coverage estimates from all simulations generated under the three baseline hazard functions. Under the three scenarios, survival submodel parameters estimates from the proposed model, i.e. the direct treatment effect on survival (ϕ) and the association parameter (α), appear to be unbiased. However, under the B-spline approach, across all scenarios we observe consistent under-estimation of the association parameter, α. This bias is eliminated under the restricted cubic spline approach. For example, under scenario 2 with the true alpha = 0.25, the percentage bias under the restricted cubic spline approach is -0.8% compared to -10% under the B-spline approach. Coverage probabilities very closely aproximate the desired 95% in all scenarios when using restricted cubic splines, even with a small number of non-adaptive quadrature nodes. For the longitudinal submodel parameters, we observe generally unbiased estimates; however, in respect to variance parameters, only when the number of non-adaptive quadrature nodes ≥ 15 or when using fully adaptive quadrature. Under non-adaptive quadrature, coverage estimates are generally below the desired 95% indicating a marked underestimation of the standard errors, compared with optimum coverage probabilities across scenarios when fully adaptive quadrature is used. Further simulations, not shown here, illustrated that 35 non-adaptive quadrature nodes were required to provide optimum coverage probabilities. Standard errors of variance parameters are not available in R so coverage could not be assessed for all parameters in the B-spline models.

Our proposed model also produces moderate bias in the variance estimate of the slope parameter when 5 point non-adaptive quadrature is used; however, this bias is eliminated under both 15 point non-adaptive and 5 point adaptive quadrature. Comparing across degrees of freedom, we observe almost identical estimates of bias and coverage probabilities between models.

Table 2 presents bias and coverage estimates for simulations generated from a 2-component mixture Weibull baseline hazard, described in Section 3.1. Results appear entirely consistent with those found when generating under a standard Weibull distribution. The underestimation of the standard errors of the longitudinal parameters remains a problem when an insufficient number of quadrature nodes are used. Despite generating data from a complex baseline hazard, the joint models fitted with only one degree of freedom appear to estimate all parameters just as effectively as with 5 degress of freedom, specifically the three treatment effects. This can perhaps be expected, as is often the case the hazard ratio can be insensitive to specification of the baseline hazard function [26].

We discuss the implications of the choice of the number of quadrature nodes, and the insensitivity to the baseline hazard in Section 5.

4. Analysis of liver cirrhosis dataset

In this section we apply the proposed joint model to the dataset introduced in Section 1, where primary interest is the effect of treatment after adjusting for the repeatedly measured prothrombin index on the time to all-cause death. A total of 488 patients had their prothrombin index measured at baseline, with further scheduled measurements at 3, 6, and 12 months, and annually thereafter. Median number of measurements was 6 (range: 1, 17). 292 (59.8%)
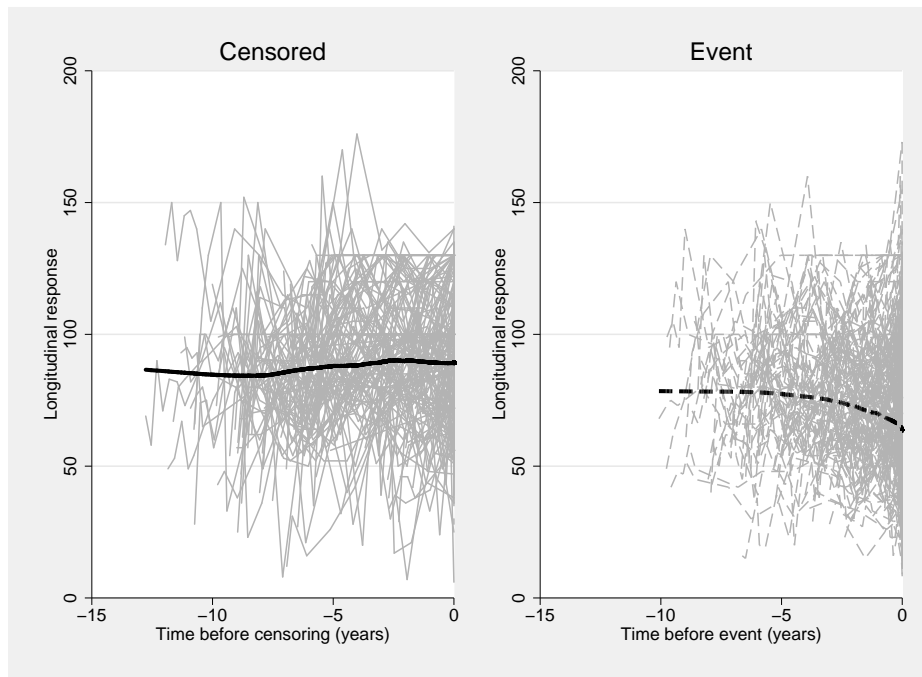
Figure 1. Longitudinal profiles of prothrombin index for patients who were censored/died. Timescale is adjusted by taking away each patients survival time. Lowess smoother overlaid.

patients died during the study. Patients were randomised to 2 treatment groups, namely prednisone or placebo. For further details regarding the dataset, we refer the reader elsewhere [19].

Figure 1 provides an initial exploration of the relationship between prothrombin trajectory and the time (in years) to death by plotting the observed longitudinal responses against observation time, where the timescale is adjusted by taking away the observed censoring/event time. A lowess smoother is overlaid. From Figure 1, it is apparent that patients who experienced the event, compared to patients who were censored, had decreasing levels of the biomarker during the 2-3 year period before death. If we assume the association between the longitudinal and survival models is based on the current value parameterisation discussed in Section 2.2, we would expect a negative association indicating a lower value of prothrombin index has an increased risk of death. This form of plot can be a useful exploratory tool in the analysis of joint longitudinal and survival data.

