Joint models of longitudinal and time-to-event data with more than one event time outcome: a review

Graeme L. Hickey[1], Pete Philipson[2], Andrea Jorgensen[1], Ruwanthi Kolamunnage-Dona*[1]

[1] Department of Biostatistics, University of Liverpool, Waterhouse Building, 1-5 Brownlow Street, Liverpool, L69 3GL, United Kingdom

[2] Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Ellison Place, Newcastle upon Tyne, NE1 8ST, United Kingdom

***Corresponding author**

Address: Department of Biostatistics, University of Liverpool,
Waterhouse Building, 1-5 Brownlow Street,
Liverpool, L69 3GL,
United Kingdom

Email: kdrr@liverpool.ac.uk

Telephone: +44(0)151 794 9736

FUNDING

This research was funded by Medical Research Council (MRC) grant MR/M013227/1.

ABSTRACT

Methodological development and clinical application of joint models of longitudinal and time-to-event outcomes have grown substantially over the past two decades. However, much of this research has concentrated on a single longitudinal outcome and a single event time outcome. In clinical and public health research, patients who are followed up over time may often experience multiple, recurrent, or a succession of clinical events. Models that utilise such multivariate event time outcomes are quite valuable in clinical decision-making. We comprehensively review the literature for implementation of joint models involving more than a single event time per subject. We consider the distributional and modelling assumptions, including the association structure, estimation approaches, software implementations, and clinical applications. Research into this area is proving highly promising, but to-date remains in its infancy.

Keywords: Joint models; multivariate data; longitudinal data; time-to-event data; recurrent events

1. INTRODUCTION

In clinical studies, measurements are often recorded about subjects at each follow-up visit; these response data give rise to longitudinal data. Subsequently, times to one or more clinically significant events are also recorded. The longitudinal data might be censored by one of these clinical events; for example, if the event was death or treatment failure. A growing field of research has emerged that seeks to jointly model these two outcomes —so-called *joint modelling*. When the outcome processes are correlated, joint modelling has been empirically demonstrated to lead to improved efficiency and reduced bias [1–3], improved prediction [4], and be applicable to outcome surrogacy [5]. The literature is extensive, with comprehensive reviews given by Hogan and Laird [6], Tsiatis and Davidian [7], Diggle et al. [8], Sousa [9], Proust-Lima et al. [10], and Gould et al. [11].

The classical joint model, from which most research has spawned, involves a single continuous longitudinal outcome and a single right-censored event time. Notwithstanding this simplicity, the joint modelling methodology has been recently extended to generalize both submodels. For the longitudinal submodel, developments include the incorporation of multiple outcomes [12], binary [13], count [14], and ordinal [15] outcomes, and extensions of the classical error and random effects distribution assumptions [16]. For the time-to-event submodel, extensions have involved the modelling of interval- [17] and left-censored [18] data, discrete event times [19], competing risks [20], parametric models [21], spline models [22], and subject- and institutional-level frailty effects [23]. Commensurate with this methodological research, there has been an increase in use of joint models in a wide –range of clinical settings [23–26] and development of several mainstream statistical software packages [27–34].

Due to current trends towards personalized medicine, models that utilise all available information more efficiently are of considerable value. In health research, patients may often experience multiple, recurrent, or a succession of clinical events, thus potentially admitting more than one event time. In this article, we comprehensively review the methodological literature for joint models involving multivariate event time data. Although the primary focus is on the ubiquitous shared random effects models, we also describe the growing framework of joint latent class models. Our review encapsulates multiple events, recurrent events (either in the presence of a terminal event, or not) and succession of events data. Although competing risks data can also be considered as multivariate time-to-event data, we do not review these models here as each subject still only admits a *single* event time. Furthermore, competing risks joint models have been extensively reviewed elsewhere in the joint model literature [35].

2. LONGITUDINAL DATA SUBMODELS

Let $Y_{ik}(t_{ijk})$ denote the j-th observed value of the k-th longitudinal outcome for subject i, measured at time t_{ijk}, for $i = 1, \ldots, N, k = 1, \ldots, K$, and $j = 1, \ldots, n_{ik}$. In some cases, only a single longitudinal outcome (i.e. $K = 1$) is considered, which greatly simplifies the model. We will consider both univariate ($K = 1$) and multivariate ($K > 1$) longitudinal data in this review, depending on the methodology presented in each article, but exclusively reserve the subscript k to denote multivariate cases.

In the framework of joint models involving more than one event time, the corresponding longitudinal measurements have predominantly been continuous. However, some models have considered binary and count data (**Table 1**). As noted earlier, some models have also considered multiple longitudinal outcomes. For a full review of joint models involving multivariate longitudinal outcomes, see Hickey et al. [12]. Król et al. [36] also considered left-censored longitudinal measurements, which is pertinent to biomarker measurements that involved minimum detection thresholds. There are a plethora of modelling approaches for multivariate longitudinal data [37]. In most cases, a generalized linear mixed model (GLMM) [38] is specified. Namely,

$$h_k\{\mathbb{E}[Y_{ik}(t_{ijk})]\} = \mu_{ik}(t_{ijk}), \quad \textbf{(1)}$$

where $h_k(\cdot)$ denotes a known one-to-one link function for the k-th outcome, $\mathbb{E}$ is the expectation operator, and $\mu_{ik}(\cdot)$ is the linear predictor:

$$\mu_{ik}(t_{ijk}) = X_{ik}^{(1)}(t_{ijk})^{\top}\beta_k^{(1)} + W_{1i}^{(k)}(t_{ijk}), \quad \textbf{(2.1)}$$

where

$$W_{1i}^{(k)}(t_{ijk}) = Z_{ik}(t_{ijk})^{\top} b_{ik}, \quad \textbf{(2.2)}$$

and $X_{ik}^{(1)}(t_{ijk})$ and $Z_{ik}(t_{ijk})$ are vectors of (possibly time-varying) covariates for subject i associated with fixed and random effects respectively, which can vary by outcome, $\beta_k^{(1)}$ is a vector of fixed effects parameters for the k-th outcome, and b_{ik} is a vector of subject-specific random effects for the k-th outcome. We denote the stacked vector of subject-specific random effects for all K outcomes by $b_i = \left(b_{i1}^T, b_{i2}^T, \ldots, b_{iK}^T\right)^T$. Some authors have considered including spline terms in $X_{ik}^{(1)}(t_{ijk})$ to capture complex functional forms between the outcome and measurement time [25,39]. On the other hand, Dantan et al. [40] specified a segmented GLMM with a random change-point, which was intrinsically linked to the time-to-event submodel through one of the transition hazard functions. Random change-points were shown to be particularly useful for capturing changes in the longitudinal trajectory of the outcome following a clinical (pre-)diagnosis.

Generally, for continuous longitudinal outcomes, independent and identically distributed normal errors are assumed. However, extensions to robust skew-normal distributed errors have also been proposed [39]. Subject-specific random effects are generally modelled as being multivariate normally distributed, reducing to a normal distribution in the case of a random-intercepts only model. Different modelling approaches have also been considered. Notably, Huang et al. [41] adopted discrete independent probability distributions. Njagi et al. [14] considered over-dispersed data, and proposed conjugate Beta and Gamma random effects for binary and count outcomes respectively. Several authors who considered multivariate longitudinal outcomes have proposed capturing the cross-sectional association between repeated measures through a correlated errors structure rather than a correlated random effects structure, i.e. $Y_{ik}(t_{ij}) = \mu_{ik}(t_{ij}) + \varepsilon_{ijk}$, with $\varepsilon_{ij.} \sim N_K(0, \Sigma)$ and $b_{ik} \sim N_{v_k}(0, \Psi_k)$ [39,42–44]. This allows for separate estimation of correlation between repeated measures and between different longitudinal outcomes.

In some cases, a semiparametric paradigm has been adopted. Within the Bayesian framework, Tang et al. [44] and Tang and Tang [39] assumed a Dirichlet process prior for the random effects, removing the need to assume a fixed parametric form, which is therefore robust to potential misspecification. Li et al. [45] suggested a time-dependent vector of random effects, which are independently and identically distributed according to an unknown multivariate distribution. The longitudinal submodels are also specified as marginal proportional rates models - namely, as **(1)** with $h_k(.)$ given by the exponential link function, and linear link functions are also suggested [46]; the time-dependent fixed effects are absorbed into an unspecified smooth baseline function.

