CrossMark

Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues

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

Abstract

Background: Available methods for the joint modelling of longitudinal and time-to-event outcomes have typically only allowed for a single longitudinal outcome and a solitary event time. In practice, clinical studies are likely to record multiple longitudinal outcomes. Incorporating all sources of data will improve the predictive capability of any model and lead to more informative inferences for the purpose of medical decision-making.

Methods: We reviewed current methodologies of joint modelling for time-to-event data and multivariate longitudinal data including the distributional and modelling assumptions, the association structures, estimation approaches, software tools for implementation and clinical applications of the methodologies.

Results: We found that a large number of different models have recently been proposed. Most considered jointly modelling linear mixed models with proportional hazard models, with correlation between multiple longitudinal outcomes accounted for through multivariate normally distributed random effects. So-called current value and random effects parameterisations are commonly used to link the models. Despite developments, software is still lacking, which has translated into limited uptake by medical researchers.

Conclusion: Although, in an era of personalized medicine, the value of multivariate joint modelling has been established, researchers are currently limited in their ability to fit these models routinely. We make a series of recommendations for future research needs.

Keywords: Joint models, Multivariate data, Longitudinal data, Time-to-event data, Software

Abbreviations: GFR, Glomerular filtration rate; GLMM, Generalised linear mixed model; JLCM, Joint latent class model; LMM, Linear mixed model; MCMC, Markov chain Monte Carlo; MVJM, Multivariate joint model; PH, Proportional hazards

Background

In many clinical studies, subjects are followed-up repeatedly and response data collected, for example a biomarker. The time to an event is also usually of interest, which might be explicit, e.g. death, or implicit, e.g. dropout. The longitudinal data may be censored by this time-to-event outcome. Modelling the longitudinal and event-time outcomes separately, for example using linear mixed models [1] or Cox regression models [2], can therefore be inefficient, and can lead to biased effect size estimates if the two outcome processes are correlated [3].

Research into joint modelling of longitudinal and time-to-event data has received considerable attention during the past two decades [3–7]. The motivation behind this active field of research has stemmed from three broad scientific objectives:

(1) Improving inference for a repeated measurement outcome subject to an informative dropout mechanism that is not of direct interest [8].

(2) Improving inference for a time-to-event outcome, whilst taking account of an intermittently and

* Correspondence: graeme.hickey@liverpool.ac.uk
[1]Department of Biostatistics, University of Liverpool, Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GL, UK
Full list of author information is available at the end of the article

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 2 of 15

error-prone measured endogenous time-dependent variable [5].

(3) Studying the relationship between the two correlated processes [6].

Depending on the objective, previous approaches to joint modelling have included fitting time-dependent Cox models [2], and two-stage models [9, 10]. In each case, joint modelling can lead to improvements in the efficiency of statistical inferences and reduces bias [11], which can yield substantial benefits when designing trials. Joint models can also be used to improve prediction [12], or to determine whether a longitudinal process is a surrogate for a time-to-event process [10]. The literature is extensive, with comprehensive reviews given by Hogan and Laird [13], Tsiatis and Davidian [14], and Gould et al. [15].

Previous research has mostly concentrated on the joint modelling of a single longitudinal outcome and a single time-to-event outcome; herein referred to as *univariate joint modelling*. Commensurate with this methodological research has been an increase in wide-ranging clinical application [16–21] and accessibility to analysis tools using mainstream statistical software packages [22–28]. In practice, however, the data collected will often be more complex, featuring multiple longitudinal outcomes and possibly multiple, recurrent or competing event times. As an example, Rizopoulos and Ghosh [29] described data on 407 patients with chronic kidney disease who underwent a renal transplantation. Each patient had 3 separate biomarker measurements repeatedly recorded: glomerular filtration rate (GFR), blood haematocrit level, and proteinuria. Each of these can be considered as markers of renal function, with the clinical interest being the time to graft failure. Harnessing all available information in a single model is advantageous and should lead to improved model predictions. This therefore makes multivariate joint modelling an attractive tool in an era of personalized medicine, as physicians can gain a better understanding of the underlying disease dynamics and ultimately choose the most optimal treatment for a patient at a particular follow-up time point. Notwithstanding the increased flexibility and better predictive capabilities, the extension of the classical univariate joint modelling framework to a multivariate setting introduces a number of technical and computational challenges.

A number of factors, many still faced by the univariate joint modelling framework, precludes integration of this more sophisticated modelling framework into routine clinical research practice [15]. The aim of this paper is to provide an overview of recent methodological developments and applications of joint models for time-to-event data and multivariate longitudinal data (MVJMs).

Methods

A MVJM is comprised of two submodels: (1) a multivariate longitudinal data model, and (2) a time-to-event data model.

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 $i = 1, ..., N$; $k = 1, ..., K$, and $j = 1, ..., n_{ik}$. There are a plethora of modelling approaches for multivariate longitudinal data [30]. A generalized linear mixed model (GLMM) is a common approach, where measurements for different outcomes can be recorded at different times between patients and outcomes, and is given by:

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

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

$$\mu_{ik}(t_{ijk}) = X_{ik}^{(1)}(t_{ijk})\beta_k^{(1)} + Z_{ik}(t_{ijk})b_{ik}, \qquad (2)$$

where $X_{ik}^{(1)}(t_{ijk})$ and $Z_{ik}(t_{ijk})$ are row-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 $b_i = (b_{i1}^T, b_{i2}^T, ..., b_{iK}^T)^T$.

Cox's proportional hazards (PH) semiparametric model [2] has been a common choice for the time-to-event submodel when modelling continuous event times. For a single failure-time per subject, the hazard function for subject i at time t is given by

$$\lambda_i(t) = \lambda_0(t) \exp\left\{X_i^{(2)}(t)\beta^{(2)} + W_i(t)\right\}, \qquad (3)$$

where $X_i^{(2)}(t)$ is a (possibly time-varying) row-vector of external covariates; $\lambda_0(t)$ denotes the baseline hazard function; $\beta^{(2)}$ is a vector of fixed effects parameters; and $W_i(t)$ is a latent process that captures the association structure between the measurement and event processes. When the PH assumption is not met, alternative modelling frameworks can be exploited, such as accelerated failure time models, which can be written directly in terms of the survival function, $S_i(t)$, as

$$S_i(t) = S_0\left(\exp\left\{X_i^{(2)}(t)\beta^{(2)} + W_i(t)\right\}t\right),$$

where $S_0(\cdot)$ is the baseline survival function that depends on the parametric family used for modelling, and all other parameters are defined as per the PH model (3). Discrete event times can also be jointly modelled with longitudinal data, particularly for selection models, which is applicable to situations of interval-censored continuous event times and predefined measurement

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 3 of 15

schedules. Models here might include logistic or probit regression models [31]. We can generally represent the discrete distribution function as a function of the latent association term, namely $F(X_i^{(2)}(t_j)\beta^{(2)} + W_i(t_j))$.

In principle, each submodel can be fitted separately. However, this can result in biased estimates and a loss of efficiency when the processes are correlated. When b_{ik} and $W_i(t)$ are jointly modelled, it leads to the so-called shared random effects joint model. A graphical representation of this model is shown in Fig. 1. Joint models might also fall under the umbrellas of pattern-mixture models or selection models depending on the factorization of the joint distribution of event time and longitudinal data [31, 32]. Furthermore, modelling approaches might also fall under the umbrellas of joint latent class models [33], semiparametric [34], and fully parametric models [35]. We note also that the topic of missing values in longitudinal data analyses has its own literature [36]. In the MVJM literature, focus has mainly been towards shared random effects models. We review the different submodels, including distributional assumptions and correlation structures, latent association structures, estimation techniques, software implementations, and give clinical examples of application.

Results

Longitudinal data submodel

The choice of model for the longitudinal outcome data will depend on the type of data measured (continuous, ordinal, discrete). Although in the development and application of MVJMs they are often restricted to the simple case of continuous outcomes only [17, 19–21, 37–53], it is conceivable that multiple outcomes might be a mixture of different outcome types. For example, in Rizopoulos and Ghosh [29], GFR and haematocrit were both continuous measures, whereas proteinuria was recorded as a binary outcome. More recent modelling approaches developed have incorporated combinations of different outcome types [4, 18, 29, 54–65]. Other models have considered multivariate binary [66], ordinal [67, 68], and continuous [0,1]-bounded data [44, 45, 47, 53]. Thiébaut et al. [41] and Guedj et al. [52] both proposed a model for multiple continuous outcomes, but also allowed for left-censoring, which is a pertinent issue with biomarker measurements as they can fall below the minimum detection level.