We now apply the joint model described in Section 2 to the liver cirrhosis dataset. In the longitudinal submodel we assume a random intercept with random effect of log(time), and also adjust for the interaction between treatment and time. In preliminary analysis log(time) showed an improved fit compared to a linear effect of time. In the survival submodel we adjust for the direct effect of treatment. We model the association between prothrombin index and time to death through the current value parameterisation. We use 5 degrees of freedom to model the baseline cumulative hazard, equivalent to 4 internal knots. Boundary knots are placed at the 0th and 100th percentiles of the uncensored log survival times. For comparison we also apply the model of Rizopoulos et al. (2009). Under the B-spline model we use cubic splines with 2 internal knots to provide a comparison of model fit with the same number of parameters used to model the baseline cumulative hazard function. As adaptive quadrature is not available for the B-spline model, we apply both models using 35 point non-adaptive quadrature.

In Table 4, comparing between our proposed approach and the B-spline model, we generally observe similar parameter estimates, in particular both models show a negative association between prothrombin index and time to death, for example under our approach we observe an association of -0.038 (95% CI: -0.045, -0.031), indicating a lower value of prothrombin index increases the risk of death. We observe a non-statistically significant direct effect of treatment on survival with a log hazard ratio of 0.210 (95% CI: -0.038, 0.457).

We now return to our primary motivation of our approach which is to effectively capture complex hazard functions. We compare the fitted marginal survival functions across models with the Kaplan-Meier estimates for the liver cirrhosis dataset, shown in Figure 2. It is evident from Figure 2 that the restricted cubic spline approach provides an improved fit compared to the B-spline approach, using the same number of parameters to model the baseline cumulative hazard function. Indeed, in Figure 3 we show the marginal survival function with an increased number of internal knots under the B-spline approach, highlighting that we need to use 5 internal knots to achieve a function

which fits as closely as the restricted cubic splines approach. In other words, we need to use 9 parameters under the B-spline approach compared to only 6 under the restricted cubic spline approach to achieve a well-fitting function.
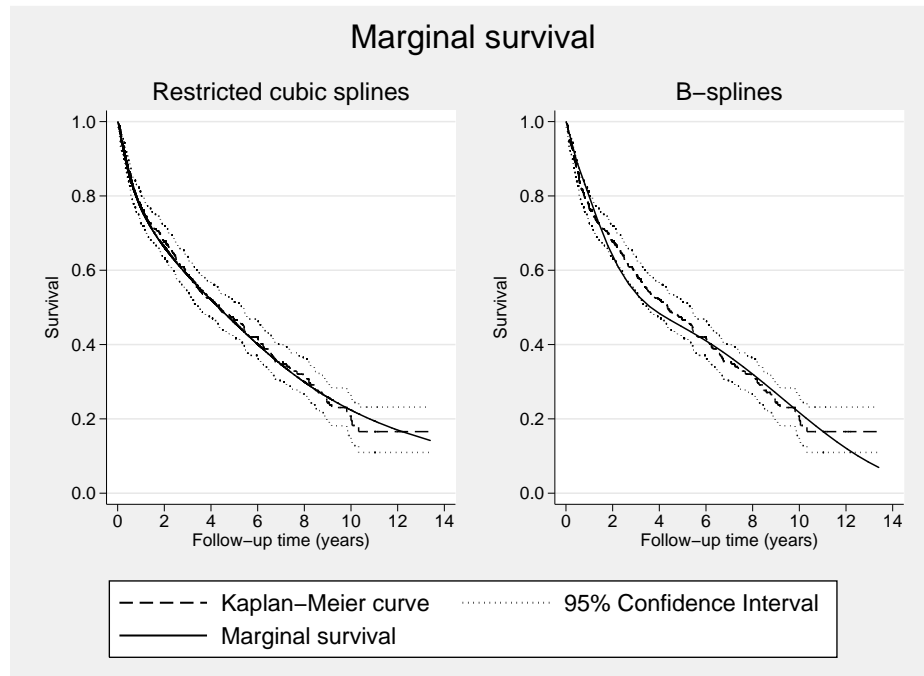

Figure 2. Marginal survival using 6 parameters to model the baseline cumulative hazard function, overlaid on the Kaplan-Meier estimate and it's associated 95% confidence interval.
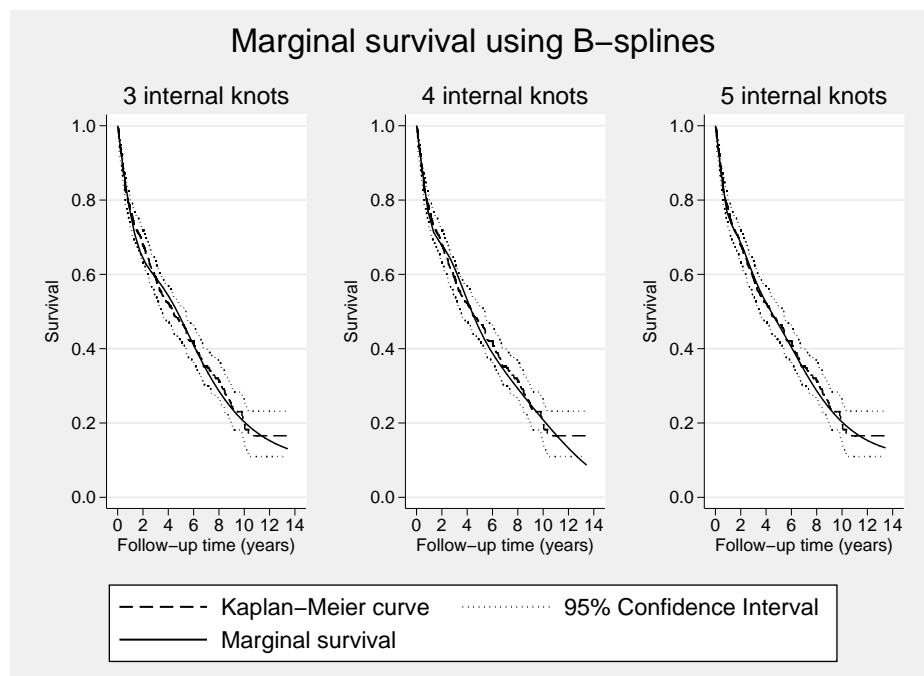

Figure 3. Marginal survival using cubic B-splines with 3, 4 or 5 internal knots, resulting in 7, 8 or 9 parameters to model the baseline cumulative hazard function, overlaid on the Kaplan-Meier estimate and it's associated 95% confidence interval.