Following Henderson et al. [47], an additional autocorrelation can be incorporated into the model by augmenting **(2.2)** to include a zero-mean stationary Gaussian process term. However, such models come with a substantially increased computational burden so it is not unexpected that very few methodological articles have considered this extension [48,49]. Zhang et al. [49], as well as considering correlation for $W_{1i}^{(k)}(t)$ in **(2.2)**, also allowed for correlation of errors within an outcome over time by letting $\varepsilon_{ik} = (\varepsilon_{i1k}, \dots, \varepsilon_{ink})^{\top}$ have zero-mean multivariate normal distribution with a *u*-lag correlation function given by

$$\rho_{1k}(\alpha_{1k}, u) = \exp\{-\alpha_{1k}|u|^{\delta}\}, \quad 0 < \delta \leq 2.$$

A summary of the longitudinal data submodels used in joint models involving multivariate time-to-event data is given in **Table 1**.

3. TIME-TO-EVENT DATA SUBMODELS

Let T_{ig}^{*} denote the g-th event time for the i-th subject ($i = 1, \dots, n$). Also, let C_i be a censoring time for the subject such that we actually observe $T_{ig} = \min\left(T_{ig}^{*}, C_i\right)$. Typically, continuous event times are observed. Two exceptions were Huang et al. [41], who considered discrete event times, and Rouanet et al. [50] who allowed one of the semi-competing event times to be interval-censored. For each subject i, let the vector $X_i^{(2)}(t)$, which may be time-varying, denote the observed covariate data, and $\beta_g^{(2)}$ denote the coefficient parameters associated with these covariates for the g-th event time. Similarly, for models involving a third submodel (e.g. a joint model of longitudinal data, recurrent and terminal events), we will use the notation $X_i^{(3)}(t)$ and $\beta^{(3)}$, as appropriate. However, in practice, there will be an overlap between baseline measurements in $X_i^{(1)}(t), X_i^{(2)}(t)$, and $X_i^{(3)}(t)$. Specification of the time-to-event model depends on the type of multivariate event time data and the association structure that gives rise to the joint model. These are described below and succinctly summarized in **Tables 2** and **3**. We will denote the association parameters by γ_g, and any extra random effects terms by θ_i.

3.1 Multiple events

Multiple (unordered) events occur when more than one event is observed, and interest lies with all of them. A joint model can be specified to capture the association between a longitudinal process and multiple failure times; for example, the time to cancer relapse in *two separate* organs.

Chi and Ibrahim [42] derived a novel yet complex bivariate survival model from first principles of latent precursor events modelled by a Poisson process. The model accommodates both zero and non-zero cure fractions, and the survival distribution is given by

$$S(t_{i1}, t_{i2}|\theta_i) = \exp\left\{-\theta_i\left[\int_0^{t_{i1}} \lambda_{i1}(u)F_1(t_{i1}-u)du + \int_0^{t_{i2}} \lambda_{i2}(u)F_2(t_{i2}-u)du\right]\right\},$$

where θ_i is a subject-specific frailty term that follows a positive stable law distribution indexed by the parameter ρ, which accounts for the correlation between the pair of event times, and $F_1(t)$ and $F_2(t)$ are distribution functions for the latent precursors, and later specified as exponential distributions. A current values parameterization was assumed to link the longitudinal and time-to-event submodels through

$$\lambda_{ig}(t) = \exp\left\{\sum_{k=1}^{K} \gamma_{gk}\mu_{ik}(t) + {X_i^{(2)}}^{\mathsf{T}}\beta_g^{(2)}\right\}.$$

It was noted that both the conditional and marginal survival function satisfies the proportional hazards property so long as the baseline covariates are modelled as per above, and $X_i^{(2)}$ is independent of time.

Zhu et al. [43], Tang et al. [44], and Tang and Tang [39] used the more ubiquitous piecewise constant proportional hazards model for the baseline hazard function, with knots placed at times v_{gq} $\{q = 1, \dots Q\}$ for the g-th time-to-event outcome, such that $0 = v_{g0} < v_{g1} < \dots < v_{gQ}$, with v_{gQ} being greater than $\max(T_{1g}, \dots, T_{ng})$; namely

$$\lambda_{0g}(t) = \sum_{q=1}^{Q} \xi_{qg} I\left(v_{g,q-1} < t \leq v_{gq}\right),$$

where $I(\cdot)$ denotes the indicator function, and ξ_{qg} denotes the value of the event-specific hazard function in the interval $(v_{g,q-1}, v_{gq}]$ for event g. The separate event time and longitudinal submodels are subsequently linked through a current values parameterisation:

$$\lambda_{ig}(t) = \lambda_{0g}(t) \exp\left\{\sum_{k=1}^{K} \gamma_{gk}\mu_{ik}(t) + {X_i^{(2)}}^{\top} \beta_g^{(2)}\right\}.$$

Huang et al. [41] adopted a discrete time hazard model of the form

$$\log\left(\frac{f_{ijg}}{S_{ijg}}\right) = X_i^{(2)}\left(t_j\right)^{\top} \beta_g^{(2)} + \gamma_g^{(1)}\eta_{ij} + \gamma_g^{(2)}\theta_i + \gamma_g^{(3)}\eta_{ij}x_i^{(3)},$$

where $f_{ijg} = P[T_{ig} = j]$, $S_{gij} = 1 - \sum_{j'=1}^{j} f_{ij'g}$ for discrete times t_j $(j = 1, \dots, J)$, and $\left\{\gamma_g^{(1)}, \gamma_g^{(2)}, \gamma_g^{(3)}\right\}$ are a set of association parameters. The first discrete random effect, η_{ij}, links the longitudinal submodel to the event process by a random effects parameterisation, which includes an interaction with one of the baseline covariates, $x_i^{(3)}$. The second discrete random effect, θ_i, captures additional association between the multivariate event times, beyond what is predicted by η_{ij}. An additional discrete multivariate distributed random effect was included in the multivariate longitudinal outcome submodel only.

3.2 Recurrent events

Recurrent (ordered) events occur when the same non-terminal event can be observed multiple times over a follow-up period. Henderson et al. [47] first presented a joint model compatible with recurrent events data, but this was ultimately simplified to the case of a single event time (i.e. a time to a single terminal event).

3.2.1 Without a terminal event. The simplest situation is when the recurrent events process is observed without a terminating process. For example, an epileptic patient can undergo multiple seizures in a day, and targeted treatments for epilepsy may be dependent on biomarker values [51]. A joint model of the recurrent events process and longitudinal outcomes data can capture this dependence.

Han et al. [51] adopted the general recurrent event model of Peña and Hollander [52] within a latent class framework, similar to that of Lin et al. [53], with the intensity function defined according to

$$r_i(t) = \theta_i r_{0r}(\mathcal{E}_i(t))\rho(N_i(t_-), a_r)\psi\left(X_i^{(2)}(t)^\top \beta^{(2)}\right),$$

where θ_i is a mean-one Gamma distributed frailty term, $r_{0r}(t)$ denotes the latent class-specific baseline intensity function (with $r = 1, \dots, R$), $\mathcal{E}_i(t)$ is the 'effective age' of subject i at time t, $N_i(t_-)$ is the effective number of accumulated events just prior to time t, $\rho(\cdot, a_r)$ is an event accumulation function parameterized by a_r, and $\psi\left(X_i^{(2)}(t)^\top \beta^{(2)}\right)$ is a function of the covariate linear predictor term, for example $\psi(x) = \exp(x)$, as in the aforementioned models. The 'effective age' is a predictable process that reflects the effect of interventions after each failure. In the simplest case, $\mathcal{E}_i(t) = t$, corresponding to a 'minimal repair'. At the other extreme, the 'effective age' may be reset to zero. The effective number of accumulated events is zero if a successful intervention is applied just prior to time t, else it equals the cumulative number of failures. The function $\rho(\cdot, a_r)$ captures the effect of recurrent events on the subject, which might be non-linear; for example, $\rho(n, a_r) = a_r^n$. The model specification is complete once a parametric distribution for $r_{0r}(t)$ is specified, which can be generalized to multiple families. The association between the longitudinal and event time processes is captured entirely through the latent class, with the class membership probabilities modelled according to a multinomial distribution. Although latent class models are distinct from shared random effects models, they can be considered as semiparametric analogues.