Model

For continuous longitudinal outcomes, the Laird and Ware [1] linear mixed model (LMM) with independent and identically normally distributed within-subject measurement errors is ubiquitous. However, this distributional assumption can be sensitive to outlier observations and heavy-tailed data, which has prompted robust modelling considerations in the univariate framework [69]. Despite this, it has translated into only limited model innovation in the MVJM framework, with Tang and Tang [49] considering using the multivariate skew-
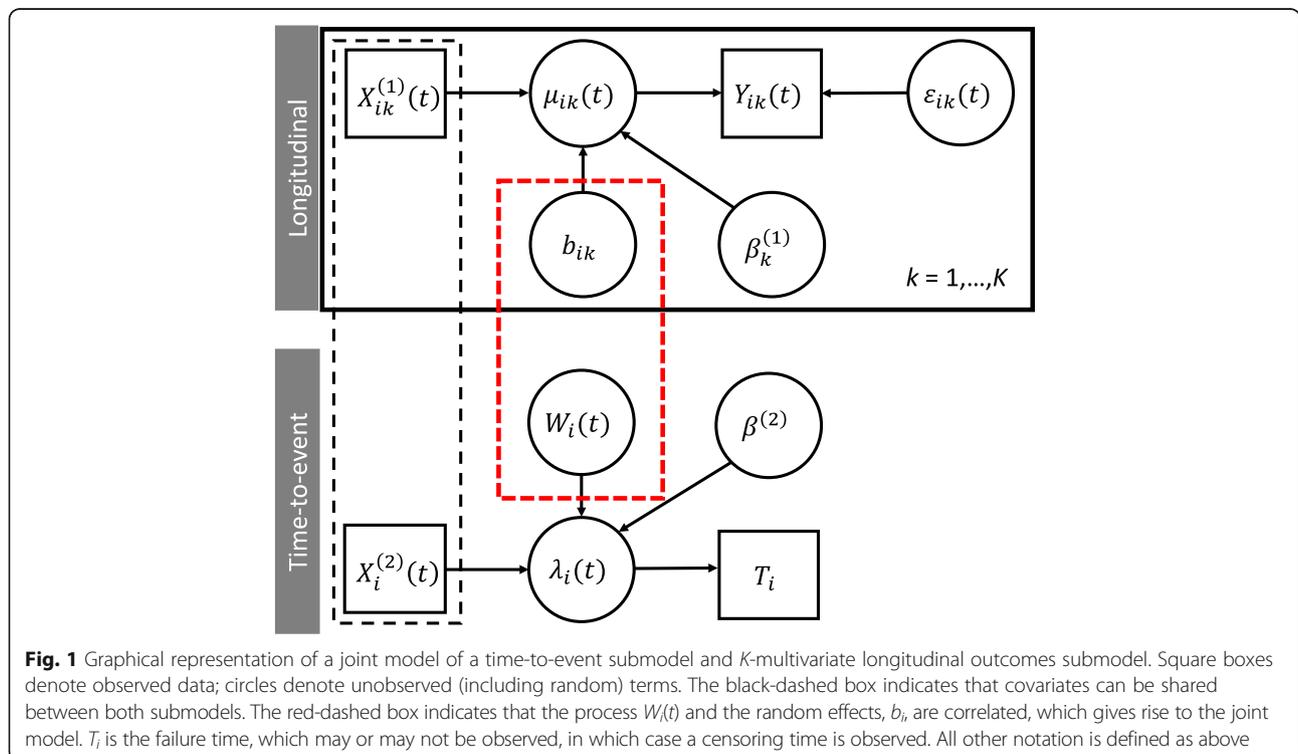

Fig. 1 Graphical representation of a joint model of a time-to-event submodel and *K*-multivariate longitudinal outcomes submodel. Square boxes denote observed data; circles denote unobserved (including random) terms. The black-dashed box indicates that covariates can be shared between both submodels. The red-dashed box indicates that the process $W_i(t)$ and the random effects, b_i, are correlated, which gives rise to the joint model. T_i is the failure time, which may or may not be observed, in which case a censoring time is observed. All other notation is defined as above

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 4 of 15

normal distribution. The robustness of model estimates to misspecification of errors, error structures, and magnitude of errors, has been examined through several simulation studies [40, 43, 49, 50]. Dantan et al. [44] and Proust-Lima et al. [45, 57] considered a model for [0, 1]-bounded continuous data using the Beta transformation link function, as it is parsimonious and offers very flexible shapes. This, however, might require a preliminary rescaling of the markers. Liu and Li [47] and Hatfield et al. [53] also considered data on the [0, 1] interval, but instead used zero-one inflated and 'zero-augmented' beta regression models, respectively. Guedj et al. [52] proposed a completely different approach by modelling multivariate continuous biomarkers using a mechanistic model based on a system of nonlinear ordinary differential equations.

For binary and count data, the standard model is the GLMM. This model has been regularly used in the MVJM framework [4, 18, 29, 54–56, 58, 60–63, 66, 70]. Moreover, this can be extended to multivariate data of different outcome types (e.g. a combination of continuous and binary measures) with correlation induced by modelling the random effects together. Huang et al. [66] considered a logistic regression model for binary outcome data, with a logistic regression model for the unobserved latent variable and a linear pairwise odds-ratio model for the association between marginal probabilities. Models for ordinal data have been considered, including the proportional odds model [18, 54, 55, 64, 67], cumulative probit model [57], partial proportional odds model [68], and the continuation ratio mixed model [56, 59].

Random effects distributions

In the univariate joint model framework, it has been reported that inferences are generally robust to misspecification of random effects [71, 72]. However, in the MVJM framework the dimension of these random effects might be greater, potentially amplifying the impact of misspecification on parameter estimates and standard errors. In general, the ubiquitous normality assumption might be too restrictive to capture individual-level variation [49]. Nonetheless, multivariate normal distributions are the standard modelling choice for random effects in the longitudinal submodel. Several simulation studies have generated data under misspecified random effects [43, 49, 60, 68]. Pantazis and Touloumi [43] explored misspecification by fitting their proposed model [40] to data simulated under a range of heavy tailed, skewed, and mixture distributions. They concluded that fixed effect parameter estimates were quite robust to misspecification with the exception of those in the time-to-event submodel, but standard errors may be underestimated for heavily skewed distributions. These findings were in agreement with those found by Li et al. [68].

Xu et al. [58] explored the sensitivity of the parameter estimates to a multivariate t-distribution for the random effects. Tang and Tang [60] investigated the effect of misspecification on their semiparametric model by simulating data with random effects drawn independently from uniform and relocated Gamma distributions. Song et al. [38] simulated random effects from a bimodal normal mixture distribution to confirm the robustness of the semiparametric estimator; whereas Rizopoulos and Ghosh [29] simulated random effects from a three-component normal mixture model to confirm the robustness of the Dirichlet process prior formulation.

To avoid misspecification, it can be advantageous to semiparametrically model the random effects. In the Bayesian paradigm, extended from the univariate framework [73], a Dirichlet process has been used to specify the random effects prior distribution [29, 49, 60]. The subject-specific random effects were assumed to be independently and identically distributed according to some unknown density function, which is modelled by the Dirichlet prior [74]. Song et al. [38] treated the random effects as nuisance parameters, for which a set of estimating equations were deduced based on a derived sufficient statistic. Li et al. [68] also proposed a method whereby the distributions of the assumed zero-mean random effects could be left completely unspecified and estimated entirely non-parametrically by exploiting the vertex-exchange method.

Discrete random effects confer an advantage in model estimation by replacing (possibly high-dimensional) numerical integration by summations that are more manageable. Bartolucci and Farcomeni [61] introduced two discrete random effects: one that follows a single first-order latent Markov chain distribution, and a second time-constant latent variable to account for unobserved heterogeneity. Huang et al. [66] also used independent discrete random effects to account for association between events and longitudinal outcomes.

Correlation structure

Multidimensional random effects are usually treated as correlated and modelled with a multivariate normal distribution with unstructured covariance matrix; however, independence and other structures have been considered. Ibrahim et al. [19] considered an unstructured correlation between the random effects, but also explored using a more parsimonious (and computationally faster) diagonal covariance matrix. Proust-Lima et al. [45, 57] included subject-class-level specific random effects in the latent variable model, and subject-outcome-level specific random intercepts in the longitudinal submodel. The subject-outcome random-intercepts were distributed normally with outcome-specific variance. The subject-class random effects, however, were distributed

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 5 of 15

multivariate normally with latent-class specific mean and an unstructured covariance matrix multiplied by a (latent) class-specific proportionality parameter. Musoro et al. [17] considered two types of random effects in the multivariate longitudinal submodel: the standard subject-specific random effects modelled by a multivariate normal distribution, and basis-outcome-specific random effects for the thin-plate spline used to model time, modelled according to independent normal distributions with basis-outcome specific variances. Li et al. [68] and Choi et al. [65] factorised the random effects as $b_{ik} = \Gamma_k b_i$, which reduces the dimension of random effects. For large K, Hatfield et al. [53] also discussed alternative simpler correlation structures for the random effects.

