Computational Statistics and Data Analysis 170 (2022) 107438

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

A fast approximate EM algorithm for joint models of survival and multivariate longitudinal data

James Murray *, Pete Philipson

School of Mathematics, Statistics and Physics, Newcastle University, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:
Received 13 August 2021
Received in revised form 21 January 2022
Accepted 21 January 2022
Available online 31 January 2022

Keywords:
Joint models
EM algorithm
Multivariate normal
Simulation
Multiple longitudinal

ABSTRACT

Joint models are an increasingly popular way to characterise the relationship between one or more longitudinal responses and an event of interest. However, for multivariate joint models the increased dimensionality and complexity of random effects present in the model specification are commensurate with increased computing time, hampering the implementation of many classic approaches. An approximate EM algorithm which ameliorates the so-called 'curse of dimensionality' is developed. The scaleability and accuracy of the proposed method are demonstrated via two simulation studies and applied to data arising from two clinical trials in the disease areas of cirrhosis and Alzheimer's disease, each with three biomarkers.

1. Introduction

Many scientific investigations, such as clinical trials, involve the repeated measurement of continuous responses over a period of follow-up as well as an event time of interest. For example, numerous biomarker measurements in a randomized drug trial and the time to mortality from the baseline visit. A raft of such studies are detailed at the LONI Image Data Archive https://ida.loni.usc.edu. When interest then falls on characterising the relationships between these longitudinal responses and the event time, joint models have emerged as a popular modelling strategy, exhibiting superiority over both separate analyses of the two data types and two-stage approaches.

At heart, a joint model consists of two or more sub-models with (at least) some shared random effects that are combined into one larger meta-model by linking the shared random effects. Following on from the early work by Wulfsohn and Tsiatis (1997), commonly adopted submodels are a linear mixed effects model (Laird and Ware, 1982) and a Cox proportional hazards model (Cox, 1972) for the longitudinal and survival components respectively. Justification for joint models over naive or two-stage approaches abound in the literature - see Ibrahim et al. (2010) for one such example.

Until recently, joint models, and their application, have predominantly appeared in the literature in the form of a joint model specification with a single longitudinal outcome ('*univariate* joint modelling'). Joint models with more than one longitudinal response ('*multivariate* joint modelling') have existed in literature for nearly two decades, but are usually focused on methodological developments (Lin et al., 2002; Song et al., 2002) and bring with them unique challenges. Simply fitting several separate univariate joint models does not take into account correlations between the longitudinal responses of interest (Lin et al., 2002) and is tantamount to omitting potentially important responses in the survival sub-model, which could be to the detriment of prediction (Hickey et al., 2018a). Whilst one review of computational methods found a relative

* Corresponding author.
E-mail addresses: j.murray7@ncl.ac.uk (J. Murray), peter.philipson1@ncl.ac.uk (P. Philipson).

https://doi.org/10.1016/j.csda.2022.107438

abundance of software for fitting univariate joint models (Furgal et al., 2019), another for multivariate joint models found software for fitting such models was lacking (Hickey et al., 2016), only recently remedied, in part, by development of new R software packages joineRML (Hickey et al., 2018a) and extensions to existing R packages, such as JMBayes (Rizopoulos, 2016). However, despite this progress, it is not clear that existing methods will prove scaleable to situations involving many longitudinal measurements, rich random effect structures, or both.

Joint models were first proposed by Wulfsohn and Tsiatis (1997) who utilised numerical integration (via low-dimensional Gauss-Hermite quadrature) as part of an EM algorithm in order to circumvent the dependence on the unobserved random effects on the parameter estimates by treating them as missing data. In their relatively simple scenario this worked well and became the de facto approach in the early years of joint models. However, when the number of longitudinal measures, the complexity of the random effects structure, or both, grows then this approach becomes less viable due to the inherent computational burden - essentially the approach is not particularly scaleable - which has led to alternative fitting procedures being utilised throughout literature: For instance Monte Carlo techniques (Henderson et al., 2000; Lin et al., 2002; Hickey et al., 2018a) and Laplace approximations (Rizopoulos et al., 2009). However, as Hickey et al. (2018a) draw attention to, the perpetually increasing volumes of data collected by clinical trials with very many longitudinal responses and increasingly complex electronic healthcare records would likely require approximate methods for the numerical integration in the E-step.

One such approximate method is a recent approach put forward by Bernhardt et al. (2015) in the context of a joint model with a binary outcome. The authors proposed a normal approximation on the distribution of the random effects *for each individual* conditional on the observed data, which has the effect of reducing the dimensionality of each required integral to be uniformally one, regardless of the complexity of the random effects, thus improving computational efficiency.

The rest of the paper establishes the proposed multivariate joint model, which is an extension to the proposed model by Bernhardt et al. (2015) to the case of a survival outcome in place of a binary outcome. We first present the multivariate joint model in Section 2 and outline the standard approach to analysis, before introducing the approximate EM algorithm. Results from two comprehensive simulation studies are presented in Section 3 before two clinical applications to primary biliary cirrhosis data and Alzheimer's data respectively are demonstrated in Section 4. As far as the authors are aware, the implementation of an approximate EM algorithm to a joint model with a time-to-event sub-model is novel.

2. Methods

In the following we set out the multivariate longitudinal and survival sub-models alongside some accompanying notation. We then define the joint likelihood we wish to maximise in the context of the EM algorithm before setting out the approximate EM algorithm.

2.1. Models and notation

For each subject $i = 1, \dots, n$ we observe $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}, \dots, \boldsymbol{Y}_{iK})$ where each $\boldsymbol{Y}_{ik},\ k = 1, \dots, K$ denotes the k^{th} longitudinal response vector of interest $\boldsymbol{Y}_{ik} = \left(y_{i1k}, \dots, y_{im_{ik}k}\right)$, where the K responses are measured m_{ik} times, which can differ between subjects and longitudinal responses. We observe a (possibly right-censored) event time $T_i = \min(T_i^*, C_i)$, where T_i^* is the true event time and C_i the potential censoring time. In addition we introduce a failure indicator Δ_i which is set equal to one if $T_i^* < C_i$ and zero otherwise. It is assumed that the censoring process is independent and non-informative.

We adopt the following linear mixed-effects model for the k^{th} longitudinal response

$$\begin{aligned} &\boldsymbol{Y}_{ik} = X_{ik}(t)\boldsymbol{\beta}_k + Z_{ik}(t)\boldsymbol{b}_{ik} + \boldsymbol{\varepsilon}_{ik} \\ &\boldsymbol{b}_{ik} \sim N(0, D_k), \qquad \boldsymbol{\varepsilon}_{ik} \sim N_{m_{ik}}(0, \sigma^2_{\varepsilon_k}), \qquad \boldsymbol{b}_{ik} \perp\!\!\!\perp \varepsilon_{ik}, \end{aligned} \tag{1}$$

where X_{ik} is the (possibly time-dependent) fixed effects design matrix and $\boldsymbol{\beta}_k$ the corresponding p_k-vector of coefficients. Likewise, Z_{ik} is a (possibly time-dependent) random effects design matrix and $\boldsymbol{b}_{ik}$ the subject-specific q_k-vector of random effects. These random effects are assumed to follow a multivariate normal distribution with zero-mean and positive-definite covariance matrix D_k.