Njagi et al. [14] considered the Weibull-gamma-normal model for recurrent events. In short, this is a Weibull regression model conditional on independent random effects $b_i \sim N(0, D)$, as per the longitudinal submodel, and $\theta_{ig} \sim \Gamma(a, b)$, a frailty term such that the intensity function can be written as

$$r_i(t_{ig}) = \lambda_g \rho_g t_{ig}^{\rho_g - 1} \theta_{ig} \exp\left\{L_{ig} - \lambda_g t_{ig}^{\rho_g} \theta_{ig} \exp\{L_{ig}\}\right\},$$

where $L_{ig} = {X_{ig}^{(2)}}^\top \beta^{(2)} + \gamma_{ig}^\top b_i$, and γ_{ig} is a vector of scale factors. The association between the event time and longitudinal submodel is captured through the shared random effects b_i, and the correlation between the recurrent events is captured by the θ_{ig}. It was noted by the authors that this model encompasses shared and correlated random effects parameterisations. In the example, the authors impose further conditions; namely, $\rho_g \equiv \rho$, $\gamma_{ig} \equiv \gamma$, and $\theta_{ig} \equiv \theta_i \sim \Gamma(a, a^{-1})$ for identifiability purposes. Efendi et al. [54] also adopted a version of this model.

Shen et al. [48] proposed modelling the recurrent events as per the model formulation in Henderson et al. [47], namely through the intensity function

$$r_i(t) = r_0(t)\exp\left\{X_i^{(2)}(t)^\top \beta^{(2)} + W_{2i}(t)\right\},$$

where $r_0(t)$ is a baseline intensity function at time t, and $W_{2i}(t)$ is a zero-mean latent process term. In general, $W_{2i}(t) = Z_i^{(2)}(t)^T b_i + V_{2i}(t)$, where $V_{2i}(t)$ is a stationary Gaussian process. The model was simplified by specifying $W_{2i}(t) = \gamma_1 b_i + \gamma_2 V_{1i}(t)$, assuming $\mu_i(t) = X_i^{(1)}(t)^\top \beta^{(1)} + b_i + V_{1i}(t)$ for the longitudinal submodel, with $V_{1i}(t)$ a second stationary Gaussian process. However, the model was ultimately reframed as a *conditional rates function*, namely $\mathbb{E}[r_i(t) \mid Y_i]$, in order to exploit and extend an estimating equations methodology approach.

Zhang et al. [49] proposed a recurrent events model with two non-absorbing states, each with separate intensity functions. Essentially, this model is a special case of the multi-state model (discussed below), known as the illness-recovery model. For states $g = 1,2$, the intensity functions were defined as

$$r_i(t) = r_{0g}\exp\left\{X_i^{(2)}(t)^T \beta_g^{(2)} + W_{2ig}(t)\right\},$$

where the baseline intensity is constant, r_{0g}, and $W_{2ig}(t) = \gamma_{0g}\theta_i + \gamma_g W_{i1}(t)$ a zero-mean Gaussian process with *u*-lag correlation function

$$\rho_2(\alpha_2, u) = \exp\{-\alpha_2|u|^\delta\}, \quad 0 < \delta \leq 2,$$

with θ_i a normally distributed subject-specific random effect, and $W_{i1}(t) \equiv W_{i1}^{(k)}(t)$ for all k.

Li [55] proposed a joint model that assumed the same intensity model as per Liu et al. [56] (with $\gamma_1 = 0$; described below). However, the repeated binary measure was modelled using a discrete-time Markov model. A joint model was formed by factorizing the likelihood into a *selection model* [9], which lies outside the scope of this review.

***3.2.2 With a terminal event*.** A natural extension to the joint model of longitudinal outcome data and a recurrent events process is to consider the situation of a terminating event process; for example, time to death. In this scenario, a third type of submodel is required to capture this additional event time, which may also be associated with the longitudinal outcomes *and* the recurrent events process.

Liu and Huang [57] and Liu et al. [56] considered a recurrent events submodel with a separate terminal event submodel. A random effects parameterization was used in both the recurrent events intensity function, $r_i(t)$, and the terminal event hazard function, $\lambda_i(t)$:

$$r_i(t) = r_0(t)\exp\left\{X_i^{(2)}(t)^T\beta^{(2)} + \gamma_1 b_{i0} + \theta_i\right\},$$

$$\lambda_i(t) = \lambda_0(t)\exp\left\{X_i^{(3)}(t)^T\beta^{(3)} + \gamma_2 b_{i0} + \gamma_3\theta_i\right\}.$$

The standard model assumption of piecewise constant baseline hazards for $r_0(t)$ and $\lambda_0(t)$ was assumed. In addition, the terminal event submodel has a random effect parameterization linking it to the recurrent events submodel, where random effect term, θ_i, captures the correlation between recurrent events independent of b_i. Rizopoulos [38] described a similar model, but only briefly described the estimation procedure, and furthermore a clinical application was not provided to illustrate the model. Król et al. [36] also adopted this model, with some slight modifications. Firstly, the baseline intensity and hazard functions were approximated by cubic M-splines on Q-knots; namely

$$r_0(t) = \sum_{q=1}^{Q+2}\xi_{rq}M_q(t) \quad \text{and} \quad \lambda_0(t) = \sum_{q=1}^{Q+2}\xi_{\lambda q}M_q(t),$$

where $\{\xi_{rq}; q = 1, \dots, Q+2\}$ and $\{\xi_{\lambda q}; q = 1, \dots, Q+2\}$ are the spline coefficients for the baseline intensity and hazard functions, respectively, corresponding to M-spline basis functions, $M_q(t)$. Secondly, the association terms with the event time submodels and the longitudinal submodel were specified more flexibly as $\gamma_1^\top f_r\left(b_i, \beta^{(1)}, Z_i(t), X_i^{(1)}(t)\right)$ and $\gamma_2^\top f_\lambda\left(b_i, \beta^{(1)}, Z_i(t), X_i^{(1)}(t)\right)$. For example, $f_r(\cdot)$ and $f_\lambda(\cdot)$ might admit the current values or random effects parameterization.

Kim et al. [58] also proposed a joint model for a longitudinal outcome and a recurrent events process with a terminal event process. The recurrent events process, modelled using a broad class of transformation models, was linked by extra random effect terms θ_i, that are correlated with b_i,

$$r_i(t) = \frac{d}{dt}F_R\left(\int_0^t r_0(s)\exp\left\{X_i^{(2)}(s)^T\beta^{(2)} + Z_i^{(2)}(s)^T\theta_i\right\}ds\right),$$

with $\eta_i = (b_i^T,\ \theta_i^T)^T$ jointly distributed, and $F_R(\cdot)$ a specified transformation function. The terminal event submodel—again modelled using a transformation model—was associated with the longitudinal and recurrent events submodels through a random effects parameterization with interaction with (possibly time-varying) subject-specific covariates:

$$\lambda_i(t) = \frac{d}{dt}F_T\left(\int_0^t \lambda_0(s)\exp\left\{X_i^{(3)}(s)^T\beta^{(3)} + Z_i^{(3)}(s)^T\gamma^T\eta_i\right\}ds\right),$$

with $F_T(\cdot)$ a separate specified transformation function. The authors explicitly used the logarithmic and Box-Cox transformation models for analysis in their data application. The baseline functions $r_0(t)$ and $\lambda_0(t)$ were modelled semiparametrically, with mass at each unique observed event time.

3.2.3 As a device for informative observation times. Joint models are usually based on the assumption of non-informative observation times for the repeated measurement process. This is generally reasonable for randomized control trials, but perhaps not so for observational data studies, where sicker patients (possibly indicated through their longitudinal measurement data) present more frequently to their physician, and whom are more likely to experience an event. Several models have been proposed to account for this potentially informative observational times protocol, which fall under the umbrella of joint models of longitudinal data and recurrent events, either with or without a separate terminal event process. In fact, the model by Liu et al. [56] was motivated by this situation, but the subject-specific shared random effects model is widely applicable to other data. This emerging field of joint modelling has its own substantive and rapidly growing literature, but clearly warrants a discussion here. In the interests of brevity, we do not review the entire literature on this particular joint model, and instead illustrate the ideas through the model proposed by Li et al. [45], which is representative of the model specification and estimation methodology in the literature. Readers should consult Li et al. [59], Han et al. [60], and references therein for more details on this model framework.

