Flexible parametric joint modelling of longitudinal and survival data

Michael J. Crowther, Keith R. Abrams and Paul C. Lambert

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. We conduct an extensive simulation study to assess the proposed model, comparing it with 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. We also describe a useful technique to simulate survival times from complex baseline hazard functions and illustrate the methods using an example data set investigating the association between longitudinal prothrombin index and survival of patients with liver cirrhosis, showing greater flexibility and improved stability with fewer parameters under the proposed model compared with the B-spline approach. We provide user-friendly Stata software. Copyright © 2012 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]. Tsiatis and Davidian [10] and Yu et al. [11] describe extensive outlines of the field.

The form of joint model, which has dominated the literature, assumes that shared random effects characterise the association between the time-to-event and longitudinal marker, and we adopt this approach. In 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 two 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 we can obtain smooth continuous time predictions and tailor them at the individual level, allowing out-of-sample predictions to be made.

The estimation of joint models has almost exclusively been implemented using expectation–maximisation 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. We adopt the second approach to fit the models. As has been discussed in Rizopoulos et al. [14], we can evaluate the score equations 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 with 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 evaluate not only the proposed joint model, but also the non-adaptive quadrature with varying numbers of nodes, and fully adaptive quadrature.

We illustrate the methods using a data set 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 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.

We organise the remainder of the paper as follows: 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 data set. We conclude the paper in Section 5 with a discussion.

2. Defining the joint model

For the $i$th 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 $y_{ij} = \{y_i(t_{ij})\}$ denote the $j$th longitudinal response measurement of a continuous biomarker for the $i$th 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. We assume both censoring and time of measurements to be non-informative.
2.1. Longitudinal submodel

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

\[ y_i(t_{ij}) = W_i(t_{ij}) + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2_e) \]  \hspace{1cm} (1)

\[ W_i(t_{ij}) = X_i'(t_{ij}) \beta + Z_i'(t_{ij}) b_i + u_i \delta \]  \hspace{1cm} (2)

with design matrices \( X_i \) and \( Z_i \) for the fixed (\( \beta \)) 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 corresponding regression coefficients, \( \delta \). We assume the error is independent from the random effects and that \( \text{cov}(e_{ij}, e_{ik}) = 0 \) (where \( j \neq k \)). We can incorporate flexibility in the longitudinal submodel 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:

\[ \log\{H(t|b_i, v_i)\} = \log\{H_0(t)\} + \alpha W_i(t_{ij}) + \phi v_i \]  \hspace{1cm} (3)

where \( H_0(t) \) is the cumulative baseline hazard function, \( \alpha \) denotes the association parameter and \( \phi \) 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 is found in Rizopoulos and Ghosh [6].

We derive the spline basis for this specification 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 \( k_0 \), as \( s(\log(t)|y, k_0) \). For example, with \( K \) knots and letting \( x = \log(t) \), we can express a restricted cubic spline function as

\[ s(x) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \cdots + \gamma_{K-1} z_{K-1} \]  \hspace{1cm} (4)

and

\[ z_1 = x \]  \hspace{1cm} (5)

\[ z_p = (x - k_p)^3 + \kappa_p (x - k_p)^3 - (1 - \kappa_p) (x - k_K)^3 \quad p = 2, \ldots, K - 1 \]  \hspace{1cm} (6)

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

\[ \log\{H(t|b_i, v_i)\} = \eta_i = s(\log(t)|y, k_0) + \alpha W_i(t) + \phi v_i \]  \hspace{1cm} (7)

We can now transform to the hazard and survival scales

\[ h(t|b_i, v_i) = \left\{ \frac{1}{t} \frac{dN(t|y, k_0)}{d \log(t)} + \alpha \frac{dW(t)}{dr} \right\} \exp(\eta_i), \quad S(t|b_i, v_i) = \exp\{-\exp(\eta_i)\} \]  \hspace{1cm} (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:

\[ \prod_{i=1}^n \left[ \int_{-\infty}^{\infty} \left( \prod_{j=1}^{m_i} f(y_i(t_{ij})|b_i, \theta) \right) f(b_1|\theta, d_i|b_1, \theta) \, db_1 \right] \]  \hspace{1cm} (9)
where

\[
    f(y_i(t_{ij})|b_i, \theta) = (2\pi \sigma^2)^{-1/2} \exp \left\{-\frac{y_i(t_{ij}) - W_i(t_{ij})}{2\sigma^2} \right\}
\]