The sub-model for the time-to-event response is

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ \boldsymbol{K}_i^T \boldsymbol{\eta} + \sum_{k=1}^{K} \gamma_k Z_{ik}(t)\boldsymbol{b}_{ik} \right\}, \tag{2}$$

where $\boldsymbol{K}_i$ is the p_S-vector of baseline covariates of interest to the event time process; $\boldsymbol{\eta}$ the corresponding vector of fixed effects and $\lambda_0(t)$ an unspecified baseline hazard. In order to establish a latent association we introduce the random effects structure from the linear mixed model (1) with association parameter vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)$ although other formulations are possible within this framework (Hickey et al., 2016). Going forward, we assume that the structure of the random effects is in the form of a random intercept and slope, $Z_{ik}\boldsymbol{b}_{ik} = b_{0ik} + b_{1ik}\boldsymbol{t}_{ik}\ \forall\ k = 1, \dots, K$ unless stated otherwise. We note that the induced association between (1) and (2) is a relatively simple one, however even with a random effects structure such as this, under traditional approaches it is expected that computational times will grow exponentially large as the number of longitudinal responses – thus the dimensionality of the random effects and subsequent integration – grows large.

2.2. Likelihood

For each subject i we construct block diagonal matrices for the covariate information, $X_i = \bigoplus_{k=1}^{K} X_{ik}$, the random effect structure $Z_i = \bigoplus_{k=1}^{K} Z_{ik}$ and error terms $V_i = \bigoplus_{k=1}^{K} \sigma^2_{\varepsilon_k} \mathbb{I}_{m_{ik}}$. Here, $\mathbb{I}_x$ is an $x \times x$ identity matrix and $\bigoplus$ denotes the direct matrix sum. Additionally we define $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}, \ldots, \boldsymbol{b}_{iK})$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ as the collection of subject-specific random effects and fixed effects, respectively, across the K longitudinal responses and D is the covariance block matrix across the K responses consisting of the D_k on the diagonal and the covariance matrices *between* responses on the off-diagonal.

The observed data likelihood for the joint model is then

$$\prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\beta}, \sigma^2_{\varepsilon_1}, \ldots, \sigma^2_{\varepsilon_K}) f(T_i, \Delta_i|\boldsymbol{b}_i; \boldsymbol{\eta}, \boldsymbol{\gamma}) f(\boldsymbol{b}_i|D) d\boldsymbol{b}_i \right\}. \tag{3}$$

Where $f(\boldsymbol{Y}_i|\cdot)$ is the density for the longitudinal responses, $f(T_i, \Delta_i|\cdot)$ that for the time to event and $f(\boldsymbol{b}_i|\cdot)$ the random effects. The detailed forms of each density present in (3) are given in Appendix A. Subsequently, we introduce the parameter vector $\boldsymbol{\Omega} = \left(\boldsymbol{\beta}, \sigma^2_{\varepsilon_1}, \ldots, \sigma^2_{\varepsilon_K}, \boldsymbol{\eta}, \boldsymbol{\gamma}, \text{vech}(D)\right)$, where $\text{vech}(D)$ denotes the half-vectorisation of covariance matrix D, thus containing all unique elements.

2.3. Parameter estimation via the EM algorithm

When faced with missing data in the form of unobserved random effects, the Expectation Maximisation (EM) algorithm is a useful and widely-used construct, allowing for maximisation of the observed data (log) likelihood via the construction of a complete data (log) likelihood - whose expected value is found in the E-step, with respect to the unobserved random effects conditional on the observed data - and a series of maximum likelihood parameter updates that form the subsequent M-step. Under a joint modelling framework, the *complete* data for subject i is $\{\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{b}_i\}$; all are observed other than the random effects $\boldsymbol{b}_i$, which, as such, are treated as missing data.

The form of the expected log-likelihood for subject i in the E-step is then

$$Q(\boldsymbol{\Omega}|\hat{\boldsymbol{\Omega}}) = \sum_{i=1}^{n} E_i \left[\log f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \hat{\boldsymbol{\Omega}}) + \log f(T_i, \Delta_i|\boldsymbol{b}_i; \hat{\boldsymbol{\Omega}}) + \log f(\boldsymbol{b}_i|\hat{\boldsymbol{\Omega}}) \right]. \tag{4}$$

Such an expectation is taken with respect to the distribution of the random effects conditional on the observed data at the current set of parameter estimates $f(\boldsymbol{b}_i|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}})$, with $\hat{\boldsymbol{\Omega}} = \left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_{\varepsilon_1}, \ldots, \hat{\sigma}^2_{\varepsilon_K}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}}, \text{vech}(\hat{D})\right)$ denoting said set.

The form of the M-step is widely reported in literature (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Lin et al., 2002; Bernhardt et al., 2015), and closed form updates exist for all parameters except for $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$, which are instead jointly updated by a one-step Newton-Raphson iteration. In practise, the M-step is formulated by computing n sets of expectations of the form $E_i \left[g(\boldsymbol{b}_i)|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}} \right]$ in the E-step, where $g(\boldsymbol{b}_i)$ denotes some function of the random effects. For each estimate in the E- and M-steps, expectations are required with respect to the distribution $f(\boldsymbol{b}_i|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}})$. Wulfsohn and Tsiatis (1997) showed that these could be calculated using Gauss-Hermite quadrature. However, as the number of random effects increases, it becomes infeasible to use this approach, which has led to the use of alternative methods for computation of the necessary expectations in the E-step (Henderson et al., 2000; Lin et al., 2002; Rizopoulos et al., 2009; Rizopoulos, 2012; Hickey et al., 2018a).

2.4. Approximate EM algorithm

As the number of longitudinal responses and/or the dimensionality of the random effects in (3) grows large, existing methods are likely to become computationally expensive. Bernhardt et al. (2015) proposed approximating the distribution of the random effects conditional on the observed data at the current set of parameter estimates as multivariate normal in a joint model with a binary outcome. This approximation then takes advantage of the fact that any linear combination of $\boldsymbol{b}_i$ would also be normal. Specifically, Bernhardt et al. (2015) use the normal approximation

$$\boldsymbol{b}_i|\boldsymbol{Y}_i, T_i, \Delta_i; \boldsymbol{\Omega} \overset{\text{appx.}}{\sim} N(\hat{\boldsymbol{b}}_i, \hat{\Sigma}_i), \tag{5}$$

where $\hat{\boldsymbol{b}}_i$ is the value of the random effects $\boldsymbol{b}_i$ which maximises the complete data log-likelihood at the current set of parameter estimates. Indeed, the posterior distribution of random effects in generalized linear mixed models have been shown to be asymptotically normal (Baghishani and Mohammadzadeh, 2012). The covariance matrix is defined as

$$\hat{\Sigma}_i = \left\{ -\frac{\partial^2 \log f(\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{b}_i; \boldsymbol{\Omega})}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T} \bigg|_{\boldsymbol{b}_i = \hat{\boldsymbol{b}}_i} \right\}^{-1}.$$

Using the approximation in (5), the expectations necessary in the E-step are calculated with respect to a one-dimensional normal distribution. For example, for the update to fixed effect parameter $\boldsymbol{\beta}$ we require the expectation $E_i[Z_i\boldsymbol{b}_i]$. In actuality this expectation – and ones to follow – are conditional upon the observed data; for ease of presentation we omit these terms. Using the approximation (5), this expectation is taken with respect to the univariate normal distribution $N(Z_i\hat{\boldsymbol{b}}_i, Z_i\hat{\Sigma}_i Z_i^T)$. Likewise, in the update for $\sigma^2_{\varepsilon_k}$ we require the expectation $E_i[X_{ik}(t)\boldsymbol{\beta}_k + Z_{ik}\boldsymbol{b}_{ik}]$ which is taken with respect to the univariate normal distribution $N(X_{ik}(t)\boldsymbol{\beta}_k + Z_{ik}\hat{\boldsymbol{b}}_{ik}, Z_{ik}\hat{\Sigma}_{ik}Z_{ik}^T)$. Here, $\hat{\Sigma}_{ik}$ refers to the matrix along the block-diagonal of $\hat{\Sigma}_i$ associated with the random effects structure on the k^{th} response. For our survival sub-model, in order to update the baseline hazard $\lambda_0(t)$, association parameters $\boldsymbol{\gamma}$ and the fixed-effect survival coefficients $\boldsymbol{\eta}$ we require the expectation $E_i\left[\exp\left\{\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^K \gamma_k Z_{ik}\boldsymbol{b}_{ik}\right\}\right]$ which once again is taken with respect to the univariate normal distribution $N(\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^K \gamma_k Z_{ik}\hat{\boldsymbol{b}}_{ik}, \sum_{k=1}^K \gamma_k^2 Z_{ik}\hat{\Sigma}_{ik}Z_{ik}^T)$.