4.1. Predictions

To illustrate the prognostic benefits of the joint modelling framework, conditional survival predictions can be tailored at the individual levels using the empirical Bayes predictions from the random effects, and appropriate sampling schemes have been proposed to calculate accurate standard errors for these predictions [17]. We adapt the approach of Rizopoulos to calculate conditional survival predictions of 2 patients with similar baseline values of prothrombin index, using the fitted restricted cubic spline based joint model, shown in Figure 4.
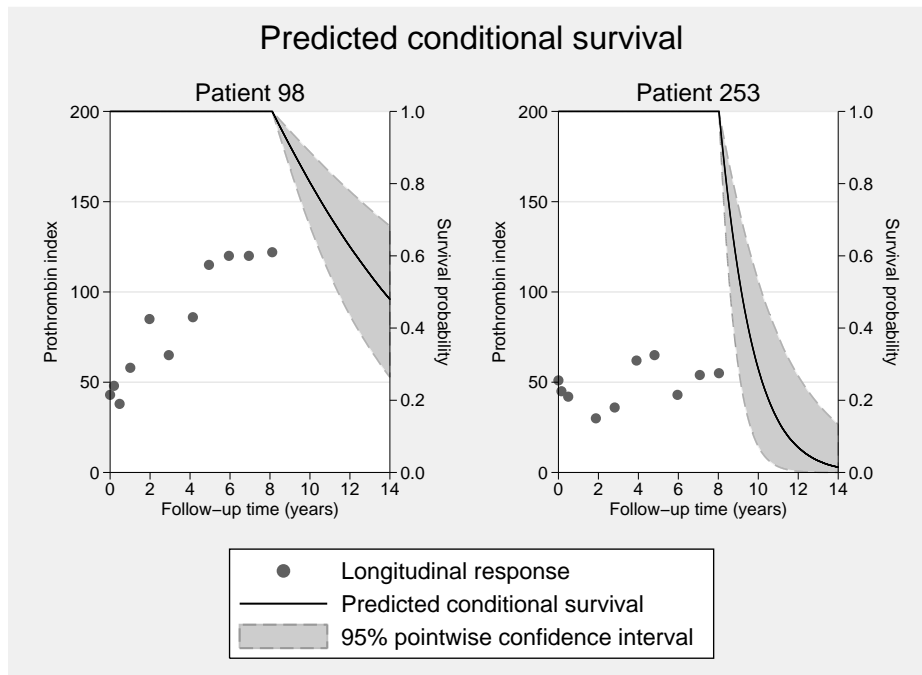

Figure 4. Predicted conditional survival probabilities for patients 98 and 253.

Given the negative association between prothrombin index and an increased risk of death, we see from Figure 4 that patient 98 has a sharply increasing pattern of prothrombin index across follow-up time, resulting in higher survival probabilities, conditional on survival at time of final measurement, when compared to Patient 253. Patient 253 maintains lower values of prothrombin index, resulting in lower survival predictions. The reliance of these predictions on accurately specifying the baseline hazard is discussed in Section 5.

4.2. Sensitivity to location and number of knots

In our experience we have found that the default knot locations, based on the distribution of uncensored event times provides the most sensible approach to modelling using spline formulations, as was found in Rizopoulos et al. [14]. This allows the data to be modelled more accurately in the areas of greatest density. Previous work within the flexible parametric survival modelling framework have shown insensitivity to knot placements [27]. Using 5 degrees of freedom (4 internal knots), we have the default knot locations of $\{0.424, 1.186, 2.894, 5.418\}$. We choose 3 other sets of internal knot locations (on the original time scale) and compare parameter estimates and predicted marginal survival curves. We have knot locations 1 of $\{0.3, 1, 3, 5\}$, locations 2 of $\{1, 3, 5, 8\}$ and locations 3 of $\{0.2, 1, 2, 9\}$. Table 5 contains the parameter estimates across models with differing knot choices, illustrating once again the robustness of parameter estimates when compared to the original results in Table 4, with only minor differences observed in the 3rd decimal place. Similarly, the left plot in Figure 5 shows very stable predicted marginal survival curves across knot choices. Furthermore, the right plot in Figure 5 illustrates the fitted marginal survival function when using 2, 3 and 5 internal knots (with locations based on equally spaced quantiles of the distribution of uncensored survival times), illustrating the stability of our proposed model. In comparison to Figure 3, we observe much more variability in the marginal survival predictions when using B-splines with varying number of knots.
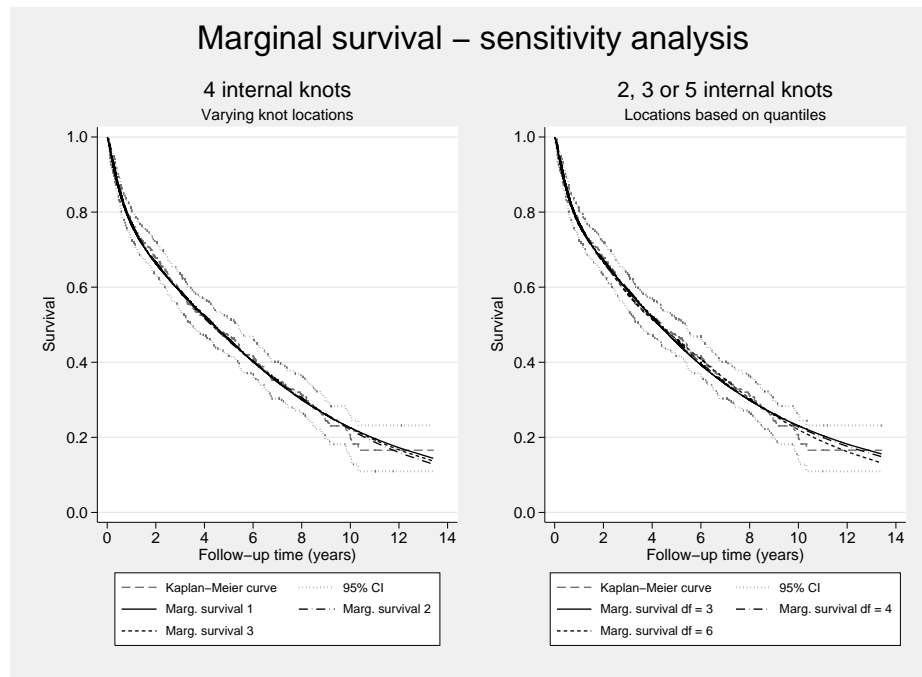
Figure 5. Fitted marginal survival function from joint models with varying knot locations and number of internal knots. Left hand plot used 6 parameters to model the baseline log cumulative hazard function, right hand plot uses 4, 5 and 7 parameters.