Although correlation between the multiple longitudinal outcomes can be modelled through the subject-specific random effects, it can alternatively be modelled through correlated error terms [19, 20, 38, 39, 42, 49, 60, 70]. Using the notation from (2) assuming a multivariate LMM with normal errors and multiple outcomes recorded according to a single measurement schedule, we have

$$Y_{ik}(t_{ij}) = X_{ik}^{(1)}(t_{ij})\beta_{k^{(1)}} + Z_{ik}(t_{ij})b_{ik} + \varepsilon_{ijk}.$$

Let also $\varepsilon_{ij\cdot} = (\varepsilon_{ij1}^T, ..., \varepsilon_{ijK}^T)^T$ denote the vector of errors for the k-th outcome. Xu and Zeger [58] considered the approach of correlated subject-specific random effects; namely

$$\varepsilon_{ijk} \sim N(0, \sigma_k^2), \text{ and } b_i \sim N_\nu(0, \Psi),$$

where σ_k^2 is an outcome-specific variance term, and Ψ is a $(\nu \times \nu)$-covariance matrix capturing the correlation between markers and repeated measures. Chi and Ibrahim [42] on the other hand considered the approach of correlated error terms; namely:

$$\varepsilon_{ij\cdot} \sim N_K(0, \Sigma), \text{ and } b_{ik} \sim N_{\nu_k}(0, \Psi_k),$$

where Σ is a covariance matrix that captures the association between longitudinal outcomes measured at the same time, and Ψ_k is a covariance matrix that captures the association between the random effects for the k-th outcome (with the vector b_{ik} of length ν_k). Chi and Ibrahim noted that their model allows for dependence between repeated measures and correlation between longitudinal outcomes to be considered separately, whereas in the Xu and Zeger model the different correlation types are conflated. Notwithstanding the inferential benefit, the latter model requires additional covariance parameters be estimated, which increases the computational challenge. Interestingly, Baghfalaki et al. [64] replaced the random effects term completely independent of outcome k, which whilst allowing for correlation between outcomes, in general will not be biologically plausible as subject-specific deviations will be on different scales for disparate longitudinal outcomes.

An alternative approach to model correlation between multiple longitudinal outcomes is to introduce a common latent variable between the models. Ibrahim et al. [19] considered the model

$$Y_{ik}(t_{ij}) = \beta_{0k} + \beta_{1k}\mu_i(t_{ij}) + \varepsilon_{ijk},$$

where $\mu_i(t_{ij})$ is the true unobserved latent variable at time t_{ij}. For example, if K different immunological measurements were recorded from a blood sample at time t_{ij}, then we might let $\mu_i(t_{ij})$ denote the 'true antibody level', which we cannot observe, but which we can infer from the multivariate outcome. The model is completed by specifying the latent variable model, e.g. as in (2), and a distribution for the error terms. For example, Ibrahim et al. considered $\varepsilon_{ij\cdot} \sim N_K(0, \Sigma)$, and Luo [55], $\varepsilon_{ijk} \sim N(0, \sigma_k^2)$. Guedj et al. [52] intrinsically accounted for correlation between multiple biomarkers through the system of dependent ODE equations that models the biological system.

Within-subject autocorrelation structures, e.g. Henderson et al. [6], are not routinely considered; although Wang et al. [67] introduced a Gaussian process model for the underlying latent variable with a power function of time modelling the correlation. Proust-Lima et al. [57] considered the inclusion of a flexible zero-mean Gaussian autocorrelated process that admits Brownian motion and autoregressive processes as special cases. Ibrahim et al. [19], Pantazis et al. [40], and Hatfield et al. [53] noted their models could be extended to autocorrelation, but at the expense of added computational burden.

In some cases, where multivariate longitudinal data has been collected, the correlation has been ignored in order to allow simpler univariate joint models to be applied. For example, Battes et al. [21] used an ad hoc approach of either summing or multiplying the three repeated continuous measures (standardized according to clinical upper reference limits of the biomarker assays), and then applying standard univariate joint models.

Time-to-event data submodel

Commonly, event times are measured continuously. If the subject does not experience the event of interest during the study observation period, then they are right-censored at the last known follow-up time. In some cases, where there is delayed entry to the study, it is necessary to account for left truncation [45, 57, 62]. In the MVJM framework, discrete event times have also been considered, [46, 50, 51, 61, 66]. Bartolucci and Farcomeni [61] noted that their estimation method could be extended to include continuous right-censored event

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 6 of 15

times at a cost of loss of information and efficiency. Dantan et al. [44] also considered discrete time, but the model applied was relevant to continuous time also. These models do not permit premature non-informative censoring before the end of the study. However, Albert and Shih [46] suggested that if subjects are censored before the end of the study, then a separate precursory model could be used to impute the event times prior to application of the joint model methodology.

In practice, clinicians might not only be interested in multivariate longitudinal outcomes, but also multivariate time-to-event data. For example, Chi and Ibrahim [42] were interested in assessing whether different quality of life measures were prognostic and predictive of breast cancer progression in a drug randomised controlled trial (RCT). In addition to the multivariate longitudinal outcomes, the study monitored patients concerning two event times: overall survival and disease-free survival. Tang et al. [60], Tang and Tang [49], and Zhu et al. [70] each proposed *multiple events* joint models, motivated by the same dataset as analysed by Chi and Ibrahim [42]. Musoro et al. [17] considered a case of *multiple recurrent events*, where each patient could become repeatedly infected with one of 9 different infections. Huang et al. [66] analysed 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. Competing risks data have recently been considered in the context MVJMs also [56, 57, 59, 68].

Model

The Cox PH model is an attractive choice as no distributional assumptions are required on the time-to-event data. In some settings, the unspecified baseline hazards function has been replaced by either a piecewise constant step-function for some preselected knots [18, 19, 29, 39, 47, 49, 54, 57, 59, 60, 70], or modelled using spline functions, including B-splines [4, 56], M- (and I-) splines [45, 57], and restricted cubic splines [62]. The position of knots is usually decided in advance, for example by taking quantiles (e.g. [59]), though advice is generally lacking. As an added degree of flexibility, Proust-Lima et al. [45, 57] consider class-specific (and cause-specific where applicable) baseline hazards that can be either stratified by class $m = 1, ..., M$, i.e. $\lambda_{0m}(t)$, or proportional by class, i.e. $\exp(\beta_m)\lambda_0(t)$, for some parameters β_m.

Parametric models considered in the MVJM framework include the Weibull [20, 45, 52, 53, 55, 57, 62, 64, 65], exponential [20, 52, 62, 65], log-normal [40, 41, 43, 55, 64], log-logistic [55], and Gompertz [57, 62]. Hu et al. [48] used a Weibull model for imputing composite event times, but used a conditional multinomial logistic

regression model to then impute the cause type. The software package by Crowther [62] also allows for the Royston-Parmar model [75] to be fitted, which is a flexible parametric model that models the log-cumulative hazard using restricted cubic splines. For models involving multivariate event time data, standard models were applied; notably, for competing risks data the cause-specific hazards model was applied [76]. Chi and Ibrahim [42] developed a novel multivariate survival model that accommodates both zero and non-zero cure fractions, and which has a PH structure conditionally and marginally under certain settings. Hu et al. [48] developed an imputation approach that first imputed a composite event time, and then imputed a cause type using a conditional multinomial regression model.

The pattern-mixture model approach by Fieuws et al. [63] used the Kaplan-Meier estimator to model the failure time, which they described as prior probabilities, which were used in an elegant Bayes rule to calculate the conditional probability of failure. Models considering discrete time data have utilised a number of discrete probability models, including the probit model [46], logistic model [50], discrete time loglinear hazard models [66], truncated geometric distribution [50, 51, 61], and discrete proportional hazards model [50].

Frailty

Random effects are also included in some time-to-event submodels, to account for correlation between different or repeated events, where they are referred to as frailty effects when multiplicative on the hazard. Chi and Ibrahim [42] proposed a power stable law distribution to account for correlation between multiple event times. Lin et al. [37] and Choi et al. [65] respectively included gamma and log-normally distributed frailty terms to allow for subject-level variability.

