ANNUAL REVIEWS

Annual Review of Statistics and Its Application

An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes

Grigorios Papageorgiou, Katya Mauff, Anirudh Tomer, and Dimitris Rizopoulos

Department of Biostatistics, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands; email: d.rizopoulos@erasmusmc.nl

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Stat. Appl. 2019. 6:223–40

First published as a Review in Advance on August 15, 2018

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-030718-105048

Copyright © 2019 by Annual Reviews. All rights reserved

All authors contributed equally to this article

Keywords

joint modeling, longitudinal, survival, dynamic predictions, association

Abstract

In this review, we present an overview of joint models for longitudinal and time-to-event data. We introduce a generalized formulation for the joint model that incorporates multiple longitudinal outcomes of varying types. We focus on extensions for the parametrization of the association structure that links the longitudinal and time-to-event outcomes, estimation techniques, and dynamic predictions. We also outline the software available for the application of these models.

1. INTRODUCTION

This article presents an overview of the suite of modeling techniques known as joint models for longitudinal and time-to-event data. These models are applicable in settings where subjects are followed over time, usually to monitor the progress of a disease or medical condition. That progression is typically evaluated via repeated measurements of a biomarker or biomarkers pertinent to the disease, and it may be of clinical interest to determine the effect of such a biomarker on the time to an event of interest (e.g., death or intervention). Since measurements of the biomarkers are taken from the subject under study, they are deemed endogenous; that is, their value at any given time point is dependent upon/may be altered by the occurrence of the event prior to that time point. They are also usually measured with error, and their complete path is unknown: Their value is only known for the specific time points at which they are measured. Standard methods for the analysis of a time-to-event outcome with a time-varying covariate, such as the extended Cox model, which assumes that the values of the time-varying covariate are constant in between measurements, therefore result in biased estimates and standard errors (Prentice 1982) when used in this context. In these same follow-up studies, it is inevitable that we should have missing data when subjects drop out or do not adhere to the scheduled visiting times. When the probability of a subject dropping out depends on their unobserved longitudinal measurements, this dropout process is defined as nonrandom or informative. This process cannot then be ignored, and valid inferences may only be made based on a joint distribution of the longitudinal measurements and the missingness process (Rizopoulos 2012b). In both cases (endogeneity and informative missingness), the joint modeling framework provides us with a solution by postulating a relative risk model for the time-to-event outcome, dependent on the true underlying value of the longitudinal outcome, which is modeled using a mixed effects model, and where estimation is based on the joint distribution of the two outcomes.

Although the most basic formulation of the model focuses on the simple association between a single time-to-event outcome and a single continuous longitudinal outcome, extensions are manifold. Multiple longitudinal outcomes (Lin et al. 2002, Brown et al. 2005, Chi & Ibrahim 2006, Rizopoulos & Ghosh 2011) and multiple recurrent or competing events (Elashoff et al. 2008, Williamson et al. 2008, Hu et al. 2009, Huang et al. 2011) are easily incorporated, as are more complex association structures (Rizopoulos 2012b, 2016; Andrinopoulou et al. 2014; Mauff et al. 2017). These models are therefore well suited to the analysis of the varying and often complex disease dynamics under study. One use of joint models that is of particular clinical interest is personalized predictive modeling, and the development of effective and personalized screening methods (Garre et al. 2008, Yu et al. 2008, Proust-Lima & Taylor 2009, Rizopoulos 2011, Tomer et al. 2017, Papageorgiou et al. 2018).

Prompted by the early work of Self & Patiwan (1992) and De Gruttola & Tu (1994) and the formative papers by Faucett & Thomas (1996) and Wulfsohn & Tsiatis (1997), there is already a substantial body of work on the topic of joint modeling, including several extensive reviews conducted by Hogan & Laird (1997), Tsiatis & Davidian (2004), and Gould et al. (2015). In the interest of brevity, in this article, we present only the shared parameter formulation of joint models, and models where the primary interest is in survival. Our main focus is on developments in dynamic predictions and those pertaining to the parameterization of the association structure. We start with a description of the primary biliary cirrhosis (PBC) data set in Section 2, which is used as an illustrative example throughout the article. Section 3 looks at the joint model, incorporating some extensions in the initial formulation, and Section 3.1 presents a detailed discussion of various association structures. Section 4 deals with estimation approaches, and Section 5 addresses dynamic predictions and measures of predictive performance. Finally, in Section 6, we outline available software.

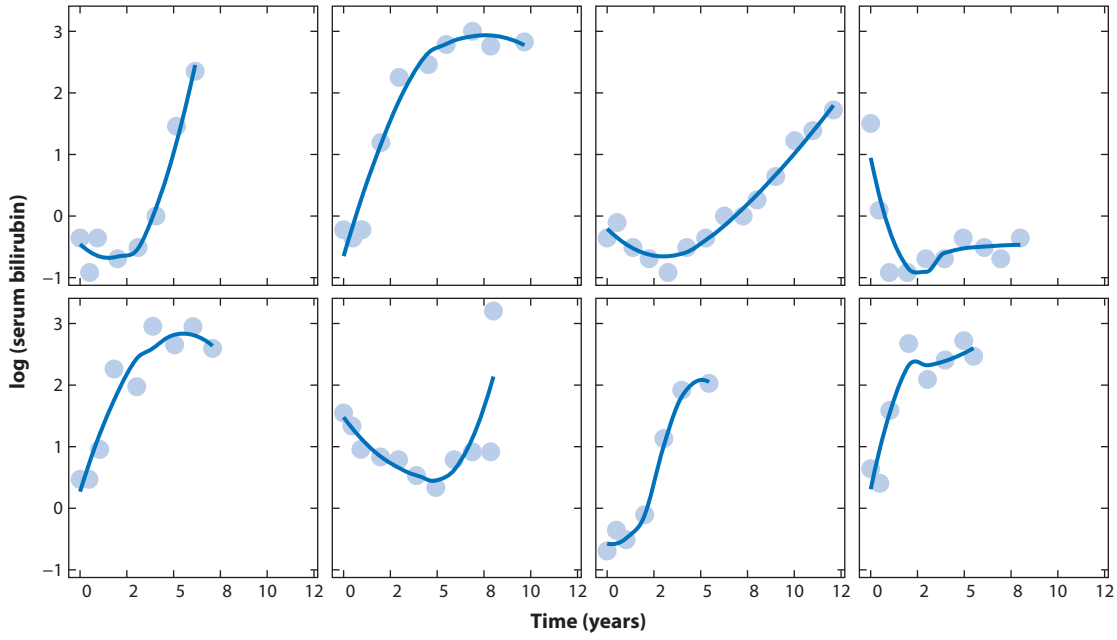

Figure 1

Longitudinal profiles of log(serum bilirubin) over time for eight randomly selected subjects from the primary biliary cirrhosis data set.

2. EXAMPLE DATA

PBC is a chronic liver disease that leads to cirrhosis and eventually death. The PBC data set (Murtaugh et al. 1994) comprises information on 312 subjects who participated in a 10-year study conducted by the Mayo Clinic. The primary outcome of interest was patient survival, and whether this was prolonged with treatment with D-penicillamine ($n = 156$) compared with a placebo ($n = 154$). This is an ideal data set to illustrate the various features of the joint model. Although it is usually analyzed as a composite, the time-to-event outcome consists of two competing events (death and transplantation). Follow-up measurements (scheduled at baseline, six months, and annually thereafter) were obtained for multiple biomarkers of varying types, such as serum bilirubin, serum cholesterol, and the presence or absence of hepatomegaly (an enlarged liver). We also note the distinct nonlinearity of the subject-specific log(serum bilirubin) profiles depicted for a randomly selected subset of patients in **Figure 1**.