Working within a semiparametric framework, a flexible proportional rates marginal model for the observation (recurrent events) process was specified by Li et al. [45]; namely

$$E\left[dN_i(t) \mid X_i^{(3)}, b_i(t)\right] = \exp\left\{{X_i^{(3)}}^{\top}\beta^{(3)} + b_{i3}(t)\right\} dr_0(t),$$

where $dr_0(t)$ is an unknown baseline rate function, and $b_i(t) = \left(b_{i1}(t), b_{i2}(t), b_{i3}(t)\right)^T$ is a vector of possibly correlated subject-specific time-dependent random effects with 3 components corresponding to the longitudinal measurements, terminal event and recurrent events, respectively. The terminal event was modelled as a semiparametric additive hazards model [45], namely,

$$\lambda_i(t) = \lambda_0(t) + {X_i^{(2)}}^{\top}\beta^{(2)} + b_{i2}(t),$$

with the baseline hazard $\lambda_0(t)$ left unspecified; however, parametric and semiparametric proportional hazards regression models could also be integrated into this framework [46,61]. Association between the submodels is induced through the joint distribution of $b_i(t)$.

3.2.4 Multiple recurrent events. Musoro et al. [25] were motivated to unify both multiple and recurrent event types (**Sections 3.1 and 3.2**) into a single joint model. For G multiple event outcomes, which can be recurrent, they specified an intensity model

$$\lambda_{ig}(t) = \lambda_{0g}(t)\exp\left\{\sum_{k=1}^{K}\gamma_{gk}\mu_{ik}(t) + X_i^{(2)}\beta_g^{(2)} + \theta_{ig} + \psi_i\right\},$$

where θ_{ig} and ψ_i are zero-mean independent Gaussian random effect terms that account for within and between event types, respectively. As above, $\lambda_{0g}(t)$ was modelled semiparametrically.

3.3 Succession of events

A *succession* of events occurs when non-fatal events can precede an absorbing state event, e.g. death. The intermediate events provide information on the disease progression, and can be viewed as transitions from one state to another. Multistate models provide a framework for analysing this data [62]. Longitudinal measurements that are collected over time may have different associations with progression between separate health states. We also note that multistate models can also be viewed as an extension of the competing risks model framework, where interest continues after the first event. Joint models of longitudinal data and standard competing risks data are described elsewhere [35].

Multistate models have also been applied in what is essentially the univariate event time joint modelling framework. For example, Deslandes and Chevret [63] discretized the longitudinal outcome space to form states that were combined with the event. However, clinical events of interest—disease progression or death—were combined into a single composite event. Hu et al. [64] also considered a multistate model where the longitudinal outcome was discretized according to quartiles to form transition states, augmented with additional states defined by competing risks data. Neither of these two articles considered an actual *succession of event times*, and therefore are not discussed further. Le Cessie et al. [65] adopted a simple model where hazard functions for disease state transitions were estimated using separate Cox proportional hazards regression models. However, the joint model was effectively constructed through a type of *pattern mixture* model, in which the conditional responses per disease state were estimated using a generalized estimating equations framework, and the disease state probabilities were combined to estimate the marginal mean response over time. Pattern mixture models (and similarly, selection models) have their own dedicated literature in the model-based literature [9].

Ferrer et al. [66] proposed a Markovian multi-state transition submodel with proportional hazards, such that the transition intensity at time t from state g to h is

$$\lambda_{igh}(t) = \lambda_{0gh}(t) \exp\left\{ {X_{ghi}^{(2)}}^{\top} \beta_{gh}^{(2)} + \gamma_{gh}^{\top} f_{gh}\left(b_i, \beta^{(1)}, Z_i(t), X_i^{(1)}(t)\right)\right\}, \quad \textbf{(3)}$$

where the baseline intensity function $\lambda_{0gh}(t)$ can be specified as a Weibull, piecewise constant, or B-splines function, and γ_{gh} are transition-specific parameters corresponding to $f_{gh}(\cdot)$—a flexible association function that links the multistate submodel to the longitudinal data submodel by any function of the random effects. Special cases include the current values parameterization, the

random slopes parameterization, and a linear combination of both aforementioned parameterizations.

Dantan et al. [40] proposed a multi-state model with transition between states specified as per **(3)**, subject to the association structures $f_{01}(\cdot) = 0$, $f_{12}(\cdot)$ a random effects parameterization, and $f_{g3}(\cdot)$ a current values parameterization, for $g = 0,1,2$, and other transitions were discounted. In addition, the baseline hazards were defined by Weibull distributions for the non-absorbing transitions, and a piecewise constant function for all transitions to the absorbing (death) state. Dantan and colleagues also extended the model to incorporate left-truncation to account for subjects already in the disease state entering the study late.

As noted earlier, competing risks data can be viewed as a special case of multistate models. In the context of multiple event times data, semi-competing risks model is of most interest. In this situation, a terminal event censors a non-terminal event, but not *vice versa*; hence, it is possible to observe more than one event time. Rouanet et al. [50] proposed two joint models for this data within a latent class framework. The first was a Markovian multi-state (or illness-death) model, as per above, with

$$\lambda_{ghri}(t) = \lambda_{ghr0}(t) \exp\left\{X_{ghi}^{(2)\top} \beta_{ghr}^{(2)}\right\},$$

where $\lambda_{ghr0}(t)$ is a baseline intensity function for the transition from states g to h in latent class r (modelled as either a Weibull function or using M-splines), and $\beta_{ghr}^{(2)}$ are class and transition-specific parameters corresponding to baseline covariates $X_{ghi}^{(2)}$. The second was a semi-Markovian model, where one specific transition (from illness to death) depends on the time spent in the illness state, i.e. $\lambda_{12ri}(t - T_{i1})$, as opposed to just the time elapsed. As per other latent class models, the association between the submodels is captured entirely through the latent classes, with class membership modelled separately.

4. MODEL ESTIMATION

Several different estimation approaches have been utilized to fit the models described above (**Table 3**). Loosely, these methods can be separated as either likelihood maximisation or Bayesian model fitting.

Extending the original joint model developments of Wulfsohn and Tsiatis [67], the expectation-maximization algorithm has been used in some cases. In the case of Han et al. [51], the latent class membership, longitudinal data submodel random effects, and the time-to-event submodel frailty terms were treated as missing data. In the case of Kim et al. [58], only the random

effects were treated as missing data, and recursive formulae used to reduce the number of model parameters required for estimation. Król et al. [36] used penalized maximum likelihood estimation using the Marquardt algorithm, with the penalization performed to obtain smooth estimates of the baseline hazard and intensity functions. Rouanet et al. [50] also utilized the Marquardt algorithm, with the number of latent classes selected according to the Bayesian Information Criterion. Dantan et al. [40] reported using a Newton-Raphson-like algorithm. Huang et al. [41] used automatic differentiation—a numerical technique for simultaneously evaluating a function and its derivatives—with a Newton-Raphson algorithm, which was purportedly faster than the EM algorithm. Njagi et al. [14] and Efendi et al. [54] used a partial marginalisation approach [68] whereby the conjugate random effects are analytically integrated out, and the normal random effects are numerically integrated using standard software. Efendi et al. [54] then exploited the ideas of Heagerty and Zeger [69] to establish marginal effects. Liu et al. [56] and Liu and Huang [57] reported using numerical likelihood maximisation via standard software. Standard errors of all these aforementioned model fits can be estimated from the inverse of the observed information matrix; however, Han et al. [51] reported using the bootstrap method.