(10)

and

\[
    f(b_i|\theta) = (2\pi |V|)^{-1/2} \exp \left\{-\frac{b_i'V^{-1}b_i}{2} \right\}
\]

(11)

with parameter vector \( \theta \) and \( V \) the variance–covariance matrix of the random effects. We can evaluate Equation (9) 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

We undertook a simulation study to assess the performance of the proposed joint model for finite sample sizes, comparing with the model of Rizopoulos et al. [14]. Under each scenario, we apply both the proposed joint model with five degrees of freedom, plus an intercept, resulting in six parameters to capture the baseline hazard and sufficient coverage and \( R \) using 1.5E-8. We fixed time of measurements at \((0, 1, 2, 3, 4)\). We also generated survival times from an exponential distribution (however, the primary motivation of the simulation study was to evaluate the Laplacian estimation method, not the survival submodel).

For each scenario, we included 300 patients in each 500 repetitions. We generated the true longitudinal profile from \( W_i = \beta_{0i} + \beta_{1i}t_{ij} + \delta u_i \), with \( \beta_{0i} \sim N(0, 1), \beta_{1i} \sim N(0, 0.25^2) \) and correlation between \((\beta_{0i}, \beta_{1i})\) of 0.25. We then generated the observed longitudinal measurements from \( Y_{ij} \sim N(W_{ij}, 0.5^2) \). We fixed time of measurements at \((0, 1, 2, 3, 4)\). We also generated survival times from \( \log(H(t|b_i)) = \log(H_0(t)) + \alpha W_i(t) + \phi u_i \), where we detail \( H_0(t) \) subsequently. We applied censoring at 5 years. We generated a binary treatment group variable from \( u_i \sim \text{Bin}(1, 0.5) \). We fix the direct treatment effect on the longitudinal response, \( \delta \), at \(-0.25\), the direct treatment effect on the time-to-event, \( \phi \), at 0.25, and vary the association parameter, \( \alpha \), 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. 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 that 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 two-component mixture Weibull distribution [25], with

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

where \(0 \leq p \leq 1\), \(\lambda_1\) and \(\lambda_2\) are scale parameters, \(\gamma_1\) and \(\gamma_2\) are shape parameters, and

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

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

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

Following the formulation of Bender et al. [24]

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

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

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

Equation (18) is analytically intractable and so cannot be directly rearranged to find \(t\); however, methods to overcome this include Newton–Raphson iterations or nonlinear 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 to overcome this include Newton–Raphson iterations or nonlinear 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.

We chose three scenarios of baseline parameters: 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 I–III 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, that is the direct treatment effect on survival (\(\phi\)) and the association parameter (\(\alpha\)), appear to be

| Table I. Simulation results from Weibull scenario 1. |
|----------------------------------|------------------|------------------|------------------|
| Parameter | True value | Model | NAQ 5 nodes | NAQ 15 nodes | AQ 5 nodes |
|-----------|------------|-------|-------------|-------------|-------------|
| \(\beta_0\) | 0 | FP (df=1) | \(-0.001\) | 66.5 | \(-0.003\) | 84.4 | \(-0.003\) | 95.6 |
| \(\sigma_0\) | 1 | FP (df=1) | \(-0.046\) | 59.8 | \(-0.002\) | 85.6 | \(-0.004\) | 94.8 |
| \(\beta_1\) | 0 | FP (df=1) | \(-0.014\) | 70.3 | \(-0.002\) | 87.2 | \(-0.001\) | 94.8 |
| \(\sigma_1\) | 0.25 | FP (df=1) | \(-0.248\) | 14.6 | \(-0.018\) | 93.0 | \(-0.009\) | 94.2 |
| \(\sigma_{01}\) | 0.25 | FP (df=1) | \(-0.247\) | 14.2 | \(-0.021\) | 92.3 | \(-0.009\) | 94.2 |
Table I. Continued.