We use these normal distributions in tandem with Gauss-Hermite quadrature using three abscissae to evaluate the necessary expectations in the E-step of the approximate EM algorithm. The form of the M-step updates for all parameters are presented in Appendix B. We did not obtain material benefit from utilising the nine abscissae originally recommended by Bernhardt et al. (2015), though we speculate this could be an artefact of the logistic sub-model used there. Although somewhat surprising, the use of three quadrature points has been verified previously (Wulfsohn and Tsiatis, 1997; McCrink et al., 2013). We recognise that Monte Carlo methods could also be used here, but they are not explored in this instance.

2.5. Starting values and convergence details

1. We begin by obtaining initial estimates for the parameter vector $\boldsymbol{\Omega}$. This deviates somewhat from that outlined in Bernhardt et al. (2015):
   i. Firstly, rather than obtaining initial conditions for the parameters present in the longitudinal sub-model from a series of univariate linear mixed model fits, we instead implement an EM algorithm for a multivariate mixed model. The (nested) initial conditions for this EM algorithm come from K linear mixed model fits, which can be obtained using standard packages, such as `nlme` (Pinheiro et al., 2021) or `lme4` (Bates et al., 2015). Additionally, each of these K model fits allow one to obtain the best linear unbiased predictor of the random effects for the k^{th} longitudinal response, $\boldsymbol{b}_k$.
   ii. These random effects estimates are then used as covariates in a time-varying Cox model (using e.g. `survival` (Therneau, 2015)), along with covariates $\boldsymbol{K}_i$ which gives rise to initial conditions for $(\boldsymbol{\gamma}, \boldsymbol{\eta})$.
2. Maximise $\log f(\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{b}_i; \boldsymbol{\Omega})$ with respect to $\boldsymbol{b}_i$ – through the use of optimisation function `ucminf` (Nielsen and Mortensen (2016), see Appendix C) – in order to obtain $\hat{\boldsymbol{b}}_i$ and subsequently $\hat{\Sigma}_i$.
3. Use approximation (5) to update parameter vector $\boldsymbol{\Omega}^{(m)} \rightarrow \boldsymbol{\Omega}^{(m+1)}$.
4. Repeat steps 2. and 3. until the algorithm converges, which we define to have occurred when for the P parameters which construct $\boldsymbol{\Omega}$,

$$\max\left(\frac{|\boldsymbol{\Omega}_1^{(m+1)} - \boldsymbol{\Omega}_1^{(m)}|}{|\boldsymbol{\Omega}_1^{(m)}| + \nu}, \ldots, \frac{|\boldsymbol{\Omega}_P^{(m+1)} - \boldsymbol{\Omega}_P^{(m)}|}{|\boldsymbol{\Omega}_P^{(m)}| + \nu}\right) < \xi,$$

   for some predetermined ξ which is sufficiently small. We introduce some small value ν to the denominators on this relative difference calculation to avoid numerical issues when parameters are close to zero. We employ the relative difference criterion due to the parameters being on different scales.

After the algorithm is deemed to have converged, standard errors could be obtained by inverting the observed information matrix at our maximum likelihood estimates $\hat{\boldsymbol{\Omega}}$. However, we note in practise that this is computationally expensive. Instead, we obtain standard errors via an approximation of the observed empirical information matrix (Lin et al., 2002; McLachlan and Krishnan, 2008; Hickey et al., 2018a)

$$\tilde{I}(\boldsymbol{\Omega}) = \sum_{i=1}^{n} s_i(\boldsymbol{\Omega}) s_i(\boldsymbol{\Omega})^T - n^{-1} S(\boldsymbol{\Omega}) S(\boldsymbol{\Omega})^T,$$

where $s_i(\boldsymbol{\Omega})$ denotes the gradient vector of the conditional expectation of the complete data log-likelihood function (4) i.e. the score statistic and $S(\boldsymbol{\Omega}) = \sum_{i=1}^n s_i(\boldsymbol{\Omega})$. McLachlan and Krishnan (2008) note that the right hand side at the MLEs $\hat{\boldsymbol{\Omega}}$ should equal zero. However, it is included for completeness' sake as $\hat{\boldsymbol{\Omega}}$ are not technically MLEs (Bernhardt et al., 2015). These standard errors allow for derivation of confidence intervals for parameters we estimate. Bootstrap methods could additionally be used at greatly increased computational expense. Hickey et al. (2018a) provide a good, brief, overview of the pros and cons of the approaches to the standard error approximation adopted here.

With the methodology established, we undertake two simulation studies followed by two application case studies: One using primary biliary cirrhosis data and another using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). In the simulation studies, we control failure rate to allow for the approximate EM algorithm to have a longer, or shorter,

longitudinal profile. We also consider a range of sample sizes and vary the dimension of the longitudinal data, with choices of K of 3, 5 and 10 in order to fully demonstrate the accuracy and utility of the proposed method. All simulations and applications in the following sections were executed on a Macbook Air 1.8 GHz Intel core i5 with 8GB RAM using `R` version 4.0.3. `R` code used to generate the simulation studies as well as to fit the approximate EM algorithm is available at https://github.com/jamesmurray7/ApproximateEM.

3. Simulations

3.1. Simulation I

We first consider a simulation with the same longitudinal, and random effect, parameter specification as that in Philipson et al. (2020). Specifically, we assume $K = 3$ longitudinal responses ('trivariate') for a simulated sample size of $n \in \{250, 500, 1000\}$. The joint model for $k = 1, \dots, K$ is specified as

$$\begin{cases} \boldsymbol{Y}_{ik} & = (\beta_{k0} + b_{ik0}) + (\beta_{k1} + b_{ik1})\boldsymbol{t}_{ik} + \beta_{k2}x_{1i} + \beta_{k3}x_{2i} + \boldsymbol{\varepsilon}_{ik} \\ \lambda_i(t) & = \lambda_0(t)\exp\left\{\eta_1 x_{1i} + \eta_2 x_{2i} + \sum_{k=1}^{K}\gamma_k(b_{ik0} + b_{ik1}t)\right\}, \end{cases}$$

with $\boldsymbol{\varepsilon}_{ik} \sim N(0, \sigma_k^2)$ and $\boldsymbol{t}_{ik}$ denoting the times the longitudinal measurement is available for subject i for response k; $\boldsymbol{t}_{ik} = (0, \dots, m_{ik})$. Additionally, x_{1i} and x_{2i} represent a set baseline variables for subject i, in the form of a continuous $N(0, 1)$ and binary $\text{Bin}(1, 0.5)$ variable. The random effects $\boldsymbol{b}_i$ are drawn from the multivariate normal $N_6(\boldsymbol{0}, D)$ and a random intercept and slope is used in each model as shown above.