3. MODEL FORMULATION

Conceptually, the joint model links a survival model with a suitable model for the endogenous longitudinal covariate. Time-varying covariates such as serum bilirubin are defined as endogenous if their value at any time point t may be altered by or depends upon the occurrence of an event at some time point $s < t$ (e.g., in the case of transplantation or death, respectively). Typically, the longitudinal outcome is modeled via the Laird & Ware (1982) linear mixed effects model for a continuous and normally distributed outcome. The subject-specific linear predictor of this mixed model (or a function of the predictor) is included in the relative risk model, assuming that a time-independent vector of shared random effects underlies both processes. The random effects account

for both the association between the longitudinal and time-to-event outcomes and the correlation between the repeated measurements in the longitudinal process. The two outcomes are then independent of one another, conditional on the random effects (conditional independence). In the analysis of the PBC data set, we could look at the association between, for example, serum bilirubin and the composite event. However, multiple longitudinal outcomes (of varying types) have been recorded. Extending the univariate joint model to accommodate these multiple longitudinal outcomes allows us to incorporate more information and thereby improve the prognostic ability of the model. We denote by T_i^* the true event time for the ith subject, and to accommodate different types of censoring, we introduce T_i and T_i^U for the observed event times and $\delta \in \{0, 1, 2, 3\}$, which denotes the event indicator, with 0 corresponding to right censoring ($T_i^* > T_i$), 1 to a true event time ($T_i^* = T_i$), 2 to left censoring ($T_i^* < T_i$), and 3 to interval censoring ($T_i < T_i^* < T_i^U$). We allow $\boldsymbol{y}_{ki}$ to denote the $n_{ki} \times 1$ longitudinal response vector for the kth outcome ($k = 1, \ldots, K$) and the ith subject, with elements y_{kil} denoting the value of the kth longitudinal outcome taken at time point t_{kil}, $l = 1, \ldots, n_{ki}$. For each outcome k, we then postulate a generalized linear mixed model (Breslow & Clayton 1993) for the conditional expectation of $\boldsymbol{y}_{ki}$ given a vector of random effects $\boldsymbol{b}_{ki}$:

$$g_k[\mathrm{E}\{y_{ki}(t) \mid \boldsymbol{b}_{ki}\}] = \eta_{ki}(t) = \boldsymbol{x}_{ki}^\top(t)\boldsymbol{\beta}_k + \boldsymbol{z}_{ki}^\top(t)\boldsymbol{b}_{ki}, \tag{1.}$$

where the $\boldsymbol{x}_{ki}(t)$ and $\boldsymbol{z}_{ki}(t)$ are the time-dependent design vectors for the fixed effects $\boldsymbol{\beta}_k$ and the random effects $\boldsymbol{b}_{ki}$, respectively, and $g_k(\cdot)$ is a generic link function. The dimensionality and composition of design vectors $\boldsymbol{x}_{ki}(t)$ and $\boldsymbol{z}_{ki}(t)$ are allowed to differ between the multiple outcomes, and they may also contain a mix of baseline and time-varying covariates. To account for the association between the multiple longitudinal outcomes, we link their corresponding random effects. More specifically, the complete vector of random effects $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}, \boldsymbol{b}_{2i}, \ldots, \boldsymbol{b}_{Ki})^T$ is assumed to follow a multivariate normal distribution with mean zero and unknown $q \times q$ variance-covariance matrix $\boldsymbol{D}$. For the survival process, we have

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i(t)) &= \lim_{\Delta t \to 0} \frac{\Pr\{t \le T_i^* < t + \Delta t \mid T_i^* \ge t, \mathcal{M}_i(t), \boldsymbol{w}_i(t)\}}{\Delta t}, \; t > 0 \\ &= h_0(t) \exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i(t) + \sum_{k=1}^{K} \sum_{l=1}^{L_K} f_{kl}\{\boldsymbol{\alpha}_{kl}, \boldsymbol{w}_i(t), \boldsymbol{b}_{ki}, \mathcal{M}_i(t)\}\right], \end{aligned} \tag{2.}$$

where $\mathcal{M}_i(t) = \{\mathcal{M}_{1i}(t), \ldots, \mathcal{M}_{Ki}(t)\}$, $\mathcal{M}_{ki}(t) = \{\eta_{ki}(s), 0 \le s < t\}$ denotes the history of the true unobserved longitudinal process up to time t, and $\boldsymbol{w}_i(t)$ is a vector of exogenous, possibly time-varying, covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Functions f_{kl}, parameterized by vector $\boldsymbol{\alpha}_{kl}$ and discussed in detail in Section 3.1, specify which components/features of each longitudinal outcome are included in the relative risk model. Some basic formulations (expressed for a single outcome) extensively covered in the literature are

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha \eta_i(t) \tag{3.}$$

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha_1 \eta_i(t) + \alpha_1 \eta_i'(t), \quad \text{with} \quad \eta_i'(t) = \frac{\mathrm{d}\eta_i(t)}{\mathrm{d}t}. \tag{4.}$$

The above formulations posit that the hazard of an event at time t may be associated with the current underlying value of the biomarker at the same time point (Equation 3), or with the current underlying value and the slope at the same time point (Equation 4). For Equation 3, α indicates the change in the log hazard for a unit change in the underlying subject-specific value of the longitudinal outcome and thus determines the strength of the association. For Equation 4, we have the standard interpretation of α (now α_1) as before, and an additional association parameter α_2, which is interpreted as the change in the log hazard for a unit increase in the slope of the

longitudinal trajectory at the same time point, provided $\eta_i(t)$ remains constant. Equation 2 also allows for a combination of L_k functional forms per longitudinal outcome, which may also interact with some covariates in the $\boldsymbol{w}_i(t)$.

Specification of the baseline hazard function, as denoted by $h_0(\cdot)$, is required in the joint model formulation. Calculation of the standard errors is otherwise complicated, given that we work with the full likelihood and not the partial likelihood as in the standard Cox model. Standard errors may also otherwise be underestimated (Hsieh et al. 2006). Since the use of a fully parametric distribution such as the Weibull or Gamma may be too restrictive, we opt for a parametric but flexible specification. Both the piecewise constant and regression spline approaches have shown satisfactory results (Rizopoulos 2012b).

These multivariate joint models have been considered by Rizopoulos & Ghosh (2011), Chi & Ibrahim (2006), Brown et al. (2005), and Lin et al. (2002), among others. Estimation of multivariate joint models is not straightforward since the dimension of the variance-covariance matrix for the random effects increases with the number of outcomes K being modeled, which translates to a high computational burden. Additionally, the more flexible the specification allowed for the $\boldsymbol{x}_{ki}(t)$ and $\boldsymbol{z}_{ki}(t)$ design matrices, for example, in the case of highly nonlinear subject-specific profiles (such as those of serum bilirubin), the higher the dimension of the random effects. Examples of flexible specifications are provided by Rizopoulos et al. (2009) and Brown et al. (2005), who both made use of B-splines with multidimensional random effects; Ding & Wang (2008), who proposed the use of B-splines with a single multiplicative effect; Rizopoulos & Ghosh (2011), who used natural cubic splines; and Umlauf et al. (2017), who used P-splines.

Options to reduce the computational burden include the use of an autocorrelation structure (Proust-Lima et al. 2016) or the factorization of the random effects, as in Li et al. (2012) and Choi et al. (2014). Alternatively, correlation between multiple longitudinal outcomes may instead be modeled via correlated error terms, as in Chi & Ibrahim (2006). One might also specify a common latent variable between the models, such as in Ibrahim et al. (2004) or Proust-Lima et al. (2009), where the longitudinal processes are considered realizations of a single latent process, which is defined in continuous time and represents the common unobserved factor that drives the observed longitudinal trajectories. The fixed effect parameters for the multivariate joint models have been shown to be fairly robust under misspecification of the random effects distribution, excepting those in the time-to-event model. Standard errors may be somewhat underestimated for heavily skewed distributions (Pantazis et al. 2005, Li et al. 2012). The use of semiparametric models for the random effects has been shown to be advantageous in avoiding this misspecification by Pantazis et al. (2005), as well as Rizopoulos & Ghosh (2011) and Tang & Tang (2015), both of whom used a Dirichlet prior for the random effects in a Bayesian framework. In the univariate case, simulation findings from Song et al. (2002) and Tsiatis & Davidian (2001), later corroborated by Rizopoulos et al. (2008) and Huang et al. (2009), suggest that parameter estimates and standard errors for the joint model are both fairly robust.