5. Discussion

We have described a highly flexible joint model for a single longitudinal continuous biomarker and an event of interest. The restricted cubic spline basis for the log cumulative baseline hazard function provides a flexible framework where often the time-to-event is of primary interest. Flexibility in the longitudinal submodel can be incorporated through the use of fixed and/or random fractional polynomials of time, which can capture a variety of shapes [20].

The simulation study conducted to assess the proposed joint model raised 3 important issues. Firstly, we observed consistent under-estimation of the association parameter, α, under the B-spline approach. This bias was eliminated when using restricted cubic splines, both with 1 and 5 degrees of freedom. Secondly, the choice of the number of quadrature nodes can have a marked impact on both parameter estimates, and in the associated standard errors. If interest is purely on the time-to-event then a lower number of quadrature nodes can be used and will obtain unbiased estimates with optimum coverage levels; however, if the longitudinal submodel is of interest then the choice of quadrature nodes and method is crucial. For example, in studies where quality of life is the longitudinal marker of interest [28], the longitudinal response profile can be of direct interest in order to be included into an economic decision model, where reliable estimates of associated standard errors can be pivotal in assessing cost-effectiveness and thus health policy decisions [29]. The simulation study highlighted the superiority of fully adaptive Gauss-Hermite quadrature in the joint model setting. The use of adaptive quadrature means we can use a much reduced number of quadrature nodes, resulting in substantial computational benefits. Finally, the simulation study showed in general how the estimates of covariate effects were insensitive to the specification of the baseline hazard. This of course can be benefical; however, one of the key benefits of the joint model framework are the predictions which can be obtained. These predictions will rely heavily on the accuracy of the model in estimating the baseline hazard function. We illustrate this in Figure 6, whereby data is simulated under a 2-component mixture Weibull baseline hazard function with a turning point. We apply joint models to the single simulated dataset, firstly with 1 degree of freedom (equivalent to a Weibull model), and then 5 degrees of freedom. We then predict the marginal survival function and compare to the Kaplan-Meier survival curve. It is evident from Figure 6 that only with a sufficient number of degrees of freedom can the baseline survival function be adequately captured.

In application to the liver cirrhosis dataset, it was found that the the restictied cubic spline approach provided improved flexibility in capturing complex baseline hazard functions when compared to a B-spline formulation with the same number of parameters, implying that we can obtain greater flexibility with fewer parameters. Of course, B-spline functions of other degrees may in fact provide well-fitting models; however, our results have shown that
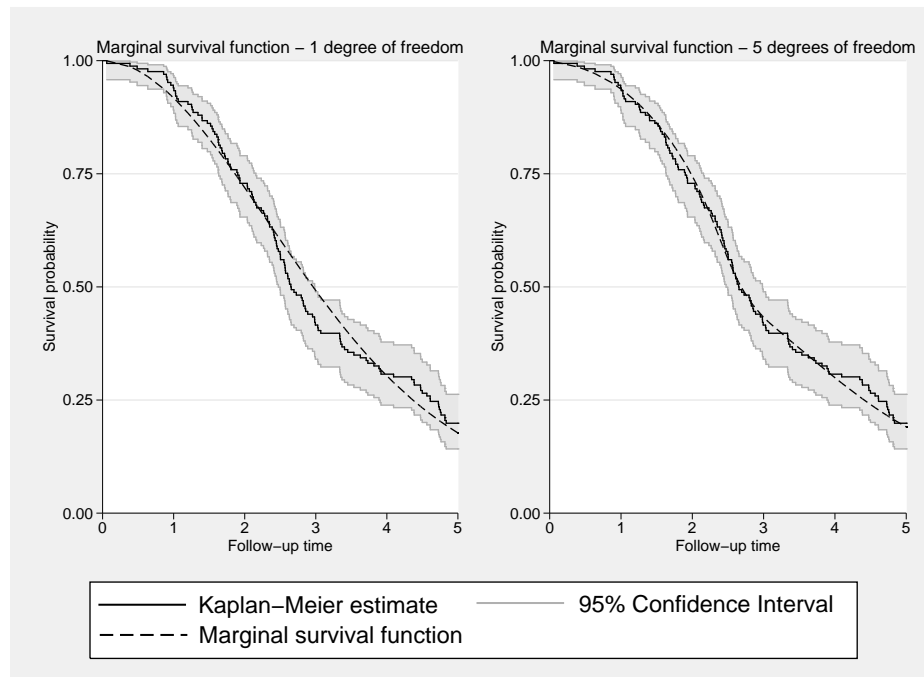
Figure 6. Fitted marginal survival function from joint models with either 1 or 5 degrees of freedom, overlaid on the Kaplan-Meier survival curve.

they can produce unstable fitted functions.