We set the true parameter value of the fixed effects $\boldsymbol{\beta}$ to be $\boldsymbol{\beta}_1 = (0, 1, 1, 1)$; $\boldsymbol{\beta}_2 = (0, -1, 0, 0.5)$ and $\boldsymbol{\beta}_3 = (0, 0, 0.5, -0.5)$; variance components $\sigma_k^2 = 0.25\ \forall\ k = 1, \dots, K$ and elements of the covariance matrix $D_{11} = D_{33} = D_{55} = 0.5^2$; $D_{22} = D_{44} = D_{66} = 0.2^2$; $D_{13} = D_{35} = -0.125$, $D_{15} = 0.125$ and all remaining elements set to 0. The fixed effects in the survival model were $\boldsymbol{\eta} = (-0.10, 0.30)$ and the association parameters were $\boldsymbol{\gamma} = (-0.50, 0.75, 0.50)$. As well as altering the simulated sample size, we alter the length of the profiles along with the proportion who fail in order to explore performance properties under profiles where increasingly more information is available in the algorithm: Simulating data under three profile lengths, one with a maximum profile length of six time-points on approximately a 50% failure rate ($\boldsymbol{t}_{ik} = (0, 1, \dots, 5)$, 'short'); another with a maximum profile length of ten ('medium') under approximately 40% failure and one further ('long') with a maximum profile length of fifteen time-points with 30% failure rate. We simulate data under each of these combinations of profile lengths and sample size one hundred times.

Fig. 1 illustrates the accuracy in parameter estimates obtained for $(\boldsymbol{\eta}, \boldsymbol{\gamma})$ by the approximate EM algorithm under the different profile lengths and sample sizes outlined above. In addition, using 95% coverage probability (CP, calculated from Wald-like confidence intervals across the one hundred simulations), we observe that the survival parameters $(\boldsymbol{\gamma}, \boldsymbol{\eta})$ have good coverage; full details on a parameter-by-parameter basis are given in the Supplementary Materials. The parameters associated with the longitudinal sub-models $\boldsymbol{\beta}$ and $\sigma_1^2 \dots, \sigma_3^2$ are well-estimated with minimal coverage 0.91 observed across simulations for these parameters. Additionally, the variance of the random effects $\boldsymbol{b}_i$ (i.e. $\text{diag}(D)$) are estimated with minimal CP 0.92. Full tabulation of all parameter estimates under the different profile lengths and sample sizes are given in the Supplementary Materials.

An additional simulation study where we inflate the baseline hazard by introducing a large $\boldsymbol{\eta}$ such that the distribution of survival times is skewed (with many early failure times) is presented in the Supplementary Materials. Here we observe that owing to $\boldsymbol{\gamma}$ being time varying, CP of the true values is much more accurate as the profile length increases from six to fifteen time-points.

3.2. Simulation II

Extending beyond simply the trivariate case, we seek to explore capabilities of the approximate EM algorithm in cases when a plethora of longitudinal biomarkers exist. To that end, we explore and compare model fitting capabilities when three, five and ten longitudinal responses are used in the multivariate joint model. We once more simulate data under the joint model

$$\begin{cases} \boldsymbol{Y}_{ik} & = (\beta_{k0} + b_{ik0}) + (\beta_{k1} + b_{ik1})\boldsymbol{t}_{ik} + \beta_{k2}x_{1i} + \beta_{k3}x_{2i} + \boldsymbol{\varepsilon}_{ik} \\ \lambda_i(t) & = \lambda_0(t)\exp\left\{\eta_1 x_{1i} + \eta_2 x_{2i} + \sum_{k=1}^{K}\gamma_k(b_{ik0} + b_{ik1}t)\right\}, \end{cases}$$

with $\boldsymbol{\varepsilon}_{ik} \sim N(0, \sigma_k^2)$ and $\boldsymbol{t}_{ik}$ denoting the times the longitudinal measurement is available for subject i for response k; $\boldsymbol{t}_{ik} = (0, \dots, m_{ik})$. Continuous (standard normal) and Bernoulli ($p = 0.5$) baseline covariates are represented via x_{1i} and x_{2i} respectively. The random effects $\boldsymbol{b}_i$ are drawn from the multivariate normal $N_{2K}(\boldsymbol{0}, D)$ and a random intercept and slope is used in each model as shown above. To obtain sensible true parameter values we use univariate model fits from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data, as a real wealth of biomarkers exist here, with their performance

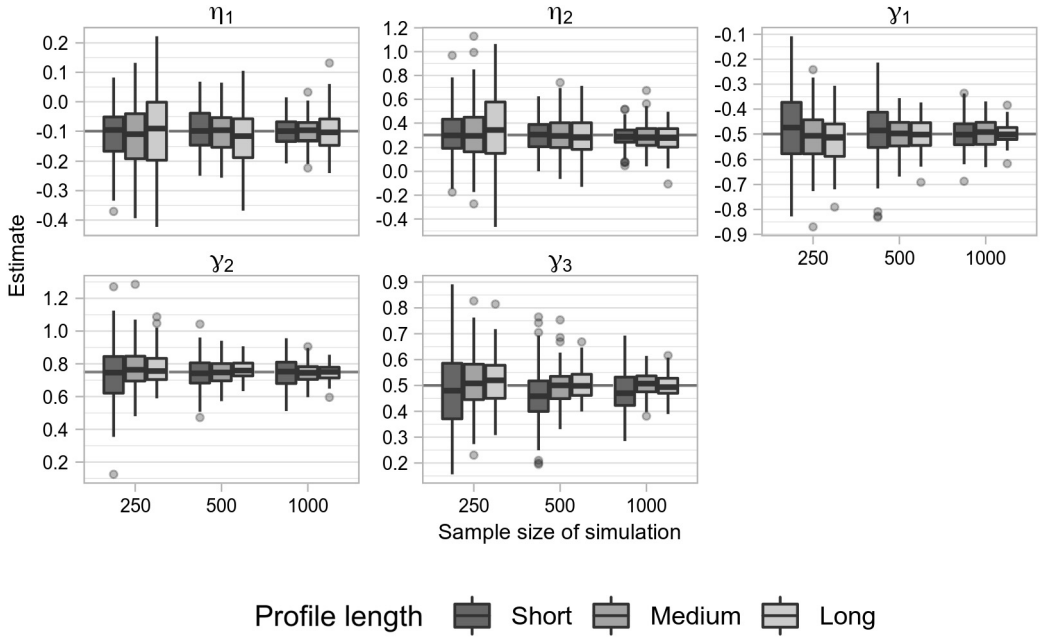

Fig. 1. Estimates for the trivariate model produced by the approximate EM algorithm for the parameters associated with the survival sub-model, $(\boldsymbol{\eta}, \boldsymbol{\gamma})$ on differing simulated sample sizes and profile lengths; horizontal lines signify the true parameter values.