The above model specification allows for multiple longitudinal outcomes, but does not consider multiple failure times, for example, the competing risks in the PBC data set. Joint models that consider a single longitudinal outcome together with competing risks are described by Rizopoulos (2012b), Elashoff et al. (2008), Williamson et al. (2008), Hu et al. (2009), and Huang et al. (2011), and approaches for the analysis of recurrent events are detailed by Liu et al. (2008), and Liu & Huang (2009). Chi & Ibrahim (2006), Tang et al. (2014), Tang & Tang (2015), and Zhu et al. (2012) address the case with both multiple longitudinal and multiple survival outcomes. Multiple recurrent events are considered by Musoro et al. (2015) and competing risks by Andrinopoulou et al. (2014, 2017b), Proust-Lima et al. (2016), and Li et al. (2012).

3.1. Functional Forms

The formulation of the relative risk model in Equation 2 specifies functions f_{kl}, allowing up to L_k functional forms for each of K longitudinal outcomes. Dissimilar biomarkers have different and often unknown disease mechanisms, and it may be necessary to combine various features of any given biomarker in order to obtain a more accurate understanding of their relationship to risk. While the current value association structure outlined in Equation 3 is simple and intuitive, it may not always be appropriate, and fitting only this association structure may lead to incorrect conclusions when a more complex form of association exists (Vacek 1997). In addition to the familiar slope parameterization (Equation 4), another simple extension to the current value allows for the use of lagged effects, where the risk of an event at time t is associated with the true value of the underlying marker at some time $t - c$, where c is the lag of interest. All of the preceding parameterizations assume that the risk of an event is associated with a feature of the longitudinal trajectory at a single time point, as demonstrated in **Figure 2*b*** and **2*c*** for the current value and slope specifications. It may, however, be beneficial to allow the risk to depend on a function of the longitudinal marker history instead (Vacek 1997, Hauptmann et al. 2000, Sylvestre & Abrahamowicz 2009). We could consider the cumulative effect, such that

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha \int_0^t \mu_i(s)\, \mathrm{d}s,$$

or, more realistically, a weighted cumulative effect (which includes both the current value and cumulative parameterizations as special cases):

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha \int_0^t \varpi(t-s)\mu_i(s)\, \mathrm{d}s.$$

The parameter α then reflects the strength of the association between the risk for an event at time t and the specific area under the longitudinal trajectory for which the weights are nonzero (**Figure 2*c***). Mauff et al. (2017) considered the use of the standard normal density as a weight function, which assigns larger weights to more recent measurements of the biomarker and allows for the estimation of the period for which the weights are nonzero, through the estimation of the scale parameter. In an analysis of the PBC data set, different periods of interest were determined for each of serum bilirubin, serum cholesterol, and hepatomegaly. An additional parameterization of interest is the random effects association structure. In this parameterization, we include only the random effects from the longitudinal model in the linear predictor of the relative risk model, which conveys a distinct computational advantage, in that no numerical integration is required (see Section 4):

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \boldsymbol{\alpha}^T \boldsymbol{b}_i.$$

For a simple random effects structure (intercept and slope), the interpretation of the random effects is then the deviation from the average intercept and the average slope, and the association parameter reflects the change in the log hazard for a unit change in this deviation. For more elaborate structures, however, this parameterization may be difficult to interpret (Rizopoulos 2012b, Gould et al. 2015). In some settings it may be plausible that the association parameter α is allowed to change during follow-up. Andrinopoulou et al. (2017a) model a time-varying coefficient using P-splines, specifying $f\{\lambda_i(t), \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\}$ (again for a specific longitudinal outcome), where

$$\lambda_i(t) = \sum_{p=1}^{P} \alpha_p B_p(t).$$

Then α_p is a set of parameters that capture the strength of association between the longitudinal and survival outcomes and $B_p(t)$ denotes the pth basis function of a B-spline. This method assumes a

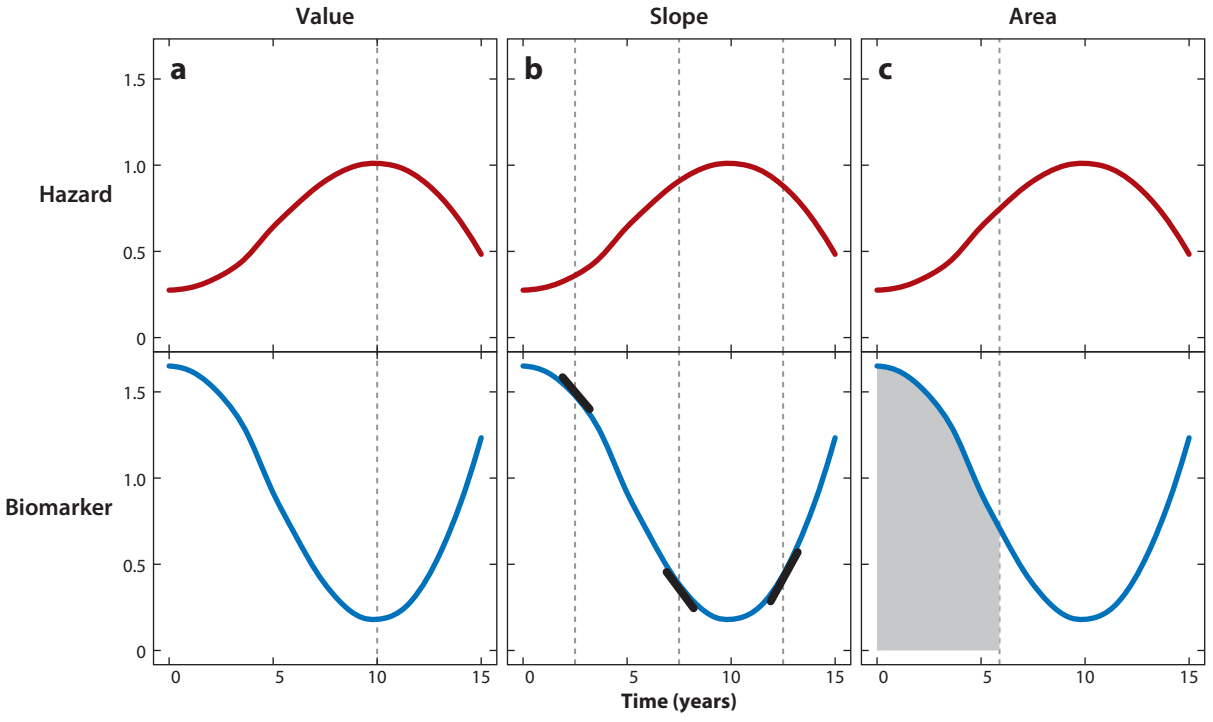

Figure 2

Illustration of the current value, current slope, and cumulative effect association structures. The thick black lines represent the slope of the biomarker. The vertical dashed lines represent the time point (e.g., in panel *a* the biomarker value at time point 10 is associated with the value of the hazard at the same time point).

high number of equally spaced knots, combined with a roughness penalty used to obtain sufficient smoothness and avoid overfitting. Furthermore, Andrinopoulou & Rizopoulos (2016) look at the use of Bayesian shrinkage methods to choose between functional forms and determine the form best suited to the analysis in question, and Andrinopoulou et al. (2017b) and Crowther (2015) allow for the use of different association structures for different (multivariate) longitudinal outcomes.