There are a multitude of extensions to this joint model framework. For example, adaption to incorporate a cure fraction can be achieved simply due to the restricted linear basis for the final spline function. Imposing the constraint that the final spline function beyond the last knot is constant has been implemented to allow for a cure fraction in population-based cancer studies within the flexible parametric framework [30]. Furthermore, extension to the competing risks setting by modelling cause-specific hazards can be accomodated, introducing cause-specific association parameters. The generalised linear mixed effects framework for the longitudinal measures submodel can be adapted to handle categorical responses [6]. Finally, a Bayesian approach to the proposed model could be investigated and contrasted [31].

In application to the liver cirrhosis dataset, a single term of observation time provided sufficient flexibility to capture the shape of subject specific longitudinal trajectories; however, further flexibility could be investigated through the use of splines [6, 32].

A reviewer and an associate editor raised concerns about ensuring the monotonicity of the cumulative hazard function. In our experience, including all scenarios of the simulation study, this is not a practical issue. If at any point in the estimation process the hazard function goes negative, then the algorithm will fail. This was not observed in any simulations, ensuring that valid cumulative hazard and subsequently survival functions were estimated.

Implementation of the model is fascilitated through user friendly Stata software [22], developed by the first author. Three choices of association are allowed; namely the current value association discussed above, the first derivative of the longitudinal submodel, and random coefficients such as the random intercept. Both non-adaptive and fully adaptive Gauss-Hermite quadrature is available. A range of other joint models can be fitted, with a variety of extensions under development, including those discussed above.

Acknowledgements

References

1. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**(1):330–339.

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

3. Tseng YK, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. *Biometrika* 2005; **92**(3):587–603.

4. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology* Jun 2010; **28**(16):2796–2801.

5. Li E, Wang N, Wang NY. Joint models for a primary endpoint and multiple longitudinal covariate processes. *Biometrics* Dec 2007; **63**(4):1068–1078.

6. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med* May 2011; **30**(12):1366–1380.

7. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG. Joint modelling of longitudinal and competing risks data. *Stat Med* Dec 2008; **27**(30):6426–6438.

8. Law NJ, Taylor JMG, Sandler H. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* Dec 2002; **3**(4):547–563.

9. Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* Jun 2006; **62**(2):432–445.

10. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**:809–834.

11. Yu M, Law NJ, Taylor JMG, Sandler H. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* 2004; **14**:835–862.

12. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 1995; **4**(1):pp. 12–35.

13. Royston P, Parmar MKB. Flexible Parametric Proportional Hazards and Proportional Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects. *Statistics in Medicine* 2002; **21**(15):2175–2197.

14. Rizopoulos D, Verbeke G, Lesaffre E. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 2009; **71**(3):637–654.

15. Durrleman S, Simon R. Flexible Regression Models with Cubic Splines. *Statistics in Medicine* 1989; **8**(5):551–561.

16. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *The Stata Journal* 2009; **9**:265–290.

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

18. Lesaffre E, Spiessens B. On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2001; **50**(3):325–335.

19. Anderson PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. New York, Springer, 1993.

20. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 1994; **43**(3):429–467.

21. Rabe-Hesketh S, Skrondal A, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* 2002; **2**:1–21.

22. Crowther MJ. STJM: Stata module to fit shared parameter joint models of longitudinal and survival data. Statistical Software Components, Boston College Department of Economics Aug 2012. URL `http://ideas.repec.org/c/boc/bocode/s457339.html`.

23. Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *Journal of Statistical Software* 7 2010; **35**(9):1–33. URL `http://www.jstatsoft.org/v35/i09`.

24. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* 2005; **24**(11):1713–1723.

25. McLachlan GJ, McGiffin DC. On the role of finite mixture models in survival analysis. *Statistical Methods in Medical Research* 1994; **3**(3):211–226.

26. Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press, 2011.

27. Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the Crude Probability of Death due to Cancer and other Causes using Relative Survival Models. *Statistics in Medicine* 2010; **29**(7-8):885–895, doi:10.1002/sim.3762. URL `http://dx.doi.org/10.1002/sim.3762`.

28. Billingham LJ, Abrams KR. Simultaneous analysis of quality of life and survival data. *Stat Methods Med Res* Feb 2002; **11**(1):25–48.

29. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* May 2000; **17**(5):479–500.

30. Andersson TM, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol* 2011; **11**:96.

31. Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* 2004; **58**(1):pp. 16–24. URL `http://www.jstor.org/stable/27643494`.

32. Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* Mar 2005; **61**(1):64–73.

Table 1. Simulation results from Weibull scenario 1. Association is varied with $\alpha = \{-0.25, 0.25\}$.