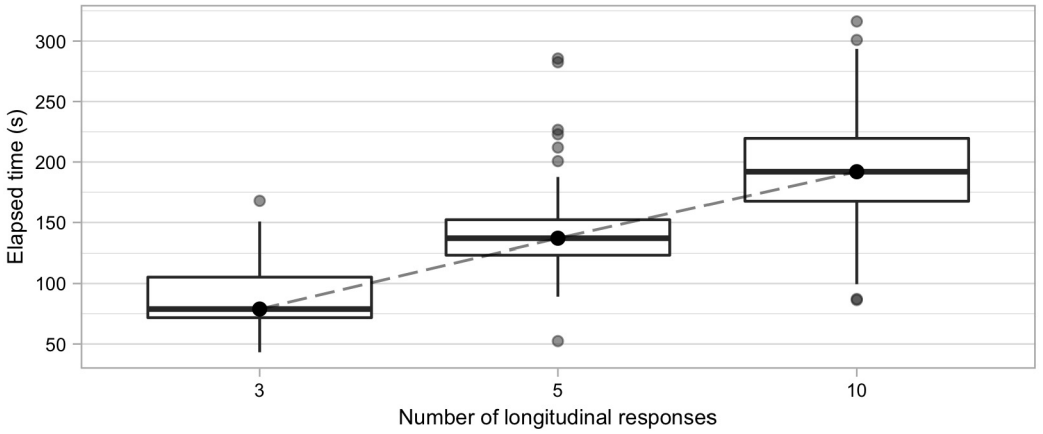

Fig. 2. Elapsed time (s) taken for the approximate EM algorithm to converge and standard error calculations to be performed for simulations on differing number of longitudinal responses, $K \in \{3, 5, 10\}$. Simulated sample size is $n = 500$ and failure rate approximately 50%.

in univariate joint models established previously (Li et al., 2017). Full details are given in the Supplementary Materials and we consider an application of the method to the ADNI data itself in Section 4.2.

For all simulations we use sample size $n = 500$ and a failure rate which mimics that of conversion to Alzheimer's in the ADNI study ($\approx 50\%$); simulated longitudinal profiles were allowed a maximum of fifteen timepoints, corresponding to the 75^{th} percentile observed in the ADNI data.

Given the available profile length of fifteen timepoints in light of the results presented in the previous section, parameters are generally well-estimated under the $K \in \{3, 5, 10\}$ simulation conditions here. A minimal coverage probability of 0.89 is observed for $(\boldsymbol{\gamma}, \boldsymbol{\eta})$ under $K = 3$ and the longitudinal parameters equally well estimated across K. Full tabulation, as well as graphical representation of these parameter estimates is given in the Supplementary Materials. Fig. 2 shows the progression of time taken for the approximate EM algorithm to converge and standard errors be calculated as the number of longitudinal responses simulated – thus the dimensionality of random effects – increases. Median time for convergence of the algorithm and calculation of the standard errors is 1.74 times slower moving from three to five responses, and only 1.40 times slower moving from five to the – much larger – ten. This lends credence to the scaleability of the approach used here.

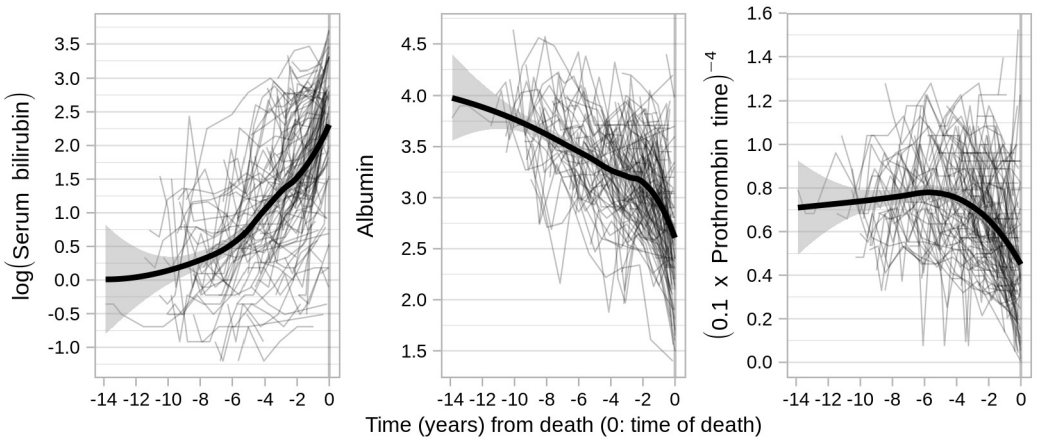

Fig. 3. Longitudinal trajectories for the three chosen biomarkers from the subset of the PBC data who experienced mortality. Black lines show individual trajectories and the thick black line show smoothed (LOESS) curve along with a shaded band representing the 95% confidence interval.

4. Data analysis

4.1. Application I

We consider the application of the approximate EM algorithm outlined in Section 2 to the set of primary biliary cirrhosis (PBC) data. The progression of PBC was studied in 312 patients between 1974 and 1984 (Murtaugh et al., 1994). The existence of multiple longitudinal biomarkers as well as information regarding a time to event of interest has led to the PBC data becoming a popular example in existing literature (Hickey et al., 2018a; Dai and Pan, 2018; Andrinopoulou and Rizopoulos, 2016; Dil and Karasoy, 2020). The PBC data has two event times of interest: Mortality or received transplant. For the purposes of this application we consider only mortality as the event time of interest.

We fit a trivariate joint model on three longitudinal responses of interest: serum bilirubin (measured in mg/dl); albumin (g/dl) and prothrombin time (seconds). The longitudinal trajectories for these three biomarkers are presented in Fig. 3. Furthermore, the profiles for serum bilirubin and prothrombin time appear to be quadratic in nature, which is reflected in our joint model specification below, wherein we additionally include a continuous covariate, (standardised) baseline age and a binary one, receiving the study drug (D-penicillamine):

$$\begin{aligned}
\log(\text{Serum Bilirubin}) &= (\beta_{10}+b_{10}) + (\beta_{11}+b_{11})\times \text{time} + (\beta_{12}+b_{12})\times \text{time}^2 \\
&\quad + \beta_{13}\times \text{age} + \beta_{14}\times \text{drug} \\
\text{Albumin} &= (\beta_{20}+b_{20}) + (\beta_{21}+b_{21})\times \text{time} + \beta_{22}\times \text{age} + \beta_{23}\times \text{drug} \\
(0.1\times \text{Prothrombin})^{-4} &= (\beta_{30}+b_{30}) + (\beta_{31}+b_{31})\times \text{time} + (\beta_{32}+b_{32})\times \text{time}^2 \\
&\quad + \beta_{33}\times \text{age} + \beta_{34}\times \text{drug} \\
\lambda(t) &= \lambda_0(t)\exp\left\{\eta_1\times \text{age} + \eta_2\times \text{drug} + \sum_k \gamma_k Z_k(t)\boldsymbol{b}_k\right\}.
\end{aligned}$$

Table 1 presents the estimates (SEs) for model parameters of interest from the above specification. Computation time required for the approximate EM algorithm to converge and standard errors be calculated was 52 seconds. The fitted values for $\boldsymbol{\beta}$ in Table 1 show that receiving the active arm does not significantly alter longitudinal trajectories for any of the three biomarkers studied, nor the overall hazard rate. Whilst log(serum bilirubin) increases over time, both albumin and transformed prothrombin decrease on average. Appraising now the values for $\boldsymbol{\gamma}$, we observe that patients who deviate upwards from the average study trajectory for log(serum bilirubin) are subject to increased hazard of experiencing the event. In contrast, those deviating upwards from the average trajectory in albumin measurements experience decreased hazard. Transformed prothrombin time does not appear to hold significant association with the hazard.

Table 1
Parameter estimates for application to primary biliary cirrhosis data. Estimates are reported along with standard errors and 95% confidence interval (CI). Biomarker names are included and correspond with transformations taken in Section 4.1. η_{age} and η_{drug} are time invariant survival parameters not corresponding to a specific biomarker and as such are reported separately.