Association structures

Fundamental to the joint modelling framework is the association structure between the longitudinal submodel and the time-to-event submodel. Rationale for selecting this association structure has received relatively little attention. McCrink et al. [32] state that the choice of association structure should reflect the study focus, namely whether it is with respect to the time-to-event or longitudinal submodel (or both). For a discussion of different association structures for univariate joint models, see Rizopoulos [4]. We consider below different representations for $W_i(t)$ applied in the MVJM framework (Table 1). MVJMs that fall outside of the ubiquitous shared random effects framework cannot be reduced to simply specifying $W_i(t)$, and so we describe

Table 1 Some association structures for joint models of time-to-event and multivariate longitudinal data

Parameterization	Latent association	Studies[i]
A1: Current value (linear predictor)	$W_i(t) = \sum_{k=1}^{K} a_k \mu_{ik}(t)$	[17, 19, 21, 29, 37, 39, 42, 48, 49, 52, 56, 58, 60, 62, 67, 70]
A2: Current value (expected value)[a]	$W_i(t) = \sum_{k=1}^{K} a_k h_k^{-1}(\mu_{ik}(t))$	[62]
A3: Interaction[b]	$W_i(t) = \sum_{k=1}^{K} \left\{ a_k \mu_{ik}(t) + \sum_{l=1}^{p} x_{il}^{(2)} \mu_{ik}(t) \gamma_{kl} \right\}$	[37]
A4: Lagged time[c]	$W_i(t) = \sum_{k=1}^{K} a_k \mu_{ik}(t - c)$	[46]
A5: General vector function[d]	$W_i(t) = a^T \psi(t, b_i)$	[38]
A6: Time-dependent slopes[e]	$W_i(t) = \sum_{k=1}^{K} \left\{ \mu_{ik}(t) + \sum_{v=1}^{V} a_{vk} \frac{d^v}{dt^v} \mu_{ik}(t) \right\}$	[29, 56]
A7: Cumulative effects	$W_i(t) = \sum_{k=1}^{K} a_k \int_0^t \mu_{ik}(s) ds$	[56]
A8: Random effects[f]	$W_i(t) = \sum_{k=1}^{K} a_k^T b_{ik}$	[18, 29, 53–55, 59, 62, 64, 65, 68]
A9: Generalised random effects + fixed effects[g]	$W_i(t) = \sum_{k=1}^{K} a_k^T r\left(b_{ik} + \tilde{\beta}_k^{(1)} \right)$	[29, 50, 51, 59, 62]
A10: Correlated random effects[h]	$W_i(t) = \theta_i \text{ with } (b_i^T, , , \theta_i)^T \sim F_a$	[29]

Notation: $\mu_{ik}(t)$ denotes the linear predictor term of the longitudinal submodel for subject i and outcome k; a_k denotes the association parameter for the k-th outcome
[a] $h_k(\cdot)$ is the link function for the k-th outcome
[b] $x_{il}^{(2)}$ denotes the l-th baseline covariates for subject i ($l = 1, …, p$) with corresponding coefficient parameters γ_{kl} for each outcome k. In practice, some γ_{kl} coefficients will be set to zero
[c] c is a lag time (with $c = 0$ returning the current value parameterization). In Albert and Shih [46], time was modelled discretely and a selection model adopted, such that $W_i(t_j) = \sum_{k=1}^{K} a_k \mu_{ik}(t_{j-1})$
[d] a is a vector of association parameters and $\psi(t, b_i)$ is a vector of time and random effects. It is assumed that $\psi(t, b_i)$ can be decomposed as $\psi(t, b_i) = \psi(t)b_i$. This general parameterization admits the current value parameterization as a special case, and leads to a number of extensions including interactions with time. In cases where $\psi(t, b_i)$ does not factorise, the authors propose using an approximation method
[e] a_{vk} denote additional association parameters for the v-th derivative (with respect to time) for the k-th longitudinal outcome mean trajectory function
[f] a_k denotes a vector of association parameters of same dimension as the number of random effects for each outcome. In practice, some elements of a_k might be forced to zero, e.g. if only random intercepts were used to link the model
[g] as per the random effects parameterization a_k denotes a vector of association parameters of same dimension as the number of random effects for each outcome. $\tilde{\beta}_k^{(1)}$ denotes the subset of coefficient parameters from $\beta_k^{(1)}$ that correspond to the random effect terms, and $r(\cdot)$ denotes a vector function. If $r(\cdot)$ is the identify function, then the standard random + fixed effects parameterization is returned
[h] F_a denotes a multivariate density function with parameters a to model correlation
[i] Rizopoulos [4] describes a general MVJM and notes that, in principle, the general association structures that are used in the R JM package [27] are applicable to the multivariate case. However, the model was only described without fitting or application, therefore we have not included these association structures here

these models separately. In some cases, one might want to use different association structures for different longitudinal outcomes. This is a greatly overlooked modelling issue, which to the best of our knowledge, only Andrinopoulou et al. [56] and Crowther [62] have considered.

Time-dependent associations

The standard joint model assumes that risk of an event at time t depends on the true value of the longitudinal profile for the same time point (Table 1, A1)—the so-called current value parameterization. The strength of the association is fully interpretable: $\exp(\alpha_k)$ is the hazard ratio for a unit increase in $\mu_{ik}(t)$, at time t. An alternative current value parameterization is to replace the linear predictor term by the expected value of the longitudinal trajectory function at time t, $h_k^{-1}(\mu_{ik}(t))$ (Table 1, A2), which is of importance for correctly modelling the functional form. The current value parameterization can be extended to incorporate additional structure, including interaction terms with the covariates (Table 1, A3), which might yield more realistic inferences as it is conceivable that different associations exist for different patient subgroups. In some

cases one might posit that the risk depends not on the current value, but on a previous value, giving rise to the time-lagged values parameterization (Table 1, A4). Current value parameterizations have been used in model frameworks not compatible with (3). Chi and Ibrahim [42] adopted a novel bivariate time-to-event model whereby covariates—including a current values parameterization—are entered by a method corresponding to a canonical link in a Poisson generalized linear model. Song et al. [38] developed a model such that $W_i(t) = \alpha^T \psi(t, b_i)$, for some vector function ψ (Table 1, A5). The estimation methodology assumed that $\psi(t, b_i)$ could be factorised into $\psi(t, b_i) = \psi(t)b_i$, which admits the current value parameterization as a special case, and leads to a number of extensions including interactions with time. In cases where $\psi(t, b_i)$ does not factorise, meaning that it is nonlinear in t, the authors propose using a linear approximation method.

Derivative terms allow one to incorporate not only the current value of the true longitudinal process, but also the rate of change, which intuitively might be associated with risk of the event. For example, two patients might have the same observed longitudinal outcome at time t,

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 8 of 15

but one patient's trajectory might be rising considerably more quickly than the other patient's. $W_i(t)$ can therefore be augmented to include the current value plus the first V derivate terms (Table 1, A6), although this model is typically only used with the first order derivatives ($V = 1$), giving rise to the time-dependent slopes parameterization. The antithesis of the time-dependent slopes parameterization is the cumulative effects parameterization (Table 1, A7). Here, a summary of the entire history of the longitudinal process up to time t is included in the hazard model, $\lambda_i(t)$. This is contrary to other association structures that relate the hazard function only to features of the longitudinal model at a fixed time point.

Random effects parameterization

The above parameterizations often require numerical integration, which presents a computational challenge. Simpler time-independent associations can overcome this. The random effects parameterization (Table 1, A8), which only includes the time-independent random effects, is therefore frequently used in joint models. This parameterization has been used by a number of authors in various ways. In the simple case of a random-intercepts and random-slopes model for (2), the random effects represent the subject-specific deviation from the average intercept and slope fixed effect terms. Nevertheless, experts have echoed caution when attempting to use these models for inference, as complex longitudinal trajectory functions, such as those modelled by polynomials or splines, lead to non-interpretable association parameters [4, 15]. On the other hand, Jaffa et al. [50, 51] were specifically interested in the multivariate LMM slopes, so there was a clear a priori rationale for this association structure. Moreover, they noted that their model does not preclude inclusion of the random-intercepts, but demonstrated that specifying the marginal time-to-event (dropout) model is only required. Random effects parameterizations are sometimes used to refer to models where the hazard is associated with random plus fixed effect terms (Table 1, A9). For example, rather than model risk as dependent on b_{ik}, one assumes it is dependent on $b_{ik} + \tilde{\beta}_k^{(1)}$, where $\tilde{\beta}_k^{(1)}$ is a subset of $\beta_k^{(1)}$ that correspond to the random effects terms in b_{ik}. This model can be generalised to include functions of random coefficients.