Zhang et al. [49] proposed a two-stage estimation strategy. In the first stage, the covariance parameters were estimated from the repeated measures marginal likelihood function, with the mean function estimated by a weighted moving average. In the second stage, the expected likelihood function for the time-to-event data were maximized by an EM algorithm, with Gibbs sampling implemented for the high-dimension numerical integration, and a Newton-Raphson step used for the M-step. Shen et al. [48] developed a two-stage conditional estimating equations approach for model fitting, followed by a bootstrap approach for standard error estimating. As a precursory step, the authors reframed the time-to-event submodel from an intensity function to a conditional rate function. For models that accounted for informative observation times, generalized estimating equations in a semiparametric framework was the standard approach , which yielded consistent estimators [45,46,61]. In these cases, theoretical results have been derived on the asymptotic normality, which is subsequently used to make inference on the estimated parameters.

Bayesian estimation of standard univariate joint models has seen increased attention over recent years [28,30], especially as it is a natural tool for dynamic prediction and model averaging [4]. Moreover, there are multiple disadvantages to the ubiquitous frequentist estimation approach, including but not limited to, computational challenges—something one would expect to be particularly burdensome in a multivariate framework, the dependence on asymptotic approximations, and the complexity of model assessment and comparison. In joint models involving multivariate longitudinal data, Liu and Li [70] compared the performance of Bayesian approaches to

maximum likelihood approaches under different strengths of association, and demonstrated superiority of the Bayesian methods with respect to bias, root-mean square error, and coverage. Of the joint models involving multivariate event time data that were estimated using Bayesian statistics [25,39,42–44], Markov chain Monte Carlo (MCMC) methods were employed in all cases with default non-informative prior distributions chosen for the parameters. As noted earlier, Tang et al. [44] and Tang and Tang [39] also assumed a Dirichlet process prior for the random effects, removing the need to assume a fixed parametric form, which is therefore robust to potential misspecification. Tang and Tang [39] explored the sensitivity of results to prior distribution inputs, showing that good prior knowledge led to marginally improved estimation. The Gibbs sampling algorithm was used in all cases, with non-standard conditionals sampled using adaptive rejection or Metropolis-Hasting algorithms. Chi and Ibrahim [42] specifically noted that hierarchical centring [71], as well as some parameter transformations were used to facilitate convergence of the MCMC algorithms. The posterior conditional distributions for each parameter were derived analytically by all authors, except Musoro et al. [25], who exploited the automation provided by the OpenBUGS software. In all cases, assessment of convergence was made using general diagnostic methods; for example, examination of trace plots, autocorrelation plots, and the Gelman-Rubin statistics [72].

5. SOFTWARE

The ability to fit the models discussed is severely limited by the availability of software packages or modifiable code. Several authors have made code available either in an appendix or online as a supplement or via an online code repository system (**Table 3**). However, many authors do not report what software was used, or make said code available. Only one article released their code in the form of a software package, namely Król et al. [36], which fits a joint model for a single longitudinal outcome, a recurrent events process, and a single terminal event, and which is available through the `trivPenal()` function in the R package `frailtypack` [73].

6. CLINICAL APPLICATIONS

Development of novel methodology of joint models of longitudinal data and multivariate event times data have predominantly been motivated by real-world clinical datasets. Here, we summarize the applications that have led to the models discussed in this review.

6.1 Multiple events

Chi and Ibrahim [42] were interested in assessing whether four different quality of life measures (appetite, mood, coping, and physical wellbeing) were prognostic and predictive of breast cancer progression in a drug randomized controlled trial (RCT). The study monitored patients concerning two different failure times: death and cancer recurrence. A joint model was constructed

to model these 4 longitudinal outcomes and 2 event time outcomes. Tang et al. [44], Tang and Tang [39], and Zhu et al. [43] each proposed multiple event joint models as per above, motivated by the same objectives and breast cancer dataset described above, but with novel model innovations including semiparametric Bayesian random effects modelling, robust errors, different association structures, and event-time submodels. Musoro et al. [25] considered a case of multiple recurrent events, where each patient could become repeatedly infected with one of 9 different infections (including upper respiratory, fungal, and parasitic infections) following kidney transplantation surgery. The objective of the study was to evaluate the effect of 4 repeatedly measured immune system biomarkers (CD4+ T cells, CD8+ T cells, natural killer cells, and B cells) on the risk of each infection type in a single joint model of multiple recurrent events and multivariate longitudinal data. This particular clinical application also falls under the umbrella of multiple events *and* recurrent events (below). Huang et al. [41] analyzed data from a complex prevention trial, with an interest on whether different interventions were associated with times to initiation of alcohol use and tobacco use. It was hypothesized that a psychiatric distress latent variable, which is reflected in multiple repeatedly measured mental health items, affects substance initiation; hence, a joint model was constructed.

6.2 Recurrent events

Njagi et al. [14] and Efendi et al. [54] were interested in jointly modelling the recurrent time to re-hospitalization and a repeated measure of heart rate from the same dataset of patients with chronic heart failure who were discharged from hospital. Efendi et al. [54] modelled heart rate as a continuous outcome, whereas Njagi et al. [14] modelled it as a count response based on the number of times the heart rate was classified as 'abnormal'. Han et al. [51] considered repeated times to seizure in an epilepsy cohort study. Serial blood measures were also recorded for 3 blood plasma lipids; however, based on clinical knowledge, a single longitudinal outcome was constructed from 2 of the biomarkers by taking a ratio at each measurement time; —the lecithin–cholesterol ratio, with the third biomarker discounted, as this ratio was believed to be elevated during periods of the day when seizures occurred. Shen et al. [48] jointly modelled time to cocaine-use relapse, a recurrent events outcome, and a repeated measure of psychiatric symptoms used to assess stress and cocaine craving levels in patients enrolled in a clinical intervention study. The primary objective was to understand whether the randomly assigned intervention (contingency management or not) treatment affects either stress or drug relapse after adjustment for demographic variables. Zhang et al. [49] were interested in investigating the health effects of air quality on respiratory symptoms. Four measures of air quality were recorded daily, as were three symptoms recorded per subject (runny nose, cough, sore throat / general sickness). Each day, subjects could be in either a

symptomatic or asymptomatic state, which they transition between (i.e. an illness-recovery model). For each symptom in turn, a recurrent events joint model with the 4 longitudinal measures was fitted.

Liu and Huang [57] hypothesized that repeatedly high CD4 cell counts in HIV positive patients are associated with low risk of opportunistic disease, which is a potentially recurring event. They further hypothesized that a higher CD4 cell count and lower rate of opportunistic disease are associated with better survival, which is a terminal event. The interplay between these three processes might, however, be motivated by different application-specific reasons. Similarly, Kim et al. [58] modelled the recurrent time to a coronary heart disease event and time to death with repeated measurements on systolic blood pressure in patients previously diagnosed with hypertension. Within the context of a clinical trial for metastatic colorectal cancer, Król et al [36] were interested in the predictive ability of tumour size (a possibly left-censored repeated measurement), and the recurrent appearance of new lesions and the terminal outcome death.

Recurrent events are a particularly attractive modelling component for observational studies. Namely, when the follow-up protocol is not pre-specified or random, one might expect that the sickest subjects are those both more likely to experience the event of interest, as well as visit their physician more regularly where they will have biomarker measurements recorded. A recurrent events process can therefore be used to account for the correlation between observation times and repeated measures process. This was the case in Liu et al. [56], who considered recurrent times to hospital visits for diagnosis or treatment of heart failure alongside time to death, with repeated measurements on medical costs. Data from a skin cancer clinical trial was analyzed in a similar fashion by Li et al. [45], with the number of observed tumours at each observation time modelled as the longitudinal outcome.

6.3 Succession of events

Ferrer et al. [66] analyzed data from a multi-centre clinical trial treated with external beam radiotherapy for localized prostate cancer. Prostate-specific antigen (PSA) was repeatedly measured during follow-up. In addition, times of transitions between different disease states were recorded: radiotherapy cessation, local recurrence, distant recurrence, initiation of hormonal therapy, and death. The association between PSA and clinical relapse is well-known from univariate joint models; however, it is also of value to clinicians and patients to be able to distinguish between the different phases of disease progression as PSA may be differently correlated at each stage.