	Parameter	Estimate	SE	95% CI
Serum bilirubin	β_{10}	0.534	0.088	[0.361, 0.707]
	β_{11}	0.160	0.026	[0.109, 0.212]
	β_{12}	0.003	0.003	[-0.003, 0.008]
	β_{13}	0.023	0.049	[-0.073, 0.120]
	β_{14}	-0.104	0.112	[-0.323, 0.115]
	σ_1^2	0.092	0.003	[0.087, 0.097]
	γ_1	1.132	0.143	[0.851, 1.413]
Albumin	β_{20}	3.541	0.035	[3.472, 3.610]
	β_{21}	-0.106	0.006	[-0.117, -0.094]
	β_{22}	-0.086	0.023	[-0.131, -0.042]
	β_{23}	0.032	0.044	[-0.054, 0.117]
	σ_2^2	0.102	0.002	[0.098, 0.106]
	γ_2	-1.452	0.483	[-2.398, -0.505]
Prothrombin	β_{30}	0.804	0.019	[0.766, 0.842]
	β_{31}	-0.031	0.008	[-0.047, -0.016]
	β_{32}	-0.003	0.001	[-0.004, -0.001]
	β_{33}	-0.042	0.012	[-0.064, -0.019]
	β_{34}	0.030	0.024	[-0.017, 0.076]
	σ_3^2	0.020	0.001	[0.018, 0.021]
	γ_3	-0.899	0.995	[-2.849, 1.051]
	η_{age}	0.703	0.096	[0.514, 0.892]
	η_{drug}	-0.087	0.182	[-0.444, 0.271]

4.2. Application II

The Alzheimer's Disease Neuroimaging Initiative (ADNI[1]) is a prospective cohort study which began in 2004. The first phase of the ADNI study (ADNI 1) recruited more than 800 individuals at differing stages of cognitive impairment (cognitively normal, mild cognitive impairment (MCI) and early Alzheimer's disease) - here we focus on the subset of patients with MCI at baseline. Patients re-attended multiple times across three years of follow-up. This gives rise to a wealth of longitudinal measures: neuropsychological assessments; brain imaging; clinical measures and other biomarkers such as cerebrospinal fluid in combination with the time to event of interest of progression to Alzheimer's disease ('AD') diagnosis.

Previous use of joint models for these data has considered a series of univariate fits, using standardised biomarkers which were subsequently ranked in terms of their significance in relation to conversion to Alzheimer's disease (Li et al., 2017). Subsequently, and recognising the difficulty in applying multivariate joint models, methods have been developed to reduce the dimensionality of the data using both functional principal component regression (Li and Luo, 2017) and partial least squares (Wang et al., 2020). The method introduced in this paper allows a multivariate joint model to be implemented directly.

We consider the trivariate model using the Alzheimer's Disease Assessment Scale-Cognitive subscale (13 item, 'ADAS-13'); Functional assessment questionnaire ('FAQ') and MRI volumetric data of the middle temporal gyrus ('MidTemp'), which were identified as the best-performing biomarkers in the cognitive, functional and neuroimaging domains, respectively, from univariate model fits (Li et al., 2017). For both ADAS13 and FAQ, a higher score indicates poorer cognition and greater functional dependence, respectively. Of the $n = 384$ of the original ADNI sample, we utilise only those with non-missing values for the three biomarkers above ($n = 359$; 93.4%). Fig. 4 displays the individual trajectories of the three biomarkers before AD conversion for the 188 (52%) patients who progressed to AD during study follow-up.

We construct each of the longitudinal sub-models to be of the following form (6) wherein one notes uniform use of the random intercept and slope. In the survival sub-model, we arbitrarily select only a continuous covariate in the form of the patient's (standardised) age at baseline visit and a binary one with presence of one or more apolipoprotein E $\varepsilon 4$ alleles ('APOE', present for 55% of subjects in the analysis sample), which we additionally adjust the longitudinal models for; time is measured as years from baseline visit.

$$\begin{cases} Y_k & = (\beta_{k0} + b_{k0}) + (\beta_{k1} + b_{k1}) \times \text{time} + \beta_{k2} \times \text{age} + \beta_{k3} \times \text{APOE}, \\ \lambda(t) & = \lambda_0(t) \exp\left\{\eta_1 \times \text{age} + \eta_2 \times \text{APOE} + \sum_{k=1}^{K} \gamma_k (b_{k0} + b_{k1} t)\right\}. \end{cases} \tag{6}$$

Table 2 presents the estimates and associated standard errors for the model parameters of interest in our application to the ADNI data. The time for the EM algorithm to converge and SEs be calculated was 98 seconds. Inspecting this trivariate

[1] Data used in preparation of this section were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

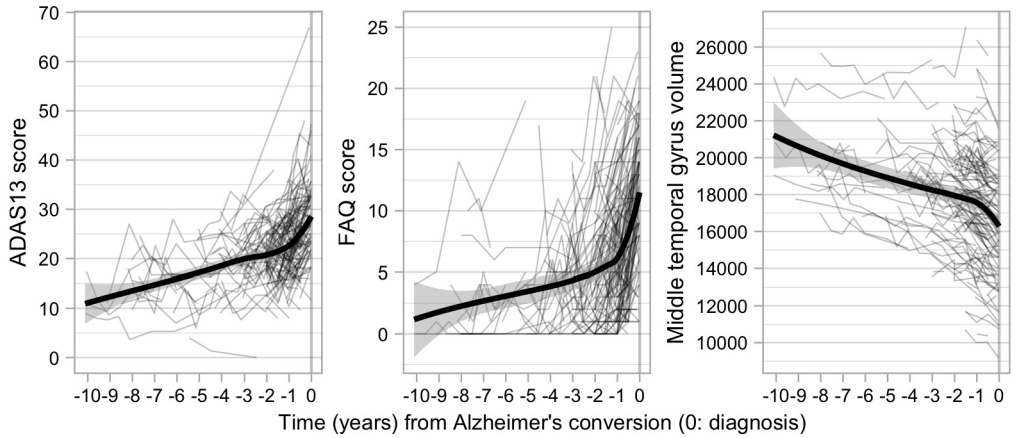

Fig. 4. Longitudinal trajectories for the three chosen biomarkers from the subset of the ADNI data who experienced AD conversion. Black lines show individual trajectories and the thick black line shows a smoothed (LOESS) curve along with a shaded band representing the 95% confidence interval.

Table 2
Parameter estimates for application to ADNI data. Estimates are reported along with standard errors and 95% confidence interval (CI). Biomarker names are included and parameter estimates correspond to their standardised forms; 'MidTemp' refers to the middle temporal gyrus volume. η_{age} and η_{APOE} are time invariant survival parameters not corresponding to a specific biomarker and as such are reported separately.