4. ESTIMATION

Estimation methods proposed in the literature fall under either the Bayesian or frequentist paradigm. Both exploit the full joint likelihood derived from the joint distribution of longitudinal and time-to-event outcomes. The expression for this likelihood function is derived under the conditional independence assumption (Section 3). Ignoring conditioning on the covariates for the sake of brevity, the expression for the likelihood is given by

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta} \mid \mathcal{D}_n) &= \prod_{i=1}^{n} p(T_i, T_i^U \delta_i, \boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{Ki}; \boldsymbol{\theta}) \\
&= \prod_{i=1}^{n} \int p(T_i, T_i^U, \delta_i, \boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{Ki}, \boldsymbol{b}_i; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{b}_i \\
&= \prod_{i=1}^{n} \int \left\{ \prod_{k=1}^{K} \prod_{l=1}^{n_{ki}} p(y_{kil} \mid \boldsymbol{b}_{ki}; \boldsymbol{\theta}) \right\} p(T_i, T_i^U, \delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}) p(\boldsymbol{b}_i; \boldsymbol{\theta}) \mathrm{d}\boldsymbol{b}_i, \qquad 5.
\end{aligned}$$

where $\mathcal{D}_n = \{T_i, T_i^U, \delta_i, \boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{Ki}; i = 1, \ldots, n\}$ are the observed data, $\boldsymbol{\theta}$ is the vector of all parameters, and $p(y_{kil} \mid \boldsymbol{b}_{ki}; \boldsymbol{\theta})$ is the contribution of the lth repeated measurement of the ith subject for the kth outcome, given by

$$p(y_{kil} \mid \boldsymbol{b}_{ki}; \boldsymbol{\theta}) = \exp\left\{\left[y_{kil}\psi_{kil}(\boldsymbol{b}_{ki}) - c_k\{\psi_{kil}(\boldsymbol{b}_{ki})\}\right]/a_k(\varphi) - d_k(y_{kil}, \varphi)\right\}.$$

$\psi_{kil}(\boldsymbol{b}_{ki})$ and φ denote the natural and dispersion parameters in the exponential family, respectively, and where $c_k(\cdot)$, $a_k(\cdot)$, and $d_k(\cdot)$ are known functions specifying the member of the exponential family. The joint density for the survival part together with the random effects is given by

$$\begin{aligned} p(T_i, T_i^U, \delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}) p(\boldsymbol{b}_i; \boldsymbol{\theta}) \;=\; & \left[h_i\{T_i \mid \mathcal{M}_i(T_i)\}\right]^{I(\delta_i=1)} \exp\left[-\int_0^{T_i} h_i\{s \mid \mathcal{M}_i(s)\}\mathrm{d}s\right]^{I(\delta_i \in \{0,1\})} \\ & \times \left\{1 - \exp\left[-\int_0^{T_i} h_i\{s \mid \mathcal{M}_i(s)\}\mathrm{d}s\right]^{I(\delta_i \in \{0,1\})}\right\}^{I(\delta_i=2)} \\ & \times \left\{\exp\left[-\int_0^{T_i} h_i\{s \mid \mathcal{M}_i(s)\}\mathrm{d}s\right] - \exp\left[-\int_0^{T_i^U} h_i\{s \mid \mathcal{M}_i(s)\}\mathrm{d}s\right]\right\}^{I(\delta_i=3)} \\ & \times \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{D})}} \exp(\boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i), \end{aligned}$$

where $I(\cdot)$ denotes the indicator function. The unidimensional integral in the definition of the survival function does not have a closed-form solution, and thus a numerical method must be employed for its evaluation. Standard options are the Gauss-Kronrod and Gauss-Legendre quadrature rules (Rizopoulos 2012b).

4.1. Frequentist Estimation

Under the frequentist paradigm, estimates of the joint model parameters can be derived by maximizing the log-likelihood function (see Equation 5). Maximum likelihood estimates (MLE) can be obtained using the expectation-maximization (EM) algorithm (Wulfsohn & Tsiatis 1997), treating the random effects as missing data, or by maximizing the log-likelihood directly using Newton-like approaches (Thiébaut et al. 2005, Guedj et al. 2011, Crowther et al. 2013). The main motivation for using the EM algorithm is the closed-form M-step updates for certain parameters of the joint model. In the E-step of the EM algorithm, the expectation of the log-likelihood function is obtained with respect to the posterior distribution of the latent random effects given the observed data and the parameters obtained in the previous M-step. The integral over the random effects in this E-step does not have a closed form and thus needs to be solved numerically for each patient in each iteration. To avoid this issue, Rizopoulos (2012a) has proposed pseudoadaptive Gaussian quadrature rules. To improve the convergence of the EM algorithm, a hybrid optimization technique (Rizopoulos 2010) has also been proposed. However, in all methods (direct maximization, EM, and the hybrid approach), in high-dimensional cases (Section 3), the integral over the random effects becomes computationally burdensome. In these settings, the use of Laplace approximation reduces this computational burden substantially (Rizopoulos et al. 2009).

Parameter estimates obtained using MLE are asymptotically unbiased, normal, and consistent. The standard error for these parameter estimates is obtained from the observed Fisher information matrix (Rizopoulos 2012b). However, inversion of the Hessian matrix to obtain standard errors may result in numerical complications when the Hessian matrix is high dimensional. This is especially the case when the baseline hazard in the joint model is left unspecified. In such a scenario, the calculation of the likelihood is based on nonparametric maximum likelihood

arguments under which the unspecified cumulative incidence function $H_0(t) = \int_0^t h_0(s)\,\mathrm{d}s$ is replaced by a high-dimensional step function with jumps at each of the unique event times (Van der Vaart 1998). Alternative approaches for the estimation of standard errors, such as the profile likelihood approach, result in underestimation (Rizopoulos 2012b). To avoid these issues, Hsieh et al. (2006) proposed the use of bootstrapping to estimate the standard errors. However, such an approach is computationally intensive in the context of the joint model. A feasible alternative is to postulate a flexible but parametric model for the baseline hazard $h_0(t)$. In this scenario, inference on parameter estimates can be done using standard asymptotic maximum likelihood theory. Consequently, the likelihood ratio, Wald, and score tests can be used for inference on the parameter estimates. For estimation and inference on the random effects, empirical Bayes estimation (estimation using the posterior distribution of the random effects, given the observed data and point MLE of the parameters) is the standard approach.

In general, parameter estimation based on maximum likelihood estimation is vulnerable to nonidentifiability, which results in unreliable parameter inference. However, for the joint model in particular, researchers investigating applications have yet to report any issues relating to identifiability beyond those concerning the individual submodels. That is, should the mixed effect and event process submodels satisfy identifiability conditions (not discussed in this article), so too should the joint model (Zeng & Cai 2005).

4.2. Bayesian Estimation

Parameter estimation for joint models using a Bayesian approach involves sampling from the posterior distribution of the parameters using standard Markov chain Monte Carlo (MCMC) or Hamiltonian Monte Carlo algorithms. Given the wide availability of BUGS (Bayesian inference using Gibbs sampling) software, the Gibbs sampling technique is extensively used. However, in practice, Gibbs sampling may be used in combination with other approaches such as the Metropolis-Hastings algorithm and slice sampling, among others (Wang & Taylor 2001, Xu & Zeger 2001, Brown & Ibrahim 2003, Rizopoulos 2016). It should be noted that under the Bayesian approach, the random effects are also considered model parameters for which we obtain a posterior sample. Consequently, solving the integral over the random effects (Equation 5) is no longer required. The complete posterior distribution of the parameters is given by

$$p(\boldsymbol{\theta}, \boldsymbol{b} \mid \mathcal{D}_n) \propto \prod_{i=1}^{n} p(T_i, T_i^U, \delta_i, \boldsymbol{y}_{1i}, \ldots, \boldsymbol{y}_{Ki} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}),$$

where $p(\boldsymbol{\theta})$ is the joint prior distribution for parameters in $\boldsymbol{\theta}$. In practice, independent univariate diffuse normal priors are typically used for each of $\boldsymbol{\beta}_k$, $\boldsymbol{\gamma}$, h_0, and $\boldsymbol{\alpha}_k$. An inverse Wishart prior is used for the covariance matrix of the random effects $\boldsymbol{D}$, and when fitting a joint model with a normally distributed longitudinal outcome, an inverse-Gamma prior is used for the variance of the error terms σ^2. The ability to specify priors under the Bayesian paradigm has certain advantages. For example, in instances where the goal is to estimate more elaborate models with many coefficients in the relative risk submodel, and the number of events are few, shrinkage methods such as lasso or Ridge can be easily implemented merely by changing the prior distribution (Andrinopoulou & Rizopoulos 2016). Parameter inference proceeds as is standard under the Bayesian approach, and model comparisons can be performed using the deviance information criterion.