Rouanet et al. [50] analyzed a cohort study of patients to model pre-dementia cognitive decline, as measured by a psychometric test score to assess verbal fluency, in the presence of semi-

competing risks of dementia onset and death. That is, the risk of dementia is null after death has occurred, but death can occur after dementia. As the diagnosis of dementia cannot be precisely recorded due to intermittent assessment, it is interval-censored, thus known to have occurred between two follow-up appointments. It is important to account for that this interval is known as the risk of dementia may be underestimated otherwise. Using data from the same cohort study, Dantan et al. [40] also analyzed the dependency of cognitive ageing—repeatedly measured using a psychometric test used to assess cognitive ability—on the progression from healthy, pre-diagnosis, illness, and death states. A fundamental difference of the latter model compared to the former is that an interim 'pre-diagnosis' state was included, which was modelled by a segmented linear mixed model with a random change point.

7. DISCUSSION

The case for use of joint models has been made already [1,74,75]. Namely, when the longitudinal and event time processes are correlated they reduce the bias obtained from simpler methods, including separate models (e.g. separate LMMs, survival models, recurrent event models, and multistate models), or even the two-stage approach. There has been a myriad of extensions in the joint modelling framework over the past few years, including extensions to multivariate longitudinal data [12] and competing risks data [35]. Relatively fewer developments have been made pertaining joint models involving more than a single event time, which includes multiple events, recurrent events, and a succession of events. Yet, as shown, there are wide-ranging clinical applications for these models. In particular, motivation has stemmed from disease areas representing cancer, infection, cardiovascular disease, neurological disease, mental health, and respiratory disease. Moreover, data were derived from both randomized controlled trials and cohort studies.

The review presented here contributes to this narrow but important topic in joint models by bringing together in a single place and juxtaposing the models and distributional assumptions, outcome types, estimation and software implementations alongside clinical applications. This is a research area of growing interest and clinical importance, and the extensions developed are necessary to appropriately analyze this complex data. However, we found that availability of mainstream statistical software to fit these models is severely limited, and this will ultimately pose problems, since the complexity of the models means that *ad hoc* programming is required. This is not unexpected as joint models are computationally difficult to fit; a problem that is exacerbated by the extension to joint models involving more than a single event time. In fact, Musoro and colleagues noted that their ambitious attempt to fit a model to 4 longitudinal outcomes and 9

recurrent event outcome types was precluded by computational time; development of approaches that reduce this computational burden are therefore of paramount importance.

The extension of joint models to more than a single event time offers not only improved inference, but also opportunity for dynamic prediction. This has received growing interest in the classical joint model framework [4], but less so in the extension of multivariate event time data. Król et al. [36] developed dynamic prediction and predictive assessment tools for their recurrent events joint model. Others have also discussed prediction in the context of joint models involving multivariate event time data [14,50,66]. Dynamic prediction is easily encompassed in a Bayesian joint model framework. Despite this, the use of Bayesian methods for model fitting has been rather limited in the methodological developments of joint models involving multivariate event time data. Moreover, there is also limited research on the role of prior distribution selection. Research to-date has been predominantly technical, and more attention is required on the interpretability of these models in clinical applications. Moreover, the complexity of these models requires further development on diagnostics that will facilitate model selection, including the choice of association structure.

FUNDING

GLH was funded by Medical Research Council (MRC) grant MR/M013227/1.

REFERENCES

1. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. J Clin Oncol. 2010;28: 2796–2801.
2. Chen LM, Ibrahim JG, Chu H. Sample size and power determination in joint modeling of longitudinal and survival data. Stat Med. 2011;30: 2295–2309.
3. Hogan JW, Laird NM. Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. Stat Methods Med Res. 1998;7: 28–48.
4. Rizopoulos D, Hatfield LA, Carlin BP, Takkenberg JJM. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. J Am Stat Assoc. 2014;109: 1385–1397.
5. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error - applications to survival and CD4 counts in patients with AIDS. J Am Stat Assoc. 1995;90: 27–37.
6. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. Stat Med. 1997;16: 259–272.
7. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin. 2004;14: 809–834.
8. Diggle PJ, Sousa I, Chetwynd AG. Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. Stat Med. 2008;27: 2981–2998.
9. Sousa I. A review on joint modelling of longitudinal measurements and time-to-event. Revstat Stat J. 2011;9: 57–81.
10. Proust-Lima C, Sene M, Taylor JMG, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. Stat Methods Med Res. 2012;23: 74–90.
11. Gould AL, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, et al. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. Stat Med. 2015;34: 2181–2195.
12. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med Res Methodol. BMC Medical Research Methodology; 2016;16: 1–15.
13. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat Med. 2011;30: 1366–1380.
14. Njagi EN, Molenberghs G, Rizopoulos D, Verbeke G, Kenward MG, Dendale P, et al. A flexible joint modeling framework for longitudinal and time-to-event data with overdispersion. Stat Methods Med Res. 2013;0: 1–16.

15. Li N, Elashoff RM, Li G, Saver J. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. Stat Med. 2010;29: 546–557.
16. Huang X, Li G, Elashoff RM, Pan J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. Lifetime Data Anal. 2011;17: 80–100.
17. Gueorguieva R, Rosenheck R, Lin H. Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. J R Stat Soc Ser A Stat Soc. 2012;175: 417–433.
18. Thiébaut R, Jacqmin-Gadda H, Babiker AG, Commenges D. 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. 2005;24: 65–82.
19. Albert PS, Shih JH. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. Ann Appl Stat. 2010;4: 1517–1532.
20. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG. Joint modelling of longitudinal and competing risks data. Stat Med. 2008;27: 6426–6438.
21. Crowther MJ, Abrams KR, Lambert PC. Flexible parametric joint modelling of longitudinal and survival data. Stat Med. 2012;31: 4456–4471. doi:10.1002/sim.5644
22. Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics. 2005;61: 64–73.
23. Luo S, Wang J. Bayesian hierarchical model for multiple repeated measures and survival data: an application to Parkinson's disease. Stat Med. 2014;33: 4279–4291.
24. Touloumi G, Pantazis N, Babiker AG, Walker SA, Katsarou O, Karafoulidou A, et al. Differences in HIV RNA levels before the initiation of antiretroviral therapy among 1864 individuals with known HIV-1 seroconversion dates. AIDS. 2004;18: 1697–1705.
25. Musoro JZ, Geskus RB, Zwinderman AH. 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. 2015;57: 185–200.
26. Ibrahim JG, Chen M-H, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. Stat Sin. 2004;14: 863–883.
27. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. Stata J. 2013;13: 165–184.
28. Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. Am Stat. 2004;58: 16–24.

29. Philipson P, Sousa I, Diggle PJ, Williamson PR, Kolamunnage-Dona R, Henderson R. Package "joineR." R Foundation for Statistical Computing; 2012.
30. Rizopoulos D. The R Package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. J Stat Softw. 2016;72: 1–45.
31. Zhang D, Chen M-H, Ibrahim JG, Boye ME, Shen W. JMFit: a SAS macro for joint models of longitudinal and survival data. J Stat Softw. 2016;71.
32. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. J Stat Softw. 2010;35: 1–33.
33. Proust-Lima C, Philipps V, Liquet B. Estimation of latent class linear mixed models: the new package lcmm. arXiv Prepr. 2015;
34. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. joineRML: Joint Modelling of Multivariate Longitudinal Data and Time-to-Event Outcomes. Comprehensive R Archive Network; 2017.
35. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. A comparison of different joint models for longitudinal and competing risks data: with application to an epilepsy drug randomised control trial. J R Stat Soc Ser A Stat Soc. 2017;
36. Król A, Ferrer L, Pignon JP, Proust-Lima C, Ducreux M, Bouché O, et al. Joint model for left-censored longitudinal data, recurrent events and terminal event: predictive abilities of tumor burden for cancer evolution with application to the FFCD 2000-05 trial. Biometrics. 2016;72: 907–916.
37. Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: A review. Stat Methods Med Res. 2012;23: 42–59.
38. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Boca Raton, FL: Chapman & Hall/CRC; 2012.
39. Tang AM, Tang NS. Semiparametric Bayesian inference on skew–normal joint modeling of multivariate longitudinal and survival data. Stat Med. 2015;34: 824–843.
40. Dantan E, Joly P, Dartigues J-F, Jacqmin-Gadda H. Joint model with latent state for longitudinal and multistate data. Biostatistics. 2012;12: 723–736.
41. Huang W, Zeger SL, Anthony JC, Garrett E. Latent variable model for joint analysis of multiple repeated measures and bivariate event times. J Am Stat Assoc. 2001;96: 906–914.
42. Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. Biometrics. 2006;62: 432–445.
43. Zhu H, Ibrahim JG, Chi YY, Tang NS. Bayesian influence measures for joint models for longitudinal and survival data. Biometrics. 2012;68: 954–964.