Correlated random effects and frailty

An alternative approach to specifying an association structure is not to directly include random effects components of the longitudinal submodel in the time-to-event submodel, but rather to include separate random effects in each, and specify a joint distribution for the latent terms (Table 1, A10). In the simplest case, one can

set $W_i(t) = \theta_i$, and then jointly model $(b_i^T, \theta_i)^T$. Rizopoulos and Ghosh [29] considered such a structure, assuming that the joint distribution was unknown, and used a Dirichlet prior to fit the model.

Correlated random effects and error

Pantazis et al. [40], Thiébaut et al. [41], and Pantazis and Touloumi [43] assumed a log-normal distributional for the event times, i.e. $\log(T_i) = X_i^{(2)}\beta^{(2)} + e_i$, where the error terms e_i are assumed to follow a normal distribution with mean 0 and variance σ_T^2. The multivariate longitudinal data submodel and log-normal event time submodel were associated by assuming that $(b_i^T, e_i)^T$ are jointly distributed. When the random effects are multivariate normally distributed, this distribution is also multivariate normal. The covariance terms $\text{cov}(b_{ik}, e_i)$ can then subsequently be used to quantify and test the strength of association.

Joint latent class models

An alternative approach to joint modelling is the use of joint latent class models (JLCMs). The assumption underlying JLCMs is that the population of subjects is heterogeneous, but consists of a number of homogeneous latent subgroups for which the subjects share the same mean longitudinal trajectories and hazard risk. A review of JLCMs is given by Proust-Lima et al. [33]. A class-specific *latent process mixed model*, which we refer to as the *latent variable model* to remain consistent with other similarly structured models, conditional on subject i being in class m, is then specified following a standard LMM: $\mu_i(t_{ijk})\big|_{c_i=m} = X_i^{(1)}(t_{ijk})\beta^{(1)} + Z(t_{ijk})b_{im}$. The fixed effects coefficients can also be made class-specific [57]. The observed multivariate longitudinal data are modelled using GLMMs or other suitable *measurement models* conditional on this latent variable: $h_k(Y_{ijk}) = \mu_i(t_{ijk}) + \alpha_{ik} + \varepsilon_{ijk}$, for some suitable link functions $h_k(\cdot)$. If $\beta^{(1)}$ is forced to be class-specific, then one might introduce additional covariates with a global fixed effects coefficient vector into the measurement model [57]. There are two sets of random effects in this model: the subject-class effects in the latent process model, and the subject-outcome effects in the longitudinal data submodel. The time-to-event submodel might be modelled as a proportional hazards model, $\lambda_i(t \mid c_i = m) = \lambda_{0m}(t)\exp(X_i^{(2)}\beta_m^{(2)})$. The class membership probabilities are modelled using multinomial regression models. As JLCMs do not require precise modelling of the association structure between the different submodels, with the association captured entirely by the latent classes, this means that they are not necessarily well suited for evaluating assumptions regarding the association between the two submodels; however, they are particularly useful for predictive modelling.

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 9 of 15

Other

Other MVJM approaches that fall outside of the ubiquitous shared random effects model framework or the emerging joint latent class model framework, lead to alternative association structures. Fieuws et al. [63] used a pattern-mixture model approach. Here, the dependency derived from the longitudinal submodels being fitted conditional on the failure times. Hu et al. [48] proposed a model that incorporates some function of the history of the longitudinal data, which reduces to the current value parameterization as a special case. Bartolucci and Farcomeni [61] proposed a discrete time event-history model with a mixed latent Markov model. A flexible association structure was obtained though the introduction of two discrete latent variables: a time-varying latent variable distributed according to a finite first-order latent Markov chain, and a time-constant latent variable.

Estimation techniques

Historically, complete likelihood analysis was precluded by the inherently complex likelihood functions, necessitating so-called two-stage models [9, 10]. However, these models have been shown to lead to biased results [77]. A number of estimation approaches have been considered for MVJMs, building on the methodological developments in the univariate joint model literature [5, 78].

Frequentist model estimation

The expectation-maximisation (EM) algorithm [79] was the original estimation approach for joint-likelihood univariate joint models [5, 6], and therefore continues to be employed in a number of MVJM approaches [37, 40, 43, 61, 68]. At the M-step, maximization was routinely implemented using both closed-form estimators and the Newton-Raphson algorithm. Some noteworthy extensions included the one-step-late algorithm [37] and restricted iterative generalized least squares [40, 43]. The E-step was typically implemented using numerical integration, including Gaussian quadrature (e.g. [adaptive] Gauss-Hermite), although Monte Carlo integration [37], extensions of the forward-and-backwards recursion method [61], and exploitations of multivariate normal and truncated normal distributions [40, 43], were also implemented. Full maximum likelihood estimation can also be implemented directly by Newtonian-like approaches. These include the Marquardt algorithm [41, 44, 45, 57], Newton-Raphson algorithm [62], and robust variance-scoring algorithm [52]. Huang et al. [66] 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.

For estimation methods based on likelihood maximization, evaluating an approximated inverse observation matrix at the maximum likelihood estimate is a standard approach [37, 41, 45, 52, 61, 62, 66] of calculating standard errors. Semiparametric time-to-event models have been noted to result in underestimation of parameter standard errors [80] in the univariate joint model framework, and can be unfeasible as the information matrix increases with sample size. One approach to overcome this is the bootstrap method, which has been adopted in MVJM approaches [46, 68]. Pantazis et al. [40] and Pantazis and Touloumi [43] estimated standard errors by refitting the model with multiply imputed data for censored survival times, which is quicker than conventional EM algorithm approaches.

Bayesian model estimation

The Bayesian approach has a number of advantages and has been previously exploited in the univariate joint modelling framework [78]. One such advantage is the ease of incorporating hierarchical data, as used by Luo and Wang [18] as part of a multicentre RCT joint model. Another advantage is the availability of Markov chain Monte Carlo (MCMC) sampling algorithms, which allow estimation from posterior probability density functions that are not analytically tractable, and which require complex multidimensional integration over the random effects. The Gibbs sampling algorithm has been the standard choice [17–19, 29, 39, 42, 47, 49, 53–56, 59, 60, 64, 65, 67, 70], applied in conjunction with the adaptive rejection algorithm, slice sampling algorithm, block sampling, and Metropolis-Hastings algorithm. The surge in Gibbs sampling can be partly explained by the use of the BUGS computer language [81], in particular WinBUGS and OpenBUGS, which reduces the need for complex analytical derivation. Deriving the likelihood function, however, is still challenging, and depending on the time-to-event model and latent association, numerical integration might be required. For example, Ibrahim et al. [19] used an approximation similar to that used by Tsiatis et al. [10] that is convenient so long as the response does not change over time too rapidly compared to the scheduled longitudinal data measurements. Brown et al. [39] on the other hand used a simple trapezoidal rule. Gaussian quadrature, such as the Gauss-Kronrod rule, has also been used [29, 56, 59, 60]. Liu and Li [47] compared the performance of Bayesian approaches to classical frequentist (maximum likelihood) approaches under different strengths of association, demonstrating superiority of the Bayesian methods with respect to bias, root-mean square error, and coverage. Another advantage to Bayesian modelling is the ability to incorporate prior knowledge. Tang and Tang [49] explored the sensitivity of results to prior distribution selection, showing that good prior knowledge led to marginally improved estimation. Uncertainty about posterior parameter estimates is readily calculated from the MCMC output without need for further complex calculations.

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 10 of 15

Other estimation approaches

Song et al. [38] extended the semiparametric conditional score estimator for the parameters in the hazard relationship, as proposed by Tsiatis and Davidian [14] in the univariate framework, which treats the random effects as nuisance parameters, and a set of estimating equations are deduced based on a derived sufficient statistic. Parameter standard errors were subsequently estimated using a sandwich matrix estimator. Li et al. [68] employed a non-trivial extension of a method proposed by Tsonka et al. [82] in order to estimate the model parameters and zero-mean random effects distribution using a modified vertex-exchange method algorithm in conjunction with an expectation-maximization algorithm. Hu et al. [48] circumvented the classical joint modelling framework by proposing a multiple imputation algorithm using either a fully conditional specification or MCMC approach, such that simple and transparent statistical approaches can be separately applied to the complete data. Rubin's rule [83] could then be used to account for the additional uncertainty in standard errors from imputation.