The speed of computation using the Bayesian approach is slower when compared with frequentist methods when a simple univariate joint model is used. It is, however, comparable

Table 1 Estimated coefficients for the parameters of the joint model fitted to the PBC data set

Parameter	ML estimate	95% Confidence interval	MCMC estimate	95% Credibility interval
β_0	0.489	(0.376; 0.603)	0.476	(0.366; 0.591)
β_1	0.190	(0.163; 0.216)	0.213	(0.137; 0.291)
γ_1	−0.049	(−0.377; 0.279)	−0.052	(−0.374; 0.28)
γ_2	0.041	(0.026; 0.057)	0.044	(0.028; 0.06)
α	1.304	(1.125; 1.483)	1.335	(1.162; 1.524)
σ^2	0.121	(0.31; 0.385)	0.115	(0.326; 0.352)

Abbreviations: MCMC, Markov chain Monte Carlo; ML, maximum likelihood; PBC, primary biliary cirrhosis.

for more complex models involving nonnormal outcomes and complex association structures (Rizopoulos 2016). Comparable computation speeds and easier implementation of more elaborate models makes the Bayesian approach an attractive option.

As an illustration of the two estimation paradigms, we fit the following model to the PBC data set:

$$\begin{cases} \log\{y_i(t)\} = \eta_i(t) + \epsilon_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\, t + \epsilon_i(t), \\ h_i(t) = h_0(t) \exp\{\gamma_1 \text{age} + \gamma_2 \text{drug} + \alpha \eta_i(t)\}, \\ \boldsymbol{b_i} \sim \mathcal{N}(0, \boldsymbol{D}), \epsilon_i(t) \sim \mathcal{N}\left(0, \sigma^2\right), \end{cases}$$

using both frequentist and Bayesian estimation. In the above formulation, $y_i(t)$ denotes the observed serum bilirubin values for the ith subject. We used a linear mixed-effects model with random intercept and slope for the longitudinal submodel and adjusted for age and treatment (placebo versus D-penicillamine) at baseline in the survival submodel. The current value association structure was used, assuming that the underlying value of log(serum bilirubin) at time t has an effect on the instantaneous risk for transplantation or death at the same time point, $h_i(t)$.

The model was fitted using function `jointModel()` from R package `JM` and function `jointModelBayes()` from package `JMbayes` for ML and MCMC estimation, respectively (Rizopoulos 2010, 2016). **Table 1** includes the parameter estimates, along with the 95% confidence and credibility intervals for ML and MCMC estimation, respectively. It should be noted, however, that the two models are not absolutely comparable because there is a difference in the approximation of the baseline hazard function between functions `jointModel()` and `jointModelBayes()`. The former uses a B-splines approximation and the latter uses a P-splines approximation. Nevertheless, we ensured that the same number of internal knots was used in both cases.

5. DYNAMIC PREDICTIONS

Thus far we have focused on the formulation and estimation of joint models and discussed various association structures. Here we focus on predictions, which exploit the dynamic nature of joint models. Personalized medicine has recently gathered much attention in medical research. Physicians need predictive tools, tailored to the individual characteristics of their patients, to improve decision making, optimize medical care, and facilitate communication of individual health risks to their patients. Moreover, it is crucial that such predictive tools can be updated with new information as soon as it becomes available. Based on a fitted joint model, subject-specific survival probabilities, as well as predictions for future biomarker levels, may be derived for a new subject j for whom a set of measurements $\mathcal{Y}_{kj}(t) = \{y_{kj}(t_{kjl}); 0 \leq t_{kjl} \leq t, l = 1, \ldots, n_j, k = 1, \ldots, K\}$ for the kth longitudinal outcome and a vector of baseline covariates $\boldsymbol{w}_j$ are available. Since the

biomarker measurements are available up to time t (except for the case of interval censoring, for which they may be available up to $s > t$), we may assume that subject j did not experience the event of interest up to this time point. We can therefore determine the conditional subject-specific survival probability that subject j survives at least up to a future time point $u > t$ given that the subject was event-free up to time t,

$$\pi_j(u \mid t) = \Pr\{T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t), \boldsymbol{w}_j\},$$

as well as the prediction for the subject's longitudinal response at u,

$$\omega_{kj}(u \mid t) = \mathrm{E}\{y_{kj}(u) \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t)\}.$$

Both $\pi_j(u \mid t)$ and $\omega_{kj}(u \mid t)$ can be updated to $\pi_j(u \mid t_{\text{new}})$ and $\omega_{kj}(u \mid t_{\text{new}})$ when new information is provided for subject j at a future time point $t_{\text{new}} > t$. To illustrate this feature, a basic joint model was fitted to the PBC data set. The subject-specific conditional survival probabilities for a new subject, based on this joint model, are shown in **Figure 3**. In **Figure 3*a***, the prediction is based on four serum bilirubin measurements recorded during the first two years of follow-up. In **Figure 3*b***, an additional value of serum bilirubin is available, leading to a steeper decrease in the survival probability over time for this subject.

Estimation of both $\pi_j(u \mid t)$ and $\omega_{kj}(u \mid t)$ is based on the corresponding posterior predictive distributions, exploiting the conditional independence assumptions that were discussed in Sections 3 and 4 (Yu et al. 2008; Proust-Lima & Taylor 2009; Rizopoulos 2011, 2012b; Taylor et al. 2013):

$$\begin{aligned}\pi_j(u \mid t) &= \int \Pr\left(T_j^* \geq u \mid T_j^* > t, \boldsymbol{b}_j, \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t), \boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{b}_j \\ &= \int \frac{S_j\{u \mid \mathcal{H}_j\left(u, \boldsymbol{b}_j\right), \boldsymbol{\theta}\}}{S_j\{t \mid \mathcal{H}_j\left(t, \boldsymbol{b}_j\right), \boldsymbol{\theta}\}} p\{\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t), \boldsymbol{\theta}\} \mathrm{d}\boldsymbol{b}_j,\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}\{y_{kj}(u) \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t), \boldsymbol{\theta}\} &= \int \mathrm{E}\{y_{kj}(u) \mid \boldsymbol{b}_j, \boldsymbol{\theta}\} p\left(\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t), \boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{b}_j \\ &= \boldsymbol{x}_{kj}^\top(u) \boldsymbol{\beta} + \boldsymbol{z}_{kj}^\top(u) \bar{\boldsymbol{b}}_j^{(t)},\end{aligned}$$

with $\bar{\boldsymbol{b}}_j^{(t)} = \int \boldsymbol{b}_j p\left(\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{1j}(t), \ldots, \mathcal{Y}_{Kj}(t), \boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{b}_j$.