	Parameter	Estimate	SE	95% CI
ADAS-13	β_{10}	-0.104	0.058	[-0.218, 0.009]
	β_{11}	0.247	0.023	[0.201, 0.292]
	β_{12}	0.092	0.046	[0.001, 0.183]
	β_{13}	0.131	0.088	[-0.041, 0.303]
	σ_1^2	0.182	0.010	[0.163, 0.201]
	γ_1	0.646	0.080	[0.489, 0.802]
FAQ	β_{20}	-0.260	0.088	[-0.433, -0.087]
	β_{21}	0.356	0.026	[0.305, 0.407]
	β_{22}	0.023	0.057	[-0.089, 0.136]
	β_{23}	0.129	0.103	[-0.073, 0.330]
	σ_2^2	0.200	0.008	[0.185, 0.215]
	γ_2	0.527	0.073	[0.384, 0.671]
MidTemp	β_{30}	0.047	0.082	[-0.113, 0.207]
	β_{31}	-0.151	0.011	[-0.173, -0.128]
	β_{32}	-0.192	0.049	[-0.288, -0.097]
	β_{33}	-0.079	0.108	[-0.290, 0.133]
	σ_3^2	0.037	0.001	[0.035, 0.040]
	γ_3	-0.132	0.077	[-0.283, 0.019]
	η_{age}	0.044	0.098	[-0.149, 0.236]
	η_{APOE}	-0.081	0.151	[-0.378, 0.216]

model fit we infer from the values of $\boldsymbol{\beta}$ that both ADAS13 and FAQ increase over time indicating a study-wide decline in cognition. Presence of the APOE allele does not appear to significantly alter the longitudinal trajectory of any of the biomarkers considered here, nor the global hazard after attenuation for the current values of the three biomarkers. The values for $\boldsymbol{\gamma}$ corroborate with the profiles shown in Fig. 3: For ADAS13 and FAQ scores, a score above study average results in increased risk of Alzheimer's conversion, whilst the same is true for lower scores in MidTemp volume, albeit in a slightly less significant manner.

5. Discussion & future work

We have presented the successful implementation of the approximate EM algorithm put forward by Bernhardt et al. (2015) – originally performed on a logistic sub-model – on a survival one and demonstrated its scaleability for up to ten biomarkers. The method performs well when the longitudinal profiles have a reasonable length, as one might expect given the underlying approximate nature of the approach. Additionally we have presented the non-exponential growth in computation time as the number of random effects grows large. As such, the approach could be well suited to electronic healthcare data, which usually spans many years with many longitudinal biomarkers available so this could be the perfect setting for this approach.

Throughout all simulations presented in Section 3 and in the application in Section 4.2, we adopted an intercept-and-slope random effects structure as well as utilising a mixture of this structure along with quadratic random effects in Section 4.1 couched in a multivariate joint model setting. However, the method may have broader use in a univariate

setting, where we note that existing methods for univariate joint models perform very well for a relatively small number of random effects (Furgal et al., 2019), but may suffer computational burden when more complex structures are used. We postulate that the approximate EM algorithm presented here should work for any reasonable random effect structure: for example inclusion of higher order terms, stationary Gaussian processes or spline terms. With this in mind, further work could include extending the method to operate under more complex random effects structures which are popular in the literature but have yet to find their way into routine use owing to computational complexity. Inclusion of spline terms could additionally enhance performance capabilities when only limited longitudinal profiles are available.

Due to the clinical importance of the survival of a specific disease, the patient's survival is predominantly the endpoint of interest in joint modelling, where it is usually jointly modelled with a continuous longitudinal response, or multiple responses, as we have solely considered here. However we note the emergence of joint models containing a mixture of generalised linear mixed models (GLMMs), where we believe the approximate EM presented here is general enough to be applied to these emerging joint models.

For instance, a continuous (i.e. Gaussian) sub-model for the longitudinal response(s) may be inadequate given the nature of a specific biomarker. For instance Zhu et al. (2018) utilise a longitudinal (zero-inflated) Poisson model for daily cigarette count and a Cox PH model for time to study dropout. Indeed, in our application to the ADNI data in Section 4.2, the biomarkers ADAS-13 and FAQ exist on a bounded scale, and so could be modelled as counts under this generalised mixed model framework.

Beyond Poisson sub-models, one may also use ordinal specifications of the longitudinal response(s). Examples of this include He and Luo (2016), where a multilevel model is used for multiple longitudinal outcomes and a Cox PH model for the event time and Li et al. (2010), as well as Alam et al. (2021), whose joint model includes a partial proportional odds model for the longitudinal response (in a multivariate setting in the latter). A much more comprehensive review of joint models which utilise a GLMM sub-model is given in Hickey et al. (2016).

We could additionally consider alternatives and extensions with respect to the assumed univariate event time. For example, there are instances where it is of importance to identify the status of a specific event, in which case the endpoint of interest is a binary variable. The joint model subsequently contains a continuous longitudinal response – modelled as per (1) – and a logistic model for the binary outcome, in place of the proportional hazards model (2) with association induced through the linear predictor. Hwang et al. (2011) introduce such a model for orthostatic hypotension with a single longitudinal response, and Bernhardt et al. (2015) employ a multivariate joint model with three continuous longitudinal responses and a logistic sub-model for survival past a certain period of follow-up.

Furthermore, a joint model with a competing risks survival sub-model (Williamson et al., 2008; Li et al., 2010) would be more useful in circumstances where patients can experience multiple events of interest: For instance death or disease recurrence; recurrent events such as re-admission to hospital; or a succession of events such as transition between (e.g. worsening of) disease states (Hickey et al., 2018b). Additionally, joint frailty-copula models could be explored (Emura et al., 2017; Peng et al., 2018; Sofeu et al., 2021), as the likelihood here also involves numerical integration with respect to the random effects.

Acknowledgements

This research was funded by an Engineering and Physical Sciences Research Council (ESPRC) Doctoral Training Partnership grant awarded to JM (grant reference EP/V520184/1). The funder had no role in writing the manuscript. The authors would like to thank the reviewers for their helpful comments, which invariably improved the manuscript.

Appendix A. Likelihood

The observed data likelihood under the joint modelling framework is presented in (3). It is generally easier to work with the log-likelihood. The complete data log-likelihood for subject i is

$$\ell = \log f(\boldsymbol{Y}_i|\boldsymbol{b}_i;\boldsymbol{\beta},\sigma^2_{\varepsilon_1},\ldots,\sigma^2_{\varepsilon_K}) + \log f(\boldsymbol{b}_i|D) + \log f(T_i,\Delta_i|\boldsymbol{b}_i;\boldsymbol{\eta},\boldsymbol{\gamma}), \tag{A.1}$$

where

$$\begin{aligned}
\log f(\boldsymbol{Y}_i|\cdot) &= -\frac{m_i}{2}\log 2\pi - \frac{1}{2}\log|V_i| - \frac{1}{2}\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\boldsymbol{b}_i\right)^T V_i^{-1}\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\boldsymbol{b}_i\right),\\
\log f(\boldsymbol{b}_i|\cdot) &= -\frac{q}{2}\log 2\pi - \frac{1}{2}\log|D| - \frac{1}{2}\boldsymbol{b}_i^T D^{-1}\boldsymbol{b}_i,\\
\log f(T_i,\Delta_i|\cdot) &= \Delta_i \log\lambda_0(T_i) + \Delta_i\left[\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^{K}\gamma_k(1,T_i)^T\boldsymbol{b}_{ik}\right] -\\
&\qquad \int_0^{T_i}\lambda_0(u)\exp\left\{\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^{K}\gamma_k(1,u)^T\boldsymbol{b}_{ik}\right\}du.
\end{aligned} \tag{A.2}$$

In the above, $m_i = \sum_{k=1}^{K} m_{ik}$ denotes the total number of longitudinal observations for subject i and q the dimensionality of random effects (i.e. $q = 6$ for three longitudinal responses each modelled with random intercept and slope).

Appendix B. M-step details

This section presents the score equations as well as parameter updates for each component of $\boldsymbol{\Omega}$, as well as the hazard, which we update at each EM iteration, but do not monitor for convergence of the algorithm.