Albert and Shih [46] proposed a novel two-stage regression calibration approach. In the first stage, conditional on each subject's event time, complete longitudinal data were simulated for each subject using normal approximations. Multivariate longitudinal models were then estimated using the approach of Fieuws and Verbeke [84], which fits all bivariate models and averages over the duplicate parameter estimates. This method is advantageous as it enables one to consider high-dimensional data, which would otherwise present numerical challenges or be computationally infeasible. Following this, an estimator was proposed for the subject-specific random effects, allowing for estimates of the mean longitudinal trajectories at each discrete time point for each subject to be obtained. In the second stage, a regression-calibration approach was then used to estimate the discrete time-to-event model parameters. The resulting parameter estimates were averaged over repeated simulations of the model-fitting algorithm. Standard errors were estimated using the bootstrap method.

Fieuws et al. [63] adopted a pattern-mixture model estimation approach, whereby multivariate longitudinal models are estimated—separately for those who experience and those who do not experience the event—again using the proposed approach of Fieuws and Verbeke [84] described above. Bayes' rule was then used to estimate the failure time distribution conditional on the longitudinal profiles, with a non-parametric survival distribution acting as the prior distribution.

Software

The adoption of joint models has been slow [15]. Among the many reasons for this includes the historically limited availability of software specific to joint models. Recently packages for mainstream statistical software, including R (R Foundation for Statistical Computing, Vienna, Austria) [24, 27], SAS (SAS Institute, Cary, NC) [26], Stata (Stata-Corp LP, College Station, TX) [22], and WinBUGS (MRC Biostatistics Unit, Cambridge, UK) [23], have allowed researchers to exploit joint modelling. However, these have been limited to univariate data. Many articles describing developments or applications of joint models involving multivariate longitudinal data have reported some details about the software used to fit the joint models.

R

Brown et al. [39] compiled their flexible B-spline model for multiple longitudinal biomarkers and time-to-event outcome (with current value association parameterization) into an R package, sjmsoft, available from the author's website: http://faculty.washington.edu/elizab/software.html [Accessed: 25 January 2016]. Battes et al. [21] used the R package JM [27], which fits univariate joint models, by reducing the multivariate longitudinal data to a univariate outcome through ad hoc techniques. R has also been used as an interface to execute JAGS [85] and WinBUGS/OpenBUGS programs [47, 53, 56, 65]. Andrinopoulou et al. [59] implemented their model using two separate software packages, one of which was R, with the code available from the authors upon request. Tang et al. [60] reported their Bayesian models were fitted using R and Matlab (The MathWorks Inc., Natick, MA), with code available from the authors upon request. Albert and Shih [46] fitted the bivariate longitudinal models using code presented elsewhere [86]. Bartolucci and Farcomeni [61] published R code as a supplemental file for their event-history extension Markov model, which depends on compiled Fortran routines.

BUGS & JAGS

Bayesian MVJMs have been coded in the BUGS language and fitted using WinBUGS and OpenBUGS [18, 29, 53–55, 59, 64, 65]—software that implements MCMC sampling. JAGS has also been used as alternative [47, 56], which shares a similar syntax to BUGS models. The flexibility of the software has permitted countless association structures, submodels, and data types. Hatfield et al. [53] also noted that they extended the WinBUGS platform to include a 'zero-augmented beta' distribution using code written in Pascal. In the majority of publications, the model script has been made available in an appendix [18, 29, 55, 59, 64] or is available on request from the authors [56, 65]. However, it was generally application specific; thus requiring adaptation for application with other datasets.

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 11 of 15

Stata

It has recently been announced that the next version release of the Stata package stjm [22, 87] will allow for multivariate longitudinal outcomes. The current version implements maximum likelihood estimation of univariate joint models with a number of different parametric time-to-event models. In addition, stjm can jointly model different outcome types, different association structures, and different random effects covariance matrix structures, alongside extensive optimization control settings, thus giving the user immense flexibility. A number of post-estimation options are available, including residual calculation and prediction. Pantazis et al. [40] also report that a Stata program is available (on request from the authors) for fitting their bivariate MVJM.

Fortran

Fortran has been used in several MVJM studies [19, 44, 57, 61]. Of particular interest is the HETMIXSURV (version 2) program available from: http://www.isped.u-bordeaux.fr/BIOSTAT [Accessed: 11 April 2006]. This Fortran 90 parallel program implements maximum likelihood estimation of the multivariate JLCM proposed by Proust-Lima et al. [57], in addition to other related models, which permits different outcome types and submodels. The R package lcmm [28] has similar capabilities, but does not currently permit multivariate longitudinal data in the JLCM framework. Dantan et al. [44] also fitted a JLCM using Fortran 90, and note the code is available on request. Ibrahim et al. [19] also report code is available for fitting their MVJM and multivariate L-statistic from the authors' website.

Other

SAS has been implemented in the MVJM literature, with direct reference to the procedure NLMIXED [20, 50, 51, 63], but without making the code available. Song et al. implemented their conditional score method using C++ [38], and Li et al. [68] fitted their bivariate ordinal model with competing risks data using C. In both cases, the code is available from the respective authors upon request. Matlab was used by Lin et al. [37] to run the one-step-ahead EM algorithm (code not provided), and by Tang et al. [60] to implement an MCMC algorithm for a Bayesian MVJM (code available on request from the authors). Huang et al. [66] developed an S-Plus (Insightful Corporation, Seattle, WA) library model, AD09, to implement the automatic differential method that enables direct maximization of the MVJM likelihood function. In addition to a Stata program, Pantazis et al. also reported that MLn (Centre for Multilevel Modelling, University of Bristol, UK) macros are available from the authors for fitting their bivariate MVJM.

Clinical examples

Most methodological developments in the MVJM framework have been motivated by real-world clinical studies. Applications have been in the disease areas of HIV/AIDS [16, 38–41, 43, 52], lung disease [65, 68], cancer [19, 37, 42, 49, 53, 60, 70], cardiovascular disease [21, 56, 59, 67], neurodegenerative disease [18, 45, 54, 55, 57], renal disease [17, 29, 48, 50, 51, 63], hepatic disease [46, 61], mental health [58, 66], and cognitive studies [20, 44]. We illustrate three examples below.

Parkinson's disease drug trial

Parkinson's disease is a chronic progressive neurodegenerative disorder. Luo and Wang [18] reported data from a clinical trial of 800 patients randomized in a 2x2 factorial design to receive double-placebo, tocopherol, deprenyl, or tocopherol with deprenyl. The latter two groups formed the treatment group, whilst the former two defined the placebo group. To investigate the effect of tocopherol in slowing the progression of Parkinson's disease, 3 longitudinal outcomes were recorded that describe progression: 1 continuous and 2 ordinal measuring different facets of the disease. A substantial number of patients (376/800) failed to complete the measurement schedule (10 measurements over a 24-month follow-up period), due to deterioration and were treated with levodopa therapy. The time to initiation of levodopa therapy was the study endpoint. It was shown that patients with shorter follow-up had worse progression measurements on average; therefore, a multivariate joint model was required to account for the informative dropout. For the multivariate longitudinal data, a multilevel item-response theory model was used, with centre effects included to account for the cluster-trial design. A piecewise constant proportional hazards regression model was used for the time-to-event data, with a shared (centre-specific and subject-specific) random effects association structure (Table 1, A8). Based on multiple model comparison methods, a random effects parameterization without centre-level effects was optimal. This model indicated a non-significant effect of tocopherol on disease progression rate and on the hazard of levodopa therapy initiation when compared to placebo. However, the association parameters were strongly significant, indicating that patients with worse disease severity and faster disease progression had an increased hazard of levodopa therapy initiation. This data was also previously analysed by Luo [55] and He and Luo [54] without consideration of the centre effects, and in the former case under different parametric time-to-event models.

Heart valve replacement observational study

Aortic valve replacement is a common cardiac surgery procedure, often required to treat aortic stenosis. Andrinopoulou et al. [59] analysed data from a cohort on

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 12 of 15

283 patients who survived aortic valve or root replacement with an allograft valve. During the routine clinical follow-up appointments, two echocardiographic measurements were recorded: valve gradient (continuous) and regurgitation (ordinal). Physiologically, increased gradient or regurgitation indicates deterioration of valve performance, which could lead to either death or necessitate reoperation—both events of interest. To investigate the effects of these parameters on the hazard of adverse events, a MVJM was proposed for the bivariate longitudinal data and a competing risks model was assumed for the event times. A LMM with B-spline functional form was used to model the non-linear gradient trajectory, and a continuation ratio mixed effects model for the regurgitation data. A simple LMM for aortic gradient was considered, but resulted in an inferior model fit. A cause-specific event-time model with piecewise constant baseline hazards was used to model the competing risks data, with a random effects (+ fixed effects) association structure (Table 1, A9). The association between the longitudinal measurements and events are presented in the form of graphical plots to facilitate interpretation, as the B-spline parameters are not straightforwardly interpretable. Andrinopoulou et al. [56] reanalysed this data with the purpose of calculating dynamic predictions. Furthermore, they extended the model to consider several different association structures, with subsequent predictions combined using Bayesian model averaging in order to account for model structure uncertainty.