Parameter	True value	Model	NAQ 5 nodes Bias	NAQ 5 nodes 95% CP	NAQ 15 nodes Bias	NAQ 15 nodes 95% CP	AQ 5 nodes Bias	AQ 5 nodes 95% CP
β_0	0	FPM (df=1)	-0.001	66.5	-0.003	84.4	-0.003	95.6
		FPM (df=5)	-0.005	67.7	-0.002	84.6	-0.003	95.6
		B-spline	0.013	52.4	0.019	81.5	-	-
σ_0	1	FPM (df=1)	-0.046	59.8	-0.002	85.6	-0.004	94.8
		FPM (df=5)	-0.046	60.5	-0.004	85.4	-0.004	94.6
		B-spline	0.012	-	-0.002	-	-	-
β_1	0	FPM (df=1)	-0.014	70.3	-0.002	87.2	-0.001	94.8
		FPM (df=5)	-0.015	69.7	-0.002	86.4	-0.001	94.2
		B-spline	-0.027	56.6	-0.024	73.0	-	-
σ_1	0.25	FPM (df=1)	-0.248	14.6	-0.018	93.0	-0.009	94.2
		FPM (df=5)	-0.247	14.2	-0.021	92.3	-0.009	94.2
		B-spline	0.022	-	-0.028	-	-	-
σ_{01}	0.25	FPM (df=1)	0.001	70.1	-0.020	72.8	0.001	94.8
		FPM (df=5)	0.001	70.5	-0.022	71.9	0.001	94.8
		B-spline	-0.040	-	-0.018	-	-	-
δ	-0.25	FPM (df=1)	-0.017	60.3	-0.004	81.2	0.001	93.8
		FPM (df=5)	-0.012	60.9	-0.009	80.6	0.001	93.8
		B-spline	0.003	49.0	-0.007	80.9	-	-
σ_e	0.5	FPM (df=1)	0.068	30.5	0.000	95.0	-0.001	95.0
		FPM (df=5)	0.068	30.3	0.000	95.0	-0.001	95.0
		B-spline	0.096	-	0.008	-	-	-
ϕ	0.25	FPM (df=1)	0.005	94.0	0.007	94.4	0.007	94.0
		FPM (df=5)	0.002	94.0	0.002	93.8	0.006	93.6
		B-spline	-0.004	95.2	-0.002	94.6	-	-
α	0.25	FPM (df=1)	0.004	95.2	0.005	95.4	0.005	96.0
		FPM (df=5)	-0.002	93.2	-0.003	91.8	0.005	95.2
		B-spline	-0.019	94.8	-0.012	95.6	-	-
β_0	0	FPM (df=1)	0.001	63.2	0.006	82.6	0.002	92.6
		FPM (df=5)	0.000	64.0	0.004	82.4	0.002	92.6
		B-spline	-0.019	56.6	-0.015	80.9	-	-
σ_0	1	FPM (df=1)	-0.048	62.4	-0.004	85.6	-0.008	95.0
		FPM (df=5)	-0.048	62.2	-0.004	84.4	-0.008	95.0
		B-spline	0.009	-	-0.005	-	-	-
β_1	0	FPM (df=1)	0.010	71.2	0.000	88.2	-0.001	95.2
		FPM (df=5)	0.010	72.0	0.001	88.2	-0.001	94.6
		B-spline	0.025	55.8	0.022	72.9	-	-
σ_1	0.25	FPM (df=1)	-0.246	17.6	-0.016	92.8	-0.010	95.2
		FPM (df=5)	-0.245	17.4	-0.012	92.6	-0.009	95.2
		B-spline	0.019	-	-0.029	-	-	-
σ_{01}	0.25	FPM (df=1)	0.001	71.6	0.000	76.6	0.016	96.4
		FPM (df=5)	0.002	71.4	-0.002	76.6	0.015	96.2
		B-spline	-0.031	-	-0.004	-	-	-
δ	-0.25	FPM (df=1)	-0.007	61.2	-0.008	80.4	-0.005	94.8
		FPM (df=5)	-0.008	61.8	-0.010	80.8	-0.005	94.8
		B-spline	-0.012	49.7	-0.012	79.7	-	-
σ_e	0.5	FPM (df=1)	0.069	29.8	0.001	93.0	0.000	94.8
		FPM (df=5)	0.070	29.8	0.001	92.6	0.000	94.8
		B-spline	0.095	-	0.007	-	-	-
ϕ	0.25	FPM (df=1)	0.001	94.0	0.002	94.0	0.002	94.2
		FPM (df=5)	0.002	93.0	-0.002	92.8	0.002	93.8
		B-spline	-0.001	93.7	-0.003	94.0	-	-
α	-0.25	FPM (df=1)	-0.004	96.0	-0.002	96.4	-0.002	96.2
		FPM (df=5)	0.001	95.6	0.008	93.8	-0.001	96.2
		B-spline	0.021	92.1	0.014	94.6	-	-

95% CP - 95% Coverage Probability, df - degrees of freedom, NAQ - Non-Adaptive Quadrature,
AQ - Adaptive Quadrature. FPM - Flexible Parametric Model usng restricted cubic splines

Table 2. Simulation results from mixture-Weibull scenario 2. Association is varied with $\alpha = \{-0.25, 0.25\}$.