The predictive performance of dynamic predictions from joint models has also received much attention in the literature. Appropriate measures that account for the dynamic nature of such predictions have been developed. These measures mainly focus on calibration (i.e., how the model performs in predicting the observed data) and discrimination (i.e., how the model performs in discriminating between patients who experienced the event and those who did not). In terms of calibration, the expected prediction error is discussed in Henderson et al. (2002) and Schoop et al. (2011). Aside from the dynamic feature of these predictions, the presence of censoring needs to be accounted for when assessing their predictive performance. This is achieved using inverse probability weighted estimators (Schoop et al. 2008, Blanche et al. 2015, Rizopoulos et al. 2017). Alternatively, information theoretic calibration measures based on the Kullback-Leibler divergence have also been proposed (Commenges et al. 2012, Proust-Lima et al. 2014). To address discrimination, the time-dependent area under the receiver operating characteristic curve and the dynamic concordance index (Rizopoulos 2011, Rizopoulos et al. 2017) are commonly used.

Aside from joint models, landmark analysis may also be used to derive dynamically updated estimates of survival probabilities. Recent work, however, has shown that dynamic predictions

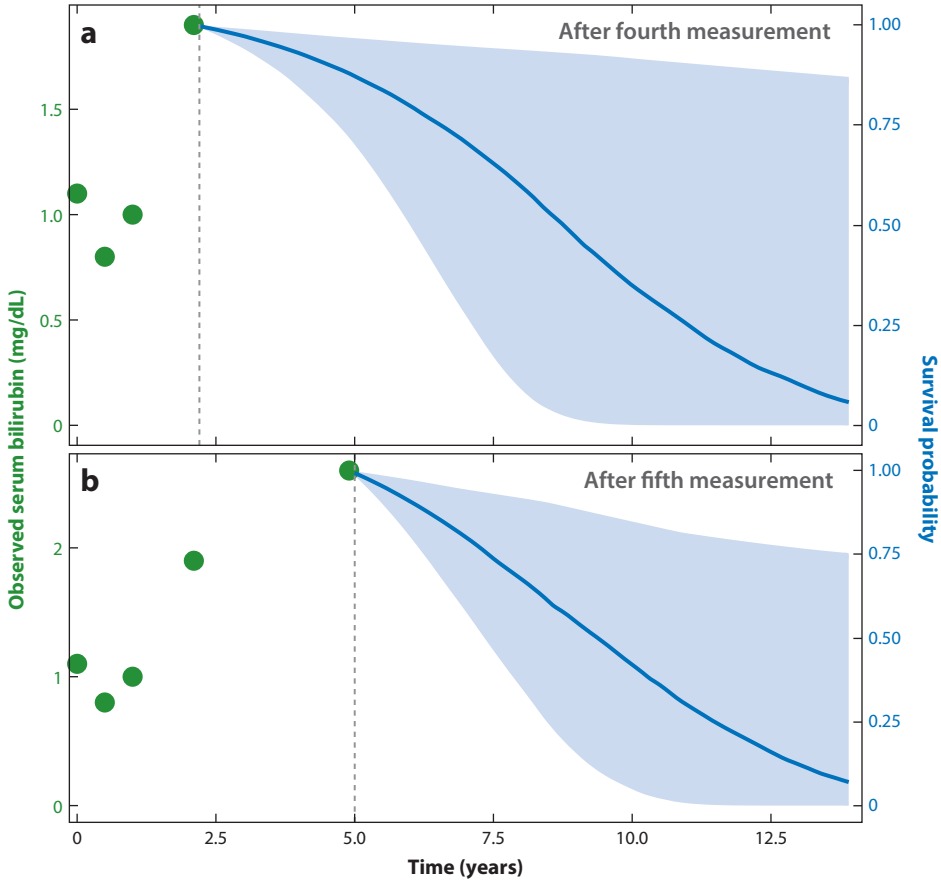

Figure 3

Conditional survival probabilities for a new subject from the primary biliary cirrhosis data set after (*a*) four and (*b*) five serum bilirubin measurements. The vertical dashed lines represent the time point of the latest biomarker measurement.

from joint models outperform the predictions based on landmark analysis (Ferrer et al. 2017, Rizopoulos et al. 2017, Suresh et al. 2017). Attention must be given to the parametrization of the joint model, especially the functional form of the association structure that links the longitudinal and survival processes, since this may impact predictive performance. To this end, Rizopoulos et al. (2014) proposed the use of Bayesian model averaging to combine dynamic predictions under different association structures. This approach was further extended for the case of multivariate joint models with competing risk survival data by Andrinopoulou et al. (2017b).

Dynamic predictions from joint models are intuitively appealing and can be used in a broad range of applications. Recent work focuses on improving their quality and on their use in novel settings. The P-splines approach of Andrinopoulou et al. (2017a), discussed in Section 3.1, aims to improve the performance of dynamic predictions. A similar approach was proposed by Köhler et al. (2017), who used P-splines in the longitudinal submodel to more precisely capture highly nonlinear subject-specific trajectories. Dynamic predictions have also been used to develop personalized schedules as an alternative to standard fixed and frequent follow-up schedules. The latter are

common for all patients and do not account for disease progression over time. To address the burden caused by such schedules, dynamic patient-specific schedules based on dynamic predictions have been proposed by Rizopoulos et al. (2015) for the measurement of biomarkers and Tomer et al. (2017) for the performance of prostate cancer biopsies. Sène et al. (2016) used dynamic predictions of prostate cancer recurrence with and without the initiation of a second treatment to assess the benefit of radiation therapy. Recently, Papageorgiou et al. (2018) proposed a flexible joint modeling framework that enables the derivation of dynamic predictions under different scenarios with respect to the occurrence of future intermediate events, such as changes in treatment and serious adverse events. This application of dynamic predictions can be very useful when planning interventions because it can be used to quantify the potential benefit (or lack thereof) of various treatment strategies on the risk of the clinical endpoint of interest for the individual patient.

6. SOFTWARE

The ongoing increase in theoretical research and application of joint models is accompanied by a respective increase in the development of appropriate statistical software. Joint modeling can be achieved with a range of tools available via standard statistical software such as R (R Core Team 2018), SAS/STAT, Stata, WinBUGS, and JAGS.

In R, there are a number of packages either dedicated to joint modeling of longitudinal and time-to-event data or capable of joint modeling. Among them, package `JM` (Rizopoulos 2010), which is based on maximum likelihood estimation employing an EM algorithm, may be used for the joint analysis of a single continuous longitudinal outcome with either a single time-to-event outcome or competing risks. Various baseline hazard specifications may be selected. Built upon the Bayesian framework, package `JMbayes` (Rizopoulos 2016) is capable of fitting a wide range of joint models using MCMC algorithms. The most notable feature of `JMbayes` is that both univariate and multivariate joint models for continuous and categorical longitudinal outcomes are available. Both `JM` and `JMbayes` allow for the flexible selection of various association structures and also include tools for model diagnostic checks, derivation of dynamic predictions, and assessment of their accuracy in terms of discrimination and calibration. In addition to these features, `JMbayes` also includes the option to use time-varying effects for the association parameters using P-splines. Another R package that uses the EM algorithm is `joineR` (Williamson et al. 2008, Philipson et al. 2018). This package can be used to fit univariate joint models for a continuous longitudinal outcome, a single time-to-event outcome with a random latent association, or a competing risks joint model.

Multivariate joint models for multiple continuous longitudinal outcomes and time-to-event data with a multivariate latent Gaussian process association are also available in the R package `joineRML` (Hickey et al. 2018), which uses a Monte Carlo EM algorithm. Package `joineRML` also includes functions to derive dynamic predictions based on the fitted joint models. The R package `frailtypack` (Rondeau & Gonzalez 2005, Rondeau et al. 2012, Król et al. 2017) has been extended and can now fit a joint model for a continuous longitudinal response and either a terminal event or recurrent events along with a terminal event (trivariate joint model). It is also able to derive dynamic predictions for the aforementioned joint models. Estimation in `frailtypack` is based on maximum likelihood using either a parametric or semiparametric approach on the penalized likelihood for estimation of the hazard functions. The recent addition of the function `stan_jm` in package `rstanarm` (Stan Development Team 2016) enables the user to fit both univariate and multivariate joint models for continuous and/or categorical longitudinal outcomes with a time-to-event outcome using the R interface to the Stan C++ library for Bayesian estimation. Various association structures may be selected, and functions for dynamic predictions are also available.