| Parameter | True value | Model       | NAQ 5 nodes |             | NAQ 15 nodes |             | AQ 5 nodes |             |
|-----------|------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|
|           |            |             | Bias  | 95% CP | Bias  | 95% CP | Bias  | 95% CP |
| 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     |        |          |
| σe       | 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    | 0.009   |          | 0.008   |          |        |          |
| φ        | 0.25       | FPM (df=1)  | 0.002   | 94.0     | 0.002   | 93.8     | 0.006   | 93.6     |
|          |            | FPM (df=5)  |        |          | 0.002   | 91.8     | 0.006   | 93.6     |
|          |            | B-spline    | 0.012   | 49.7     | 0.012   | 79.7     | 0.012   | 79.7     |
| α        | 0.25       | FPM (df=1)  | 0.001   | 94.8     | −0.012  | 95.6     |        |          |
|          |            | FPM (df=5)  | 0.001   | 94.8     | −0.012  | 95.6     |        |          |
| β0       | 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     |        |          |
| σ0       | 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  |          |        |          |
| β1       | 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     |        |          |
| σ1       | 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  |          |        |          |
| σ01      | 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     |        |          |
| σe       | 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     |        |          |

Association is varied with \( \alpha = \{-0.25, 0.25\}\).

95% CP. 95% coverage probability; df, degrees of freedom; NAQ, non-adaptive quadrature; AQ, adaptive quadrature; FPM, flexible parametric model using restricted cubic splines.

Table II. Simulation results from mixture-Weibull scenario 2.

| Parameter | True value | Model       | NAQ 5 nodes |             | NAQ 15 nodes |             | AQ 5 nodes |             |
|-----------|------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|
|           |            |             | Bias  | 95% CP | Bias  | 95% CP | Bias  | 95% CP |
| β0        | 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     |        |          |
| σ0        | 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  |          |          |        |          |
Table II. Continued.

| Parameter | True value | Model         | NAQ 5 nodes | NAQ 15 nodes | AQ 5 nodes |
|-----------|------------|---------------|-------------|--------------|------------|
|           |            |               | Bias   | 95% CP | Bias   | 95% CP | Bias   | 95% CP |
| $\beta_1$ | 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  | −      | −      |
| $\sigma_1$ | 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 | −      | −      | −      |
| $\sigma_{01}$ | 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 | −      | −      | −      |
| $\delta$ | −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  | −      | −      |
| $\sigma_e$ | 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  | −      | −      | −      |
| $\phi$   | 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  | −      | −      |
| $\alpha$ | 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  | −      | −      |
| $\beta_0$ | 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  | −      | −      |
| $\sigma_0$ | 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  | −      | −      | −      |
| $\beta_1$ | 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  | −      | −      |
| $\sigma_1$ | 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 | −      | −      | −      |
| $\sigma_{01}$ | 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 | −      | −      | −      |
| $\delta$ | −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  | −      | −      |
| $\sigma_e$ | 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  | −      | −      | −      |
| $\phi$   | 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 | 94.2  |
|           |            | B-spline      | −0.004 | 95.2  | −0.004 | 95.0  | −      | −      |
| $\alpha$ | −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  | −      | −      |

Association is varied with $\alpha = \{-0.25, 0.25\}$.

95% CP, 95% coverage probability; df, degrees of freedom; NAQ, non-adaptive quadrature; AQ, adaptive quadrature; FPM, flexible parametric model using restricted cubic splines.

unbiased. However, under the B-spline approach, across all scenarios, we observe consistent underestimation of the association parameter, $\alpha$. 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 with −10% under the B-spline approach. Coverage probabilities
Table III. Simulation results from Weibull scenario 3.

| Parameter | True value | Model | NAQ 5 nodes | NAQ 15 nodes | AQ 5 nodes |
|-----------|------------|-------|-------------|--------------|------------|
|           |            |       | Bias        | 95% CP       | Bias        | 95% CP       | Bias        | 95% CP       |
| $\beta_0$ | 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        | -         | -           |
|           |            | B-spline | 0.006 | -          | -0.006 | -         | -         | -           |
| $\sigma_0$ | 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 | -         | -         | -           |
| $\beta_1$ | 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        | -         | -           |
|           |            | B-spline | 0.040 | -          | 0.012  | -         | -         | -           |
| $\sigma_1$ | 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 | -         | -         | -           |
|           |            | B-spline | 0.046 | 20.6       | -0.020 | 68.5        | 0.011  | 95.2        |
| $\sigma_{01}$ | 0.25     | FPM (df=1) | -0.046 | 20.4       | -0.023 | 69.7        | 0.009  | 94.7        |
|           |            | FPM (df=5) | -0.046 | 20.4       | -0.023 | 69.7        | 0.009  | 94.7        |
| $\delta$ | -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        | -         | -           |
|           |            | B-spline | 0.133 | -          | 0.010  | -         | -         | -           |
| $\phi$ | 0.25       | 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 | 0.013 | 35.3       | 0.000 | 71.3        | -         | -           |
|           |            | B-spline | 0.133 | -          | 0.010  | -         | -         | -           |
| $\sigma_e$ | 0.5        | 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        | -         | -           |
| $\alpha$ | 0.25       | FPM (df=1) | 0.010 | 94.0        | 0.004 | 94.6        | 0.004  | 94.4        |
|           |            | FPM (df=5) | 0.010 | 93.2       | 0.001 | 93.7        | -0.016 | 90.7        |
|           |            | B-spline | -0.025 | 96.0       | -0.018 | 96.4        | -         | -           |