Cancer drug trial

Lin et al. [37] reanalysed data from a placebo-controlled randomized trial to establish whether supplemental beta-carotene reduces recurrence of the primary tumours in patients cured from a recent early-stage head and neck cancer. The primary endpoint was all-cause mortality, with 63 subjects dying during follow-up. During the trial, blood samples were taken from 264 patients at baseline, 3-months, 1-year, and annually thereafter until 5-years total follow-up was attained. Two continuous biomarkers—both plasma nutrient concentrations—were of interest: lycopene and lutein + zeaxanthin. Previous univariate studies had demonstrated that increased biomarker concentrations were associated with reduced mortality. Interest here was specifically on whether both biomarkers were simultaneously associated with mortality. To investigate this, a MVJM was developed with LMMs used to model the biomarkers, and a Cox PH model to model the time to death, with association modelled using an interaction parameterization (Table 1, A3). Additionally, a gamma-distributed subject-level frailty term was included to capture over-dispersion. It was found that the effect of lutein + zeaxanthin was diminished when both biomarkers were included, and moreover the sign was opposite to that

of the univariate joint model. By exploring the correlation between the biomarkers, the authors suggested that the effect of lutein + zeaxanthin on all-cause mortality appeared to be mediated only through the association with lycopene. Lin et al. also demonstrated the necessity of joint modelling by fitting (1) a separate event time model with baseline biomarker concentrations; (2) a time-varying covariate Cox model (equivalent to a last observation carried-forward model); and (3) a two-stage joint model, and juxtaposing each to the fitted MVJM.

Discussion

The case for use of joint models has been made in the univariate data framework [3]. Similarly in the MVJM literature, several studies have demonstrated the potential bias from ignoring the correlation between the outcomes by comparing joint models to separately fitted submodels in simulation analyses [18, 40, 42, 47, 54, 55]. With an increased focus on personalised medicine, the need to implement models that account for multiple longitudinal outcomes is necessary. Despite this, joint models have predominantly focused on univariate data. A consequence of this has been researchers fitting multiple separate univariate joint models [88], which is inefficient and can affect inferences [37]. Here we have presented an overview of MVJM literature. The models developed in the literature showcase broad classes of longitudinal and time-to-event data and countless association structures, with bespoke models often developed to meet the demands of the clinical data at hand. A small number of models have even considered simultaneously multivariate longitudinal data and multivariate time-to-event joint models [17, 42, 48, 49, 56, 59, 60, 66, 68, 70].

One advantage of joint models is that they might improve power and efficiency over separate fitted models [3]. Nonetheless, few applications of joint models to clinical trial data have derived from a planned joint modelling design. For joint models to become embedded in clinical research, researchers will need guidance on study design with reference to the different models and estimation proposals available. However, issues such as statistical power have attracted little attention in the univariate joint model literature [11, 89]. Furthermore, developments in the field of joint modelling—particularly MVJMs—have primarily focused on modelling and estimation. Practitioners will naturally require diagnostic tools to evaluate model fits and compare models. Articles considering MVJMs have utilised many different model comparison methods, goodness-of-fit tests, and diagnostics. In some cases, these statistical methods have been developed specifically for the multivariate data joint modelling framework. We illustrate some of the approaches adopted in the literature in Additional file 1: Table S1.

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 13 of 15

The extension from univariate data modelling to multivariate data modelling is, in principle, straightforward [4]. The main barrier to applied researchers is the lack of readily available software that can implement multivariate joint model estimation approaches; this is despite an emergence of software relevant to univariate joint models. The main limitation to statisticians is the inherent computational challenge that arises from the increasing dimensionality of the random effects component. Despite many multivariate models being presented in full generality, computational power, limits in numerical estimation, (effective) sample size, and clinical data availability will, unavoidably in practice, preclude analyses involving large dimensions. Of the MVJM methodologies encountered to-date, bivariate data is the most commonly encountered during application [19, 20, 37–41, 43, 48, 50, 53, 56, 57, 59, 61, 64, 68], followed by trivariate data [18, 21, 29, 44–47, 51, 58, 66]. A few articles have considered 4 longitudinal outcomes [17, 42, 49, 54, 55, 60, 63, 70]. Jaffa et al. [51] explored converge properties of their MVJM for up to 10 longitudinal outcomes, and reported good performance. It was also noted that the model should straightforwardly extended to higher dimensions. Of all the studies involving multiple event types or competing risks, all were limited to the bivariate case/two competing risk events, with the exception of one study [17], which considered 9 separate recurrent events along with 4 longitudinal biomarkers. However, in the latter case it was reported that a single joint model could not actually be fitted due to computational limitations, with the authors opting instead to fit multiple pairwise models.

The flexibility of a model to be applied to a given dataset will depend on the ability to include complex functional forms and covariate adjustment options within the submodels. Complex functional forms, whilst perhaps better capturing non-linear trajectories in the longitudinal submodel that may be observed in biological data, increase computational requirements to fit as the dimensions of the random effects increases. Notwithstanding these issues, some have developed models which includes complex smoothing functions within the multivariate framework, including B-splines [39, 59], natural cubic splines [29, 56, 62], fractional polynomials [62], P-splines [49], thin-plate splines [17], I-splines [57], and parametric non-linear functions [43]. Albert and Shih [46] also noted that their approach could be extended to nonlinear models with an additional approximation. The mechanistic model proposed by Guedj et al. [52] is able to model complex dynamics of biomarkers by virtue of the complex system of differential equations specified, which captures knowledge of actual biological processes.

Conclusions and Recommendations

Gould et al. [15] concluded that it was going to be a challenge to encourage stakeholders to adopt joint models. This challenge is further exacerbated when multivariate longitudinal data are incorporated. However, as demonstrated in this review there is a solid methodological foundation to implement these methods. We have also observed the early phase of infiltration into the non-statistical biomedical literature [16, 20, 21], which highlights the potential of the MVJM framework. For MVJMs to be integrated into the applied biostatistician's tool belt, further developments are required, including:

- Development of statistical software capable of fitting of MVJMs. Univariate joint models have seen a surge in software development, which includes wider integration with multivariate event-time data; however, lack of software for multivariate longitudinal data will preclude their use by the vast majority of applied statisticians.
- Reporting of the maximum dimension limits for multiple longitudinal outcomes. Whilst many models are presented in full generality, computational limitations will preclude large numbers of random effects, and therefore large numbers of biomarkers. Currently, little is understood about this with the exception of short commentaries by Jaffa et al. [51] and Musoro et al. [17].
- Guidance on the underlying model types, distributional assumptions, and choice of association structure. Despite methodological developments, a coherent and flexible modelling framework that encapsulates the general multivariate joint model is lacking, which precludes penetration into the biomedical arena due to a lack of understanding.
- Development of model diagnostics and selection devices compatible with the MVJM framework. A number of useful ancillary statistical tools have been recently presented, however clarity and demonstration of their suitability, as well as integration into the aforementioned software developments, are required.
- Study design guidance, including but not limited to: sample size and power calculations; selection of appropriate latent association structures; applicability of models according to different intra- and inter-patient measurement protocols, and outcome types; surrogacy; and dynamic prediction.
- Further research into MJVMs with multivariate time-to-event submodels, including competing risks, recurrent events, multiples events, and multistate data.

Additional file

Additional file 1: Table S1. Goodness-of-fit tests, model diagnostics, comparison instruments, and other model assessment tools. (DOCX 162 kb)

Availability of data and materials
Not applicable. No datasets were used in this study.

Authors' contributions
RKD conceived the intellectual idea. GLH, PP, AJ and RKD designed the study. GLH wrote the manuscript. All authors reviewed drafts of the manuscript, offered interpretation and critical comment, and approved the final version.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
[1]Department of Biostatistics, University of Liverpool, Waterhouse Building, 1-5 Brownlow Street, Liverpool L69 3GL, UK. [2]Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Ellison Place, Newcastle upon Tyne NE1 8ST, UK.