Parameter	True value	Model	NAQ 5 nodes Bias	NAQ 5 nodes 95% CP	NAQ 15 nodes Bias	NAQ 15 nodes 95% CP	AQ 5 nodes Bias	AQ 5 nodes 95% CP
β_0	0	FPM (df=1)	-0.007	64.2	0.002	82.1	0.001	94.8
		FPM (df=5)	-0.004	64.8	0.003	80.9	0.002	94.7
		B-spline	0.028	52.5	0.022	80.4	-	-
σ_0	1	FPM (df=1)	-0.050	58.2	-0.002	86.5	-0.004	95.2
		FPM (df=5)	-0.050	58.8	-0.002	86.4	-0.004	95.3
		B-spline	0.011	-	-0.001	-	-	-
β_1	0	FPM (df=1)	-0.012	67.0	-0.001	86.9	0.000	94.8
		FPM (df=5)	-0.013	66.8	-0.001	85.6	-0.001	94.3
		B-spline	-0.024	57.6	-0.022	70.9	-	-
σ_1	0.25	FPM (df=1)	-0.250	11.2	-0.020	93.2	-0.010	94.6
		FPM (df=5)	-0.252	10.4	-0.020	92.9	-0.011	94.3
		B-spline	0.014	-	-0.028	-	-	-
σ_{01}	0.25	FPM (df=1)	-0.007	68.2	-0.008	75.5	0.012	95.6
		FPM (df=5)	-0.012	68.6	-0.007	74.8	0.012	95.5
		B-spline	-0.022	-	-0.002	-	-	-
δ	-0.25	FPM (df=1)	0.001	62.0	-0.007	79.9	-0.005	96.6
		FPM (df=5)	0.000	61.0	-0.008	78.5	-0.005	96.6
		B-spline	-0.005	50.9	-0.006	78.1	-	-
σ_e	0.5	FPM (df=1)	0.068	27.2	0.000	95.0	-0.001	96.4
		FPM (df=5)	0.068	27.2	0.000	95.5	-0.001	96.4
		B-spline	0.101	-	0.006	-	-	-
ϕ	0.25	FPM (df=1)	-0.002	95.4	-0.002	95.6	-0.002	95.4
		FPM (df=5)	-0.003	95.4	-0.004	94.9	-0.003	95.1
		B-spline	-0.011	95.8	-0.012	94.9	-	-
α	0.25	FPM (df=1)	0.001	93.6	0.000	93.4	0.000	93.4
		FPM (df=5)	-0.002	93.0	-0.002	91.7	-0.001	93.3
		B-spline	-0.025	90.1	-0.020	91.7	-	-
β_0	0	FPM (df=1)	-0.010	64.3	0.003	81.1	-0.002	94.4
		FPM (df=5)	-0.011	63.3	-0.001	81.7	-0.002	94.4
		B-spline	-0.040	46.6	-0.019	78.1	-	-
σ_0	1	FPM (df=1)	-0.042	60.7	-0.001	84.5	-0.003	94.8
		FPM (df=5)	-0.043	60.7	-0.001	85.3	-0.003	94.8
		B-spline	0.012	-	0.003	-	-	-
β_1	0	FPM (df=1)	0.009	74.4	0.000	87.6	-0.001	96.4
		FPM (df=5)	0.009	74.8	0.001	88.2	-0.001	96.2
		B-spline	0.023	58.1	0.022	75.1	-	-
σ_1	0.25	FPM (df=1)	-0.247	14.0	-0.018	95.6	-0.008	94.0
		FPM (df=5)	-0.248	13.4	-0.018	95.2	-0.008	94.0
		B-spline	0.017	-	-0.028	-	-	-
σ_{01}	0.25	FPM (df=1)	0.002	67.7	-0.002	69.9	0.016	94.2
		FPM (df=5)	0.003	67.1	-0.003	70.1	0.017	94.4
		B-spline	-0.017	-	-0.002	-	-	-
δ	-0.25	FPM (df=1)	0.014	56.3	-0.002	81.9	0.002	95.0
		FPM (df=5)	0.014	57.7	0.001	81.9	0.002	95.0
		B-spline	0.023	43.8	-0.003	79.7	-	-
σ_e	0.5	FPM (df=1)	0.069	25.7	0.000	92.8	-0.002	93.8
		FPM (df=5)	0.069	25.0	0.000	92.8	-0.001	94.0
		B-spline	0.100	-	0.006	-	-	-
ϕ	0.25	FPM (df=1)	-0.004	94.4	-0.001	94.4	-0.002	94.4
		FPM (df=5)	-0.004	94.2	-0.007	94.2	-0.002	94.8
		B-spline	-0.004	95.2	-0.004	95.0	-	-
α	-0.25	FPM (df=1)	-0.008	96.2	-0.004	96.4	-0.005	96.4
		FPM (df=5)	-0.005	94.0	0.003	93.0	-0.005	95.2
		B-spline	0.020	94.0	0.014	94.2	-	-

95% CP - 95% Coverage Probability, df - degrees of freedom, NAQ - Non-Adaptive Quadrature, AQ - Adaptive Quadrature. FPM - Flexible Parametric Model usng restricted cubic splines

Table 3. Simulation results from Weibull scenario 3. Association is varied with $\alpha = \{-0.25, 0.25\}$.