Association is varied with $\alpha = \{-0.25, 0.25\}$.  
95% CP, 95% coverage probability; df, degrees of freedom; NAQ, non-adaptive quadrature; AQ, adaptive quadrature; FPM, flexible parametric model using restricted cubic splines.
very closely approximate 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 $\geq 15$ 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 we use five-point non-adaptive quadrature; however, we eliminate this bias 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 II presents bias and coverage estimates for simulations generated from a two-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 we use an insufficient number of quadrature nodes. 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 five degrees 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 data set

In this section, we apply the proposed joint model to the data set 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). Two hundred ninety-two (59.8%) patients died during the study. Patients were randomised to two treatment groups, namely prednisone and placebo. For further details regarding the data set, 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 we adjust the timescale by taking away the observed censoring/event time. We overlay a lowess smoother. From Figure 1, it is apparent that patients who experienced the event, compared with patients who were censored, had decreasing levels of the biomarker during the 2–3-year period before death. If we assume that 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 data set. 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 with 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 five degrees of freedom to model the baseline cumulative hazard, equivalent to four 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. Under the B-spline model, we use cubic splines with two 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 IV, 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 data set, shown in Figure 2. It is evident from Figure 2 that the restricted cubic spline approach provides an improved fit compared with 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 five internal knots to achieve a function that fits as closely as the restricted cubic splines approach. In other words, we need to use nine parameters under

| Table IV. Results from applying the two-stage and full joint models to the liver cirrhosis data set. |
|---------------------------------------------------|----------------|----------------|----------------|----------------|
| Parameter                                    | RCS Estimate | 95% CI         | B-spline Estimate | 95% CI         |
| Longitudinal:                                |               |                |                 |                |
| t1                                           | 0.872         | 0.388          | 1.356           |                |
| t1*treatment                                 | 0.272         | -0.354         | 0.899           |                |
| Intercept                                    | 75.252        | 73.033         | 77.471          |                |
| sd(t1)                                       | 2.333         | 1.953          | 2.786           |                |
| sd(intercept)                                | 21.839        | 20.152         | 23.667          |                |
| corr(t1,intercept)                           | 0.610         | 0.467          | 0.722           |                |
| sd(Residual)                                 | 17.612        | 17.092         | 18.148          |                |
| Survival:                                    |               |                |                 |                |
| Association                                  | -0.038        | -0.045         | -0.031          | -0.039         |
| Treatment                                    | 0.210         | -0.038         | 0.457           | 0.241          |

t1 = log(time + 0.00273)

RCS, restricted cubic splines; CI, confidence interval.
Figure 2. Marginal survival using six parameters to model the baseline cumulative hazard function, overlaid on the Kaplan–Meier estimate and its associated 95% confidence interval.

Figure 3. Marginal survival using cubic B-splines with three, four or five internal knots, resulting in seven, eight or nine parameters to model the baseline cumulative hazard function, overlaid on the Kaplan–Meier estimate and its associated 95% confidence interval.