658 44. Tang NS, Tang AM, Pan DD. Semiparametric Bayesian joint models of multivariate
659 longitudinal and survival data. Comput Stat Data Anal. 2014;77: 113–129.
660 45. Li Y, He X, Wang H, Sun J. Regression analysis of longitudinal data with correlated censoring
661 and observation times. Lifetime Data Anal. 2016;22: 343–362.
662 46. Sun L, Song X, Zhou J, Liu L. Joint analysis of longitudinal data with informative observation
663 times and a dependent terminal event. J Am Stat Assoc. 2012;107: 688–700.
664 47. Henderson R, Diggle PJ, Dobson A. Joint modelling of longitudinal measurements and event
665 time data. Biostatistics. 2000;1: 465–480.
666 48. Shen Y, Huang H, Guan Y. A conditional estimating equation approach for recurrent event
667 data with additional longitudinal information. Stat Med. 2016;
668 49. Zhang H, Ye Y, Diggle PJ, Shi J. Joint modeling of time series measures and recurrent events
669 and analysis of the effects of air quality on respiratory symptoms. J Am Stat Assoc. 2008;103:
670 48–60.
671 50. Rouanet A, Joly P, Dartigues J-F, Proust-Lima C, Jacqmin-Gadda H. Joint latent class model for
672 longitudinal data and interval-censored semi-competing events: application to alzheimer's
673 disease. Biometrics. 2016; 1–22.
674 51. Han J, Slate EH, Pena EA. Parametric latent class joint model for a longitudinal biomarker and
675 recurrent events. Stat Med. 2007;26: 5285–5302.
676 52. Pena EA, Hollander M. Models for recurrent events in reliability and survival analysis. In:
677 Soyer R, Mazzuchi T, Singpurwalla N, editors. Mathematical Reliability: An Expository
678 Perspective. Dordrecht: Kluwer Academic Publishers; 2004.
679 53. Lin H, Turnbull BW, McCulloch CE, Slate EH. Latent class models for joint analysis of
680 longitudinal biomarker and event process data: application to longitudinal prostate-specific
681 antigen readings and prostate cancer. J Am Stat Assoc. 2002;97: 53–65.
682 54. Efendi A, Molenberghs G, Njagi EN, Dendale P. A joint model for longitudinal continuous and
683 time-to-event outcomes with direct marginal interpretation. Biometrical J. 2013;55: 572–588.
684 55. Li S. Joint modeling of recurrent event processes and intermittently observed time-varying
685 binary covariate processes. Lifetime Data Anal. Springer US; 2016;22: 145–160.
686 56. Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative
687 observational times and a dependent terminal event, with application to medical cost data.
688 Biometrics. 2008;64: 950–958.
689 57. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes
690 in the presence of death, with application to a study on acquired immune deficiency
691 syndrome. J R Stat Soc Ser C Appl Stat. 2009;58: 65–81.

692 58. Kim S, Zeng D, Chambless L, Li Y. Joint models of longitudinal data and recurrent events with
693 informative terminal event. Stat Biosci. 2012;4: 262–281.
694 59. Li Y, He X, Wang H, Sun J. Joint analysis of longitudinal data and informative observation
695 times with time-dependent random effects. In: Jin Z, Liu M, Luo X, editors. New
696 Developments in Statistical Modeling, Inference and Application. Switzerland: Springer
697 International Publishing; 2016.
698 60. Han M, Song X, Sun L. Joint modeling of longitudinal data with informative observation times
699 and dropouts. Stat Sin. 2014;24: 1487–1504.
700 61. Miao R, Chen X, Sun L. Analyzing longitudinal data with informative observation and terminal
701 event times. Acta Math Appl Sin. 2016;32: 1035–1052.
702 62. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state
703 models. Stat Med. 2007;26: 2389–2430.
704 63. Deslandes E, Chevret S. Assessing surrogacy from the joint modelling of multivariate
705 longitudinal data and survival: application to clinical trial data on chronic lymphocytic
706 leukaemia. Stat Med. 2007;26: 5411–5421.
707 64. Hu B, Li L, Wang X, Greene T. Nonparametric multistate representations of survival and
708 longitudinal data with measurement error. Stat Med. 2012;31: 2303–17.
709 65. le Cessie S, de Vries EGE, Buijs C, Post WJ. Analyzing longitudinal data with patients in
710 different disease states during follow-up and death as final state. Stat Med. 2009;28: 3829–
711 3843.
712 66. Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin-Gadda H, Proust-Lima C. Joint modelling of
713 longitudinal and multi-state processes: application to clinical progressions in prostate cancer.
714 Stat Med. 2016;
715 67. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with
716 error. Biometrics. 1997;53: 330–339.
717 68. Molenberghs G, Verbeke G, Demétrio CGB. An extended random-effects approach to
718 modeling repeated, overdispersed count data. Lifetime Data Anal. 2007;13: 513–531.
719 69. Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference. Stat Sci.
720 2000;15: 1–26.
721 70. Liu F, Li Q. A Bayesian model for joint analysis of multivariate repeated measures and time to
722 event data in crossover trials. Stat Methods Med Res. 2014;0: 1–13.
723 71. Gelfand AE, Sahu SK, Carlin BP. Efficient parameterizations for normal linear mixed models.
724 72. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd
725 Editio. Boca Raton, FL: CRC Press; 2013.

73. Rondeau V, Gonzalez JR, Mazroui Y, Mauguen A, Król A, Diakite A, et al. R package "frailtypack." 2016.
74. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical J. 2011;53: 750–763.
75. Wang P, Shen W, Boye ME. Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. Heal Serv Outcomes Res Methodol. 2012;12: 182–199.
76. Molenberghs G, Verbeke G, Demétrio CGB, Vieira AMC. A family of generalized linear models for repeated measures with normal and conjugate random effects. Stat Sci. 2011;25: 325–347.

Table 1. Summary of longitudinal submodels.

Article	Ref.	Multivariate	Outcome types	Model	Error distribution	Random effects distribution
Multiple events						
Huang et al. (2001)	[41]	Yes	Binary	• Logistic regression model given the latent variable • Marginal log-odds model for the longitudinal latent process	N/A	Discrete independent probability distributions
Chi & Ibrahim (2006)	[42]	Yes	Continuous	LMM	MVN	MVN
Zhang et al. (2008)	[49]	Yes	Continuous	LMM	MVN – stationary Gaussian process with exponential correlation	MVN – stationary Gaussian process with exponential correlation
Zhu et al. (2012)	[43]	Yes	Continuous and/or discrete	GLMM	MVN for continuous outcomes	MVN
Tang et al. (2014)	[44]	Yes	Continuous and/or discrete	GLMM	MVN for continuous outcomes	Unspecified distribution modelled with a Dirichlet process prior (with MVN base distribution)
Tang & Tang (2015)	[39]	Yes	Continuous	LMM + P-splines	Multivariate skew-normal	Unspecified distribution modelled with a Dirichlet process prior (with MVN base distribution)
Multiple events + recurrent events						
Musoro et al. (2015)	[25]	Yes	Continuous	LMM + thin-plate splines	Normal	MVN + normal for thin-plate spline effects
Recurrent events						
Han et al. (2007)	[51]	No	Continuous	LMM	Normal	MVN
Liu et al. (2008)	[56]	No	Continuous	LMM	Normal	Normal
Liu & Huang (2009)	[57]	No	Continuous	LMM	Normal	Normal
Kim et al. (2012)	[58]	No	Continuous	LMM	Normal	MVN
Efendi et al. (2013)	[54]	No	Continuous	LMM	Normal	MVN
Njagi et al. (2013)	[14]	No	Continuous, binary or count	• LMM for continuous outcomes • Probit for binary outcomes • Poisson for count outcomes	Normal for continuous outcomes	MVN Separate Beta or Gamma random effects for binary or count outcomes,

						respectively
Król et al. (2014)	[36]	No	Continuous	LMM	Normal	MVN
Li et al. (2016)	[45]	No	Continuous	Marginal proportional means model	N/A	Multivariate – left unspecified
Shen et al. (2016)	[48]	No	Continuous	LMM	Normal	MVN + stationary Gaussian process
Succession of events						
Dantan et al. (2012)	[40]	No	Continuous	Segmented LMM with random change-point	Normal	MVN
Ferrer et al. (2016)	[66]	No	Continuous	LMM	Normal	MVN
Rouanet et al. (2016)	[50]	Yes*	Continuous (normal and non-normal)	LMM*	Normal	MVN

Abbreviations: LMM = linear mixed model, GLMM = generalized linear mixed model, MVN = multivariate normal, N/A = not applicable

* The primary model was developed for a univariate continuous outcome, but the extension to multivariate non-Gaussian longitudinal outcomes through a latent variable process model with parametric monotonic link function was also detailed.