Parameter	True value	Model	NAQ 5 nodes		NAQ 15 nodes		AQ 5 nodes	
			Bias	95% CP	Bias	95% CP	Bias	95% CP
β_0	0	FPM (df=1)	0.006	57.3	-0.001	78.5	-0.001	95.0
		FPM (df=5)	0.009	57.1	0.000	78.0	-0.005	94.9
		B-spline	-0.002	42.7	0.006	74.6	-	-
σ_0	1	FPM (df=1)	-0.063	49.9	-0.010	79.5	-0.010	96.2
		FPM (df=5)	-0.063	50.3	-0.010	79.4	-0.011	95.6
		B-spline	0.006	-	-0.006	-	-	-
β_1	0	FPM (df=1)	0.000	79.8	0.001	85.7	0.000	94.6
		FPM (df=5)	0.001	78.2	0.001	85.3	-0.002	94.1
		B-spline	0.000	68.1	0.000	83.0	-	-
σ_1	0.25	FPM (df=1)	-0.217	6.6	-0.016	92.6	-0.003	93.4
		FPM (df=5)	-0.218	6.8	-0.016	92.3	-0.005	93.9
		B-spline	0.012	-	-0.004	-	-	-
σ_{01}	0.25	FPM (df=1)	-0.046	20.6	-0.020	68.5	0.011	95.2
		FPM (df=5)	-0.046	20.4	-0.023	69.7	0.009	94.7
		B-spline	0.040	-	0.012	-	-	-
δ	-0.25	FPM (df=1)	0.006	95.8	0.008	77.3	0.000	96.6
		FPM (df=5)	-0.012	94.8	0.007	76.6	-0.001	96.6
		B-spline	0.013	35.3	0.000	71.3	-	-
σ_e	0.5	FPM (df=1)	0.073	9.8	0.003	93.0	0.000	93.8
		FPM (df=5)	0.073	10.0	0.003	92.9	0.000	94.3
		B-spline	0.133	-	0.010	-	-	-
ϕ	0.25	FPM (df=1)	0.006	95.8	0.007	95.6	0.006	95.4
		FPM (df=5)	-0.012	94.8	0.006	95.6	-0.012	95.1
		B-spline	-0.005	95.4	-0.002	95.8	-	-
α	0.25	FPM (df=1)	0.010	94.0	0.004	94.6	0.004	94.4
		FPM (df=5)	0.001	93.2	0.001	93.7	-0.016	90.7
		B-spline	-0.025	96.0	-0.018	96.4	-	-
β_0	0	FPM (df=1)	0.008	56.6	0.006	73.4	0.003	94.8
		FPM (df=5)	0.009	56.8	0.006	74.0	0.005	94.8
		B-spline	0.009	39.1	0.002	70.4	-	-
σ_0	1	FPM (df=1)	-0.061	45.6	-0.009	75.8	-0.006	95.8
		FPM (df=5)	-0.061	45.8	-0.009	76.2	-0.007	95.2
		B-spline	0.011	-	-0.003	-	-	-
β_1	0	FPM (df=1)	0.000	80.0	0.000	87.6	-0.001	95.2
		FPM (df=5)	0.000	80.4	0.000	88.2	0.000	95.4
		B-spline	0.000	70.9	0.000	87.8	-	-
σ_1	0.25	FPM (df=1)	-0.219	5.2	-0.019	93.6	-0.004	94.3
		FPM (df=5)	-0.219	4.8	-0.018	93.2	-0.005	93.6
		B-spline	0.010	-	-0.005	-	-	-
σ_{01}	0.25	FPM (df=1)	-0.036	23.4	-0.021	64.5	0.010	94.1
		FPM (df=5)	-0.034	23.4	-0.020	64.8	0.009	94.4
		B-spline	0.037	-	0.009	-	-	-
δ	-0.25	FPM (df=1)	-0.011	52.6	-0.015	74.6	-0.009	94.8
		FPM (df=5)	-0.010	52.6	-0.015	75.4	-0.009	94.8
		B-spline	-0.023	36.3	-0.015	68.8	-	-
σ_e	0.5	FPM (df=1)	0.073	8.6	0.002	93.4	0.000	94.3
		FPM (df=5)	0.074	8.2	0.003	93.6	-0.001	94.4
		B-spline	0.133	-	0.010	-	-	-
ϕ	0.25	FPM (df=1)	-0.006	95.2	-0.003	95.2	-0.003	95.2
		FPM (df=5)	-0.010	95.4	-0.009	94.8	-0.019	94.2
		B-spline	0.000	95.4	-0.001	95.4	-	-
α	-0.25	FPM (df=1)	-0.007	95.0	-0.001	95.0	0.000	94.8
		FPM (df=5)	0.002	93.4	0.000	94.2	0.013	92.0
		B-spline	0.028	94.2	0.021	95.0	-	-

95% CP - 95% Coverage Probability, df - degrees of freedom, NAQ - Non-Adaptive Quadrature,
AQ - Adaptive Quadrature. FPM - Flexible Parametric Model usng restricted cubic splines

Table 4. Results from applying the two-stage and full joint models to the PBC dataset.

Parameter	Joint model: RCS			Joint model: B-spline		
	Estimate	95% CI		Estimate	95% CI	
Longitudinal:						
t1	0.872	0.388	1.356	1.169	0.694	1.643
t1*treatment	0.272	-0.354	0.899	0.299	-0.327	0.925
Intercept	75.252	73.033	77.471	75.901	73.745	78.056
sd(t1)	2.333	1.953	2.786	2.176	-	-
sd(intercept)	21.839	20.152	23.667	21.327	-	-
corr(t1,intercept)	0.610	0.467	0.722	0.567	-	-
sd(Residual)	17.612	17.092	18.148	17.658	-	-
Survival:						
Association	-0.038	-0.045	-0.031	-0.039	-0.046	-0.031
Treatment	0.210	-0.038	0.457	0.241	-0.007	0.489

where t1 = log(time + 0.00273)
RCS - Restricted Cubic Splines, CI - Confidence Interval

Table 5. Results from joint models with varying knot locations.

Parameter	Knot locations 1			Knot locations 2			Knot locations 3		
	Estimate	95% CI		Estimate	95% CI		Estimate	95% CI	
Longitudinal:									
t1	0.877	0.393	1.361	0.873	0.390	1.357	0.874	0.391	1.358
t1*treatment	0.275	-0.351	0.902	0.275	-0.352	0.901	0.273	-0.353	0.900
Intercept	75.304	73.088	77.519	75.261	73.044	77.479	75.262	73.045	77.481
sd(t1)	2.333	1.954	2.786	2.331	1.952	2.784	2.331	1.951	2.784
sd(intercept)	21.843	20.132	23.700	21.830	20.143	23.658	21.829	20.142	23.657
corr(t1,intercept)	0.609	0.466	0.721	0.609	0.466	0.722	0.609	0.466	0.722
sd(Residual)	17.611	17.091	18.147	17.613	17.093	18.148	17.612	17.093	18.148
Survival:									
Association	-0.038	-0.045	-0.031	-0.038	-0.045	-0.031	-0.038	-0.045	-0.031
Treatment	0.209	-0.038	0.456	0.212	-0.035	0.459	0.210	-0.037	0.457