Package `bamlss` (Umlauf et al. 2017) includes functions to fit and derive predictions for a flexible additive joint model for a continuous longitudinal outcome and single time-to-event outcome under the Bayesian framework. Finally, package `lcmm` (Proust-Lima et al. 2017a,b) may be used to fit joint latent class models. Estimation in `lcmm` is based on the frequentist approach, and it features predictive tools and the EPOCE estimator for assessing predictive accuracy.

In SAS, there are two notable macros that may be used to fit joint models for longitudinal and time-to-event data. Macro `%JM` can be used to fit a shared parameter model for the joint analysis of longitudinal and time-to-event outcomes. Gaussian and non-Gaussian distributions, along with linear and nonlinear evolutions over time using splines, are supported for the mixed-effects submodel. Various options for the association structure are also available. SAS macro `JMfit` (Zhang et al. 2016) can also be used to fit a variety of popular joint models while providing several measures of assessment, including the decomposition of AIC (the Akaike information criterion) and BIC (the Bayesian information criterion) as well as ΔAIC and ΔBIC (Zhang et al. 2014).

In Stata, package `stjm` (Crowther 2012) can be used to fit a shared parameter joint model for a single continuous longitudinal outcome and a single time-to-event outcome under the maximum likelihood framework. Flexible parametrization of the linear mixed-effects submodel is supported using either polynomials or splines along with several options for the specification of the survival submodel. In addition, the user has the option to choose between three different association structures to link the two processes.

This section has focused on the packages described above because they are state-of-the-art software tools available in popular statistical environments, they are accompanied by detailed reference manuals and documentation regarding their usage, and they are actively maintained and likely to be extended in the future. However, the full range of software tools available is far more extensive. Many examples of joint models use other software such as WinBUGS and JAGS, with the code available as supplementary material or upon request from the author (Rizopoulos & Ghosh 2011; Andrinopoulou et al. 2014, 2017b; Baghfalaki et al. 2014; Liu & Li 2016).

7. DISCUSSION

The joint analysis of longitudinal and survival data has received much attention during the past two decades, yielding a large body of research with numerous clinical applications. As a result, a broad range of joint models has been developed under both the frequentist and Bayesian paradigms. The parallel development of statistical software tools that are able to implement joint models is increasing their accessibility while steadily decreasing the computational time required to run them. Motivated, in most cases, by the improved quality and volume of data, as well as the focus on personalized medicine, recent work on joint models has covered topics such as multivariate joint models, personalized screening schedules, methods for the selection of the association structure, methods for improving the accuracy of dynamic predictions and utilizing joint models and dynamic predictions in optimizing individualized treatment strategies, and medical decision making.

Joint models have been widely used in health studies offering many advantages, such as (*a*) addressing measurement error and missing data in time-dependent covariates in survival regression models, (*b*) increasing efficiency in statistical inference by utilizing both the longitudinal and survival data simultaneously, (*c*) understanding and quantifying the association between longitudinal and survival processes, and (*d*) developing dynamic prediction tools for risk assessment. Their disadvantages are of a practical nature: More complex models are still computationally expensive and may be difficult to interpret and communicate to a broader audience. However, we believe that the benefits of using joint models surpasses their practical disadvantages and that they will continue to grow in popularity in the near future as more user-friendly software becomes available.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Andrinopoulou ER, Eilers PH, Takkenberg JJ, Rizopoulos D. 2017a. Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics* 74:685–93

Andrinopoulou ER, Rizopoulos D. 2016. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Stat. Med.* 35:4813–23

Andrinopoulou ER, Rizopoulos D, Takkenberg JJ, Lesaffre E. 2014. Joint modeling of two longitudinal outcomes and competing risk data. *Stat. Med.* 33:3167–78

Andrinopoulou ER, Rizopoulos D, Takkenberg JJ, Lesaffre E. 2017b. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Stat. Methods Med. Res.* 26:1787–801

Baghfalaki T, Ganjali M, Berridge D. 2014. Joint modeling of multivariate longitudinal mixed measurements and time to event data using a Bayesian approach. *J. Appl. Stat.* 41:1934–55

Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, Jacqmin-Gadda H. 2015. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics* 71:102–13

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.* 88:9–25

Brown ER, Ibrahim JG. 2003. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 59:221–28

Brown ER, Ibrahim JG, DeGruttola V. 2005. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics* 61:64–73

Chi YY, Ibrahim JG. 2006. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* 62:432–45

Choi J, Anderson SJ, Richards TJ, Thompson WK. 2014. Prediction of transplant-free survival in idiopathic pulmonary fibrosis patients using joint models for event times and mixed multivariate longitudinal data. *J. Appl. Stat.* 41:2192–205

Commenges D, Liquet B, Proust-Lima C. 2012. Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback–Leibler risks. *Biometrics* 68:380–87

Crowther MJ. 2012. STJM: Stata module to fit shared parameter joint models of longitudinal and survival data. *Statistical software*. Stat. Softw. Compon. S457502, Boston Coll. Dep. Econ. **https://ideas.repec.org/c/boc/bocode/s457502.html**

Crowther MJ. 2015. Extensions to the STJM package. *Statistical software*. Univ. Leicester. **https://ideas.repec.org/c/boc/bocode/s457502.html**

Crowther MJ, Abrams KR, Lambert PC. 2013. Joint modeling of longitudinal and survival data. *Stata J.* 13:165–84

De Gruttola V, Tu XM. 1994. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* 50:1003–14

Ding J, Wang JL. 2008. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* 64:546–56

Elashoff RM, Li G, Li N. 2008. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* 64:762–71

Faucett CL, Thomas DC. 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Stat. Med.* 15:1663–85

Ferrer L, Putter H, Proust-Lima C. 2017. Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment. arXiv:1707.03706 [stat.AP]

Garre FG, Zwinderman AH, Geskus RB, Sijpkens YW. 2008. A joint latent class changepoint model to improve the prediction of time to graft failure. *J. R. Stat. Soc. A* 171:299–308

Gould AL, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, et al. 2015. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Stat. Med.* 34:2181–95

Guedj J, Thiébaut R, Commenges D. 2011. Joint modeling of the clinical progression and of the biomarkers' dynamics using a mechanistic model. *Biometrics* 67:59–66

Hauptmann M, Wellmann J, Lubin JH, Rosenberg PS, Kreienbrock L. 2000. Analysis of exposure-time-response relationships using a spline weight function. *Biometrics* 56:1105–8

Henderson R, Diggle P, Dobson A. 2002. Identification and efficacy of longitudinal markers for survival. *Biostatistics* 3:33–50

Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. 2018. `joineRML`: joint modelling of multivariate longitudinal data and time-to-event outcomes. *R package version 0.4.1*

Hogan JW, Laird NM. 1997. Mixture models for the joint distribution of repeated measures and event times. *Stat. Med.* 16:239–57

Hsieh F, Tseng YK, Wang JL. 2006. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* 62:1037–43

Hu W, Li G, Li N. 2009. A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat. Med.* 28:1601–19

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

Huang X, Stefanski LA, Davidian M. 2009. Latent-model robustness in joint models for a primary endpoint and a longitudinal process. *Biometrics* 65:719–27

Ibrahim JG, Chen MH, Sinha D. 2004. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Stat. Sin.* 14:863–83

Köhler M, Umlauf N, Beyerlein A, Winkler C, Ziegler AG, Greven S. 2017. Flexible Bayesian additive joint models with an application to type 1 diabetes research. *Biometrical J.* 59:1144–65

Król A, Mauguen A, Mazroui Y, Laurent A, Michiels S, Rondeau V. 2017. Tutorial in joint modeling and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *J. Stat. Softw.* 81:1–52

Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. *Biometrics* 38:963–74

Li N, Elashoff RM, Li G, Tseng CH. 2012. Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Stat. Med.* 31:1707–21