Table 2. Summary of time-to-event submodels.

Article	Ref.	Multiple events	Recurrent events	Succession of events	Model	Random effects distribution[&]
Huang et al. (2001)	[41]	✓	X	X	Discrete-time hazard log-linear models	Discrete probability
Chi & Ibrahim (2006)	[42]	✓	X	X	Novel time-to-event joint model with conditional and marginal proportional hazards structure, and capable of accommodating zero- and non-zero cure rate fractions	Positive stable law[§]
Han et al. (2007)	[51]	X	✓	X	General recurrent events model of Peña and Hollander [52]	Gamma[§]
Liu et al. (2008)	[56]	X	✓	X	Proportional hazards with piecewise constant baseline hazard and intensity functions	Normal
Zhang et al. (2008)	[49]	✓	X	X	Constant baseline intensities	Normal
Liu & Huang (2009)	[57]	X	✓	X	Proportional hazards with piecewise constant baseline hazard and intensity functions	Normal
Dantan et al. (2012)	[40]	X	X	✓	Proportional transition intensity model with Weibull and piecewise constant baseline functions	N/A
Kim et al. (2012)	[58]	X	✓	X	Transformation models	Normal
Zhu et al. (2012)	[43]	✓	X	X	Proportional hazards with piecewise constant baseline hazard functions	N/A
Efendi et al. (2013)	[54]	X	✓	X	Weibull-gamma-normal model	Gamma[§]
Njagi et al. (2013)	[14]	X	✓	X	Weibull-gamma-normal model	Gamma[§]
Tang et al. (2014)	[44]	✓	X	X	Proportional hazards with piecewise constant baseline hazard functions	N/A
Musoro et al. (2015)	[25]	✓	✓	X	Proportional semiparametric intensity model	Independent normal (two random effects present for within and between event types)
Tang & Tang (2015)	[39]	✓	X	X	Proportional hazards with piecewise constant baseline hazard functions	N/A
Ferrer et al. (2016)	[66]	X	X	✓	A proportional hazards Markovian intensity model (with Weibull, piecewise constant, or B-spline baseline intensity function)	N/A
Król et al. (2016)	[36]	X	✓	X	Proportional hazards with cubic M-spline baseline hazard and intensity functions	Normal
Li et al. (2016)[$]	[45]	X	✓	X	**Terminal event**: additive hazards with unspecified baseline hazard function **Recurrent events**: marginal proportional rates model	Left unspecified

Rouanet et al. (2016)	[50]	X	X	✓[#]	Two models proposed: 1. A proportional hazards Markovian intensity model (with Weibull or M-spline baseline intensity function) 2. A semi-Markovian model where transition intensity to death from disease state depends on time with illness	N/A
Shen et al. (2016)	[48]	X	✓	X	Proportional semiparametric intensity model, which was reframed as a conditional rate function for the purpose of estimation	N/A*

Abbreviations: N/A = not applicable

* In principle, separate normal frailty terms can be included, as per Henderson et al. [47].

[#] This model was a semi-competing events model.

[&] Random effects in the time-to-event submodels *other* than those shared with the longitudinal data submodel.

[§] Denotes distributions of frailties that act *multiplicatively* on the hazard. All other distributions correspond to random effects that act *additively* on the log-hazard scale.

[$] This methodological article is representative of a vast research literature on the use of marginal joint models with informative observation times, modelled according to some intensity function. In the interests of brevity, we only include a single article here.

Table 3. Summary of association structure, estimation method, and software implementation.

Article	Ref.	Association structure*	Estimation method	Software implementation & availability
Huang et al. (2001)	[41]	Current value of true latent variable + interaction terms with external covariates	**MLE**: Newton-Raphson algorithm with automatic differentiation and iterative proportional fitting	**S-Plus**: AD09 module available online to implement automatic differentiation and Newton-Raphson algorithm[1]
Chi & Ibrahim (2006)	[42]	Current value parameterization	**Bayesian MCMC**: Gibbs sampling algorithm (with adaptive rejection algorithm and Metropolis algorithm)	N/S
Han et al. (2007)	[51]	Latent class membership	**MLE**: EM algorithm	N/S
Liu et al. (2008)	[56]	Random effects parameterization Both recurrent and terminal time-to-event models additionally correlated through common frailty, which is independent of longitudinal process	**MLE**: Gaussian quadrature tools in standard statistical packages	**SAS**: code provided online
Zhang et al. (2008)	[49]	Random effects parameterization	**MLE**: two-stage approach with one component estimated using the EM algorithm	N/S
Liu & Huang (2009)	[57]	Random effects parameterization	**MLE**: Gaussian quadrature tools in standard statistical packages	**SAS**: code provided online[2]
Dantan et al. (2012)	[40]			
Kim et al. (2012)	[58]	Correlated random effects between longitudinal and recurrent events submodels, with time-dependent covariate vector interactions	**MLE**: EM algorithm with a recursive formula proposed to reduce the number of parameters to be maximised	**R**: code provided online
Zhu et al. (2012)	[43]	Current value parameterization	**Bayesian MCMC**: Gibbs sampling algorithm (with Metropolis-Hastings algorithm)	N/S
Efendi et al. (2013)	[54]	Random effects parametrization	**MLE**: via partial marginalization [76]; i.e. where the conjugate random effects are analytically integrated out, followed by numerical integration of shared normal random effects	**SAS**: code provided in the Appendix

[1] Code reported as being available on two websites, but neither URL appears to still be available

[2] Code reported as being available on authors website, but URL no longer appears to be active.

Njagi et al. (2013)	[14]	Random effects parameterization	**MLE**: via partial marginalization [68]; i.e. where the conjugate random effects are analytically integrated out, followed by numerical integration of shared normal random effects	**SAS**: code provided in the Appendix
Tang et al. (2014)	[44]	Current value parameterization	**Bayesian MCMC**: Block Gibbs sampling algorithm (with Metropolis-Hastings algorithm)	**R** and **Matlab**: code available on request from the authors
Musoro et al. (2015)	[25]	Current value parameterization	**Bayesian MCMC**: Gibbs sampling algorithm	**OpenBUGS**: code not provided
Tang & Tang (2015)	[39]	Current value parameterization	**Bayesian MCMC**: Block Gibbs sampling algorithm (with Metropolis-Hastings algorithm)	N/S
Ferrer et al. (2016)	[66]	Current value parameterization, Time-dependent slopes parameterization, both, or any other function of the random effects	**MLE**: hybrid algorithm that begins with an EM algorithm and switches to a quasi-Newton algorithm if the convergence is not achieved	**R**: code provided online and in Appendix
Król et al. (2016)	[36]	Current value parameterization, Time-dependent slopes parameterization, both, or any other function of the random effects	**MLE**: penalized maximum likelihood estimation using the Marquardt algorithm	**R**: implemented in the `frailtypack` package (v2.8) and code provided in the Appendix
Li et al. (2016)	[45]	Correlated random effects	Estimating equations	N/S
Rouanet et al. (2016)	[50]	Latent class membership	**MLE**: Marquardt algorithm	**R**: code provided online
Shen et al. (2016)	[48]	Random effects parameterization, with separate coefficients for the time-independent and –dependent random effects	Two-stage conditional estimating equation approach	N/S

Abbreviations: MLE = maximum likelihood estimation, MCMC= Markov chain Monte Carlo, N/S = not specified

* Association structure between the longitudinal data sub-model and the event time sub-model.