4.1. Predictions

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

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 with patient 253. Patient 253 maintains lower values of prothrombin index, resulting in lower survival predictions. We discuss the reliance of these predictions on accurately specifying the baseline hazard 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

Table V. 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 − 1.357    | 0.874 − 1.358      |                  |                  |
| t1*treatment    | 0.275 − 0.351| 0.902        | 0.275 − 0.352    | 0.273 − 0.353      | 0.273 − 0.352    | 0.900 − 0.900    |
| Intercept       | 75.304 − 73.088| 77.519        | 75.261 − 73.044  | 77.479 − 73.045    | 75.262 − 73.045  | 77.481 − 77.481  |
| sd(t1)          | 2.333 − 2.786| 1.954        | 2.331 − 2.784    | 2.331 − 2.784      | 2.331 − 2.784    | 2.784 − 2.784    |
| sd(intercept)   | 21.843 − 23.700| 20.132        | 21.830 − 20.143  | 21.829 − 20.142    | 21.829 − 20.142  | 23.657 − 23.657  |
| corr(t1,intercept) | 0.609 − 0.721| 0.466        | 0.609 − 0.722    | 0.609 − 0.722      | 0.609 − 0.722    | 0.722 − 0.722    |
| sd(Residual)    | 17.611 − 18.147| 17.091        | 17.613 − 18.148  | 17.612 − 18.148    | 17.612 − 18.148  | 18.148 − 18.148  |
| Survival:       |              |              |                  |                  |                  |                  |
| Association     | −0.038 − 0.045| −0.031       | −0.038 − 0.045   | −0.038 − 0.045     | −0.038 − 0.045   | −0.031 − 0.045   |
| Treatment       | 0.209 − 0.456| 0.456        | 0.212 − 0.459    | 0.210 − 0.457      | 0.210 − 0.457    | 0.457 − 0.457   |
density. Previous work within the flexible parametric survival modelling framework have shown insensitivity to knot placements [27]. Using five degrees of freedom (four internal knots), we have the default knot locations of \{0.424, 1.186, 2.894, 5.418\}. We choose three 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 V contains the parameter estimates across models with differing knot choices, illustrating once again the robustness of parameter estimates when compared with the original results in Table IV, with only minor differences observed in the third 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 two, three and five 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 with Figure 3, we observe much more variability in the marginal survival predictions when using B-splines with varying number of knots.

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. We can incorporate flexibility in the longitudinal submodel 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 three important issues. Firstly, we observed consistent underestimation of the association parameter, \(\alpha\), under the B-spline approach. We eliminated this bias when using restricted cubic splines, both with one and five 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 we can use a lower number of quadrature nodes and 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

![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.](image-url)
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 beneficial; 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 are simulated under a two-component mixture Weibull baseline hazard function with a turning point. We apply joint models to the single simulated data set, firstly with one degree of freedom (equivalent to a Weibull model) and then five degrees of freedom. We then predict the marginal survival function and compare with 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 data set, we found that the restricted cubic spline approach provided improved flexibility in capturing complex baseline hazard functions when compared with 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 they can produce unstable fitted functions.

There are a multitude of extensions to this joint model framework. For example, we can incorporate a cure fraction simply because of 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 accommodated, introducing cause-specific association parameters. We can adapt the generalised linear mixed effects framework for the longitudinal measures submodel to handle categorical responses [6]. Finally, we could investigate and contrast a Bayesian approach to the proposed model [31].

In application to the liver cirrhosis data set, a single term of observation time provided sufficient flexibility to capture the shape of subject specific longitudinal trajectories; however, we could investigate further flexibility 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

![Figure 6. Fitted marginal survival function from joint models with either one or five degrees of freedom, overlaid on the Kaplan–Meier survival curve.](image-url)
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 we estimated valid cumulative hazard and subsequently survival functions.

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

Acknowledgements

The authors would like to thank two anonymous reviewers and an associate editor for their comments and suggestions that greatly improved the manuscript. MJC was funded by a National Institute for Health Research (NIHR) Methods Fellowship (RP-PG-0407-10314) and KRA is partially supported as an NIHR Senior Investigator (NI-51-0508-10061).

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 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 2010; 28(16):2796–2801.
5. Li E, Wang N, Wang NY. Joint models for a primary endpoint and multiple longitudinal covariate processes. Biometrics 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. Statistics in Medicine 2011; 30(12):1366–1380.
7. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG. Joint modelling of longitudinal and competing risks data. Statistics in Medicine 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 2002; 3(4):547–563.
9. Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. Biometrics 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):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 2011; 67(3):819–829.
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. Springer: New York, 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, August 2012.
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.
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 GI, 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: College Station, Texas, 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.
28. Billingham LJ, Abrams KR. Simultaneous analysis of quality of life and survival data. *Statistical Methods in Medical Research* 2002; 11(1):25–48.
29. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 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 Medical Research Methodology* 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):16–24.
32. Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* 2005; 61(1):64–73.