References
1. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38:963–74.
2. Cox DR. Regression models and life-tables. J R Stat Soc Ser B Stat Methodol. 1972;34:187–220.
3. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. J Clin Oncol. 2010;28:2796–801.
4. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Boca Raton: Chapman & Hall/CRC; 2012.
5. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997;53:330–9.
6. Henderson R, Diggle PJ, Dobson A. Joint modelling of longitudinal measurements and event time data. Biostatistics. 2000;1:465–80.
7. Asar O, Ritchie J, Kalra P, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. Int. J. Epidemiol. 2015; 1–11.
8. Dupuy J-F, Mesbah M. Joint modeling of event time and nonignorable missing longitudinal data. Lifetime Data Anal. 2002;8:99–115.
9. Self S, Pawitan Y. Modeling a marker of disease progression and onset of disease. In: Jewell NP, Dietz K, Farewell VT, editors. AIDS Epidemiol. Methodol Issues. Boston: Birkhauser; 1992.
10. 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.
11. 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.
12. 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. 2013.
13. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. Stat Med. 1997;16:259–72.
14. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Stat Sin. 2004;14:809–34.
15. Gould AL, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, Bois FY. 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;34:2181–95.
16. Touloumi G, Pantazis N, Babiker AG, Walker SA, Katsarou O, Karafoulidou A, Hatzakis A, Porter K. Differences in HIV RNA levels before the initiation of antiretroviral therapy among 1864 individuals with known HIV-1 seroconversion dates. AIDS. 2004;18:1697–705.
17. 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. 2014;57:185–200.
18. Luo S, Wang J. Bayesian hierarchical model for multiple repeated measures and survival data: an application to Parkinson's disease. Stat Med. 2014;33:4279–91.
19. 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;14:863–83.
20. Ghisletta P. Application of a joint multivariate longitudinal-survival analysis to examine the terminal decline hypothesis in the Swiss Interdisciplinary Longitudinal Study on the Oldest Old. J Gerontol B Psychol Sci Soc Sci. 2008;63:185–P192.
21. Battes LC, Caliskan K, Rizopoulos D, Constantinescu AA, Robertus JL, Akkerhuis M, Manintveld OC, Boersma E, Kardys I. Repeated measurements of NT-pro-B-type natriuretic peptide, troponin T or C-reactive protein do not predict future allograft rejection in heart transplant recipients. Transplantation. 2015;99:580–5.
22. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. Stata J. 2013;13:165–84.
23. Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. Am Stat. 2004;58:16–24.
24. Philipson P, Sousa I, Diggle PJ, Williamson PR, Kolamunnage-Dona R, Henderson R. Package "joineR" [Internet]. R Foundation for Statistical Computing; 2012. Available from: https://cran.r-project.org/package=joineR.
25. Rizopoulos D. The R Package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *arXiv Prepr.* 2014; arXiv:1404.
26. 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. 2009;30:1–3.
27. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. J Stat Softw. 2010;35:1–33.
28. Proust-Lima C, Philipps V, Liquet B. Estimation of latent class linear mixed models: the new package lcmm. *arXiv Prepr.* [Internet] 2015; Available from: http://arxiv.org/pdf/1503.00890.pdf.
29. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat Med. 2011;30:1366–80.
30. Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: A review. Stat Methods Med Res. 2012;23:42–59.
31. Sousa I. A review on joint modelling of longitudinal measurements and time-to-event. Revstat Stat J. 2011;9:57–81.
32. McCrink LM, Marshall AH, Cairns KJ. Advances in joint modelling: A review of recent developments with application to the survival of end stage renal disease patients. Int Stat Rev. 2013;81:249–69.
33. 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.
34. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. J Am Stat Assoc. 1999;94:1096–120.
35. Diggle PJ, Sousa I, Chetwynd AG. Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. Stat Med. 2008;27:2981–98.
36. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. J. Am. Stat. Assoc. 1995. 1112–1121.
37. Lin H, McCulloch CE, Mayne ST. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. Stat Med. 2002;21:2369–82.
38. Song X, Davidian M, Tsiatis AA. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics. 2002;3:511–28.

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117

Page 15 of 15

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

40. Pantazis N, Touloumi G, Walker SA, Babiker AG. 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: Ser C: Appl Stat. 2005;54:405–23.

41. 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.

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

43. Pantazis N, Touloumi G. Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: a simulation study. Stat Med. 2007;26:5473–85.

44. Dantan E, Proust-Lima C, Letenneur L, Jacqmin-Gadda H. Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. Int J Biostat. 2008;4:1–26.

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

46. Albert PS, Shih JH. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. Ann Appl Stat. 2010;4:1517–32.

47. Liu F, Li Q. A Bayesian model for joint analysis of multivariate repeated measures and time to event data in crossover trials. Stat Methods Med Res. 2014;0:1–13.

48. Hu B, Li L, Greene T. Joint multiple imputation for longitudinal outcomes and clinical events that truncate longitudinal follow-up. Stat. Med. 2015

49. Tang AM, Tang NS. Semiparametric Bayesian inference on skew–normal joint modeling of multivariate longitudinal and survival data. Stat Med. 2015; 34: 824–843.

50. Jaffa MA, Woolson RF, Lipsitz SR. Slope estimation for bivariate longitudinal outcomes adjusting for informative right censoring by using a discrete survival model: application to the renal transplant cohort. J R Stat Soc Ser A Stat Soc. 2011;174:387–402.

51. Jaffa MA, Gebregziabher M, Jaffa AA. A Joint modeling approach for right censored high dimensiondal multivariate longitudinal data. J Biometrics Biostat. 2014; 5.

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

53. Hatfield LA, Boye ME, Carlin BP. Joint modeling of multiple longitudinal patient-reported outcomes and survival. J Biopharm Stat. 2011;21:971–91.

54. He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. Stat Methods Med Res. 2013; 0: 1–13.

55. Luo S. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. Stat Med. 2014;33:580–94.

56. Andrinopoulou E-R, Rizopoulos D, Takkenberg JJM, Lesaffre E. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. Stat Methods Med Res. 2015;0:1–18.

57. Proust-Lima C, Dartigues J-F, Jacqmin-Gadda H. Joint modelling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. Stat. Med. 2015. In press.

58. Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. Biometrics. 2001;57:81–7.

59. Andrinopoulou E-R, Rizopoulos D, Takkenberg JJM, Lesaffre E. Joint modeling of two longitudinal outcomes and competing risk data. Stat Med. 2014;33:3167–78.

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

61. Bartolucci F, Farcomeni A. A discrete time event-history approach to informative drop-out in multivariate latent Markov models with covariates. Biometrics. 2015;71:80–9.

62. Crowther MJ. Extensions to the stjm package. Leicester University; 2015.

63. Fieuws S, Verbeke G, Maes B, Vanrenterghem Y. Predicting renal graft failure using multivariate longitudinal profiles. Biostatistics. 2008;9:419–31.

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

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

66. 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;96:906–14.

67. Wang C, Douglas J, Anderson S. Item response models for joint analysis of quality of life and survival. Stat Med. 2002;21:129–42.

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

69. Baghfalaki T, Ganjali M, Berridge D. Robust joint modeling of longitudinal measurements and time to event data using normal/independent distributions: a Bayesian approach. Biometrical J. 2013;55:844–65.

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

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

72. Rizopoulos D, Verbeke G, Molenberghs G. Shared parameter models under random-effects misspecification. Biometrika. 2008;95:63–74.

73. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics. 2003;59:221–8.

74. Escobar MD. Estimating normal means with a Dirichiet process prior. J Am Stat Assoc. 2012;89:268–77.

75. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21:2175–97.

76. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007;26:2389–430.

77. 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;53:750–63.

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

79. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Stat Methodol. 1977;39:1–38.

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

81. Lunn D, Spiegelhalter DJ, Thomas A, Best N. The BUGS project: evolution, critique and future directions. Stat Med. 2009;28:3049–67.

82. Tsonaka R, Verbeke G, Lesaffre E. A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. Biometrics. 2009;65:81–7.

83. Little RJA, Rubin DB. Statistical Analysis With Missing Data. Wiley Ser. Probab. Math. Stat. New York: Wiley; 1987.

84. Fieuws S, Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics. 2006;62:424–31.

85. Plummer M. JAGS : A program for analysis of Bayesian graphical models using Gibbs sampling. In: Hornik K, Leisch F, Zeileis A, editors. Proc. 3rd Int. Work. Distrib. Stat. Comput. Vienna, Austria; 2003.

86. Doran HC, Lockwood JR. Fitting value-added models in R. J Educ Behav Stat. 2006;31:205–30.

87. Crowther MJ. The stjm package in Stata: Joint modeling of longitudinal and survival data. Jt. Stat. Meet. Seattle; 2015.

88. 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;12:182–99.

89. Chen LM, Ibrahim JG, Chu H. Sample size and power determination in joint modeling of longitudinal and survival data. Stat Med. 2011;30:2295–309.