Lin H, McCulloch CE, Mayne ST. 2002. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat. Med.* 21:2369–82

Liu F, Li Q. 2016. A Bayesian model for joint analysis of multivariate repeated measures and time to event data in crossover trials. *Stat. Methods Med. Res.* 25:2180–92

Liu L, Huang X. 2009. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *J. R. Stat. Soc. C* 58:65–81

Liu L, Huang X, O'Quigley J. 2008. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* 64:950–58

Mauff K, Steyerberg EW, Nijpels G, Heijden AA, Rizopoulos D. 2017. Extension of the association structure in joint models to include weighted cumulative effects. *Stat. Med.* 36:3746–59

Murtaugh P, Dickson E, Dam GV, Malincho M, Grambsch P, et al. 1994. Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology* 20:126–34

Musoro JZ, Geskus RB, Zwinderman AH. 2015. A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant. *Biometrical J.* 57:185–200

Pantazis N, Touloumi G, Walker A, Babiker A. 2005. Bivariate modelling of longitudinal measurements of two human immunodeficiency type 1 disease progression markers in the presence of informative drop-outs. *J. R. Stat. Soc. C* 54:405–23

Papageorgiou G, Mokhles MM, Takkenberg JJM, Rizopoulos D. 2018. Individualized dynamic prediction of survival under time-varying treatment strategies. arXiv:1804.02334 [stat.AP]

Philipson P, Sousa I, Diggle P, Williamson P, Kolamunnage-Dona R, et al. 2018. joineR: joint modelling of repeated measurements and time-to-event data. *R package version 1.2.3*

Prentice R. 1982. Covariate measurement errors and parameter estimates in a failure time regression model. *Biometrika* 69:331–42

Proust-Lima C, Dartigues JF, Jacqmin-Gadda H. 2016. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Stat. Med.* 35:382–98

Proust-Lima C, Joly P, Dartigues JF, Jacqmin-Gadda H. 2009. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Comput. Stat. Data Anal.* 53:1142–54

Proust-Lima C, Philipps V, Diakite A, Liquet B. 2017a. lcmm: extended mixed models using latent classes and latent processes. *R package version 1.7.8*

Proust-Lima C, Philipps V, Liquet B. 2017b. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *J. Stat. Softw.* 78:1–56

Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. 2014. Joint latent class models for longitudinal and time-to-event data: a review. *Stat. Methods Med. Res.* 23:74–90

Proust-Lima C, Taylor J. 2009. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post treatment PSA: a joint modeling approach. *Biostatistics* 10:535–49

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Found. Stat. Comput.

Rizopoulos D. 2010. JM: An R package for the joint modelling of longitudinal and time-to-event data. *J. Stat. Softw.* 35:1–33

Rizopoulos D. 2011. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 67:819–29

Rizopoulos D. 2012a. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Comput. Stat. Data Anal.* 56:491–501

Rizopoulos D. 2012b. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Boca Raton, FL: Chapman & Hall/CRC

Rizopoulos D. 2016. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *J. Stat. Softw.* 72:1–45

Rizopoulos D, Ghosh P. 2011. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat. Med.* 30:1366–80

Rizopoulos D, Hatfield L, Carlin B, Takkenberg J. 2014. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *J. Am. Stat. Assoc.* 109:1385–97

Rizopoulos D, Molenberghs G, Lesaffre E. 2017. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical J.* 59:1261–76

Rizopoulos D, Taylor J, Rosmalen JV, Steyerberg E, Takkenberg J. 2015. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics* 17:149–64

Rizopoulos D, Verbeke G, Lesaffre E. 2008. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics* 64:611–19

Rizopoulos D, Verbeke G, Lesaffre E. 2009. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *J. R. Stat. Soc. B* 71:637–54

Rondeau V, Gonzalez J. 2005. frailtypack: A computer program for the analysis of correlated failure time data using penalized likelihood estimation. *Comput. Methods Programs Biomed.* 80:154–64

Rondeau V, Mazroui Y, Gonzalez J. 2012. frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *J. Stat. Softw.* 47:1–28

Schoop R, Graf E, Schumacher M. 2008. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics* 64:603–10

Schoop R, Schumacher M, Graf E. 2011. Measures of prediction error for survival data with longitudinal covariates. *Biometrical J.* 53:275–93

Self S, Patiwan Y. 1992. Modeling a marker of disease progression and onset of disease. In *AIDS Epidemiology: Methodological Issues*, ed. NP Jewell, K Dietz, VT Farewell, pp. 231–55. Boston: Birkhäuser

Sène M, Taylor JM, Dignam JJ, Jacqmin-Gadda H, Proust-Lima C. 2016. Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: development and validation. *Stat. Methods Med. Res.* 25:2972–91

Song X, Davidian M, Tsiatis AA. 2002. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 58:742–53

Stan Development Team. 2016. `rstanarm`: Bayesian applied regression modeling via Stan. *R package version 2.13.1*

Suresh K, Taylor JMG, Spratt DE, Daignault S, Tsodikov A. 2017. Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometrical J.* 59:1277–300

Sylvestre MP, Abrahamowicz M. 2009. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Stat. Med.* 28:3437–53

Tang AM, Tang NS. 2015. Semiparametric Bayesian inference on skew–normal joint modeling of multivariate longitudinal and survival data. *Stat. Med.* 34:824–43

Tang NS, Tang AM, Pan DD. 2014. Semiparametric Bayesian joint models of multivariate longitudinal and survival data. *Comput. Stat. Data Anal.* 77:113–29

Taylor JM, Park Y, Ankerst DP, Proust-Lima C, Williams S, et al. 2013. Realtime individual predictions of prostate cancer recurrence using joint models. *Biometrics* 69:206–13

Thiébaut R, Jacqmin-Gadda H, Babiker A, Commenges D. 2005. Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Stat. Med.* 24:65–82

Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. 2017. Personalized schedules for surveillance of low risk prostate cancer patients. arXiv:1711.00285 [stat.AP]

Tsiatis AA, Davidian M. 2001. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 88:447–58

Tsiatis AA, Davidian M. 2004. Joint modeling of longitudinal and time-to-event data: an overview. *Stat. Sin.* 14:809–34

Umlauf N, Klein N, Zeileis A. 2017. BAMLSS: Bayesian additive models for location, scale and shape (and beyond). *J. Comput. Graph. Stat.* **https://doi.org/10.1080/10618600.2017.1407325**

Vacek PM. 1997. Assessing the effect of intensity when exposure varies over time. *Stat. Med.* 16:505–13

Van der Vaart AW. 1998. *Asymptotic Statistics*. Cambridge, UK: Cambridge Univ. Press

Wang Y, Taylor JMG. 2001. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Am. Stat. Assoc.* 96:895–905

Williamson P, Kolamunnage-Dona R, Philipson P, Marson A. 2008. Joint modelling of longitudinal and competing risks data. *Stat. Med.* 27:6426–38

Wulfsohn MS, Tsiatis AA. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics* 53:330–39

Xu J, Zeger SL. 2001. Joint analysis of longitudinal data comprising repeated measures and times to events. *J. R. Stat. Soc. C* 50:375–87

Yu M, Taylor JM, Sandler HM. 2008. Individualized prediction in prostate cancer studies using a joint longitudinal-survival-cure model. *J. Am. Stat. Assoc.* 103:178–87

Zeng D, Cai J. 2005. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. 33:2132–63

Zhang D, Chen MH, Ibrahim JG, Boye ME, Wang P, Shen W. 2014. Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Stat. Med.* 33:4715–33

Zhang D, Chen MH, Ibrahim JG, Boye ME, Wang P, Shen W. 2016. JMFit: A SAS macro for joint models of longitudinal and survival data. *J. Stat. Softw.* 71:1–24

Zhu H, Ibrahim JG, Chi YY, Tang N. 2012. Bayesian influence measures for joint models for longitudinal and survival data. *Biometrics* 68:954–64