B.1. Update for D

The update for the covariance matrix of the random effects D does not require any integration. The expectation on the relevant part of the log-likelihood from (A.1) is

$$\begin{aligned} E_i\left[\log f(\boldsymbol{b}_i|D)\right] &= E_i\left[-\frac{q}{2}\log 2\pi - \frac{1}{2}\log|D| - \frac{1}{2}\boldsymbol{b}_i^T D^{-1}\boldsymbol{b}_i\right] \\ &= -\frac{q}{2}\log 2\pi + \frac{1}{2}\log|D^{-1}| - \frac{1}{2}\mathrm{Tr}\left\{D^{-1}E_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^T\right]\right\}, \end{aligned}$$

with score equation for D^{-1}

$$\frac{\partial E_i\left[\log f(\boldsymbol{b}_i|D)\right]}{\partial D^{-1}} = -\frac{1}{2}D - \frac{1}{2}E_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^T\right] \implies \hat{D} = E_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^T\right].$$

Using the normal approximation (5) to evaluate $E_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^T\right]$

$$\mathrm{Var}\left[\boldsymbol{b}_i\right] = \hat{\Sigma}_i = E_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^T\right] - E_i\left[\boldsymbol{b}_i\right]E_i\left[\boldsymbol{b}_i\right]^T \implies E_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^T\right] = \hat{\Sigma}_i + \hat{\boldsymbol{b}}_i\hat{\boldsymbol{b}}_i^T.$$

Thus the parameter update for D is

$$\hat{D} = \sum_{i=1}^{n}\hat{\Sigma}_i + \hat{\boldsymbol{b}}_i\hat{\boldsymbol{b}}_i^T/n.$$

B.2. Update for β

The coefficients for the fixed effects for covariates in the longitudinal sub-model, $\boldsymbol{\beta}$, appears in $\log f(\boldsymbol{Y}_i|\cdot)$, the expectation of which is then

$$\begin{aligned} E_i\left[\log f(\boldsymbol{Y}_i|\cdot)\right] &\underset{\boldsymbol{\beta}}{\propto} -\frac{1}{2}E_i\left[\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\boldsymbol{b}_i\right)^T V_i^{-1}\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\boldsymbol{b}_i\right)\right] \\ &= -\frac{1}{2}\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - E_i[Z_i\boldsymbol{b}_i]\right)^T V_i^{-1}\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - E_i[Z_i\boldsymbol{b}_i]\right). \end{aligned}$$

Then, making use of normal approximation (5),

$$Z_i\boldsymbol{b}_i \overset{\text{appx.}}{\sim} N(Z_i\hat{\boldsymbol{b}}_i, Z_i\hat{\Sigma}_i Z_i^T) = N(\mu_i, \tau_i^2),$$

we then obtain

$$\tilde{E}_i\left[\ell_{\boldsymbol{\beta}}\right] = -\frac{1}{2}\sum_{\ell=1}^{\rho} w_\ell\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\hat{\boldsymbol{b}}_i - \tau_i v_\ell\right)^T V_i^{-1}\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\hat{\boldsymbol{b}}_i - \tau_i v_\ell\right),$$

where w_ℓ and v_ℓ, $\ell = 1, \ldots, \rho$ are the quadrature weights and abscissae respectively obtained from the `gauss.quad.prob` function from the `R` library `statmod` (Smyth, 2005). We can then form the score, and subsequently parameter update, for $\boldsymbol{\beta}$.

$$\frac{\partial\tilde{E}_i[\ell_{\boldsymbol{\beta}}]}{\partial\boldsymbol{\beta}} = X_i^T V_i^{-1}\sum_{\ell=1}^{\rho} w_\ell\left(\boldsymbol{Y}_i - X_i\boldsymbol{\beta} - Z_i\hat{\boldsymbol{b}}_i - \tau_i v_\ell\right) = 0$$

$$X_i\boldsymbol{\beta}\left(X_i^T V_i^{-1}\right) = X_i^T V_i^{-1}\sum_{\ell=1}^{\rho} w_\ell\left(\boldsymbol{Y}_i - Z_i\hat{\boldsymbol{b}}_i - \tau_i v_\ell\right)$$

$$X_i^T X_i \boldsymbol{\beta} = X_i^T \left(\boldsymbol{Y}_i - Z_i \hat{\boldsymbol{b}}_i - \sum_{\ell=1}^{\rho} w_\ell \tau_i v_\ell \right),$$
$$\implies \hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} X_i^T X_i \right)^{-1} \sum_{i=1}^{n} X_i^T \left(\boldsymbol{Y}_i - Z_i \hat{\boldsymbol{b}}_i - \sum_{\ell=1}^{\rho} w_\ell \tau_i v_\ell \right).$$

B.3. Update for $\sigma^2_{\varepsilon_k}$

As with the update for $\boldsymbol{\beta}$, the variance terms only appear in $\log f(\boldsymbol{Y}_i|\cdot)$. We make use of the fact that $V_i = \sigma^2_{\varepsilon_k} \mathbb{I}_{m_{ik}}$ and formulate the expectation

$$\begin{aligned} E_i\left[\log f(\boldsymbol{Y}_{ik}|\cdot)\right] &\underset{\sigma^2_{\varepsilon_k}}{\propto} E_i\left[-\frac{1}{2} m_{ik} \log \sigma^2_{\varepsilon_k} - \frac{1}{\sigma^2_{\varepsilon_k}} \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\boldsymbol{b}_{ik}\right)^T \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\boldsymbol{b}_{ik}\right) \right] \\ &= -\frac{1}{2} m_{ik} \log \sigma^2_{\varepsilon_k} - \frac{1}{\sigma^2_{\varepsilon_k}} \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - E_i[Z_{ik}\boldsymbol{b}_{ik}]\right)^T \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - E_i[Z_{ik}\boldsymbol{b}_{ik}]\right). \end{aligned}$$

Next, we make use of the normal approximation (5) (and, essentially repeat same formulation in the update for $\boldsymbol{\beta}$)

$$Z_{ik}\boldsymbol{b}_{ik} \overset{\text{appx.}}{\sim} N(Z_{ik}\hat{\boldsymbol{b}}_{ik}, Z_{ik}\hat{\Sigma}_{ik}Z_{ik}^T) = N(\mu_{ik}, \tau^2_{ik}).$$

Thus, we obtain

$$\begin{aligned} \tilde{E}_i[\ell_{\sigma^2_{\varepsilon_k}}] = &-\frac{1}{2} m_{ik} \log \sigma^2_{\varepsilon_k} \\ &- \frac{1}{\sigma^2_{\varepsilon_k}} \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\hat{\boldsymbol{b}}_{ik} - \sum_{\ell=1}^{\rho} w_\ell \tau_{ik} v_\ell \right)^T \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\hat{\boldsymbol{b}}_{ik} - \sum_{\ell=1}^{\rho} w_\ell \tau_{ik} v_\ell \right) \end{aligned}$$

And formulate the score equation for $\sigma^2_{\varepsilon_k}$ and subsequently its update, denoting $\mathcal{A} = \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\hat{\boldsymbol{b}}_{ik} - \sum_{\ell=1}^{\rho} w_\ell \tau_{ik} v_\ell \right)$ for brevity.

$$\begin{aligned} \frac{\partial \tilde{E}_i[\ell_{\sigma^2_{\varepsilon_k}}]}{\partial \sigma^2_{\varepsilon_k}} &= -\frac{m_{ik}}{2\sigma^2_{\varepsilon_k}} + \frac{1}{2\sigma^4_{\varepsilon_k}} \mathcal{A}^T\mathcal{A} = 0 \\ \frac{m_{ik}}{2\sigma^2_{\varepsilon_k}} &= \frac{1}{2\sigma^4_{\varepsilon_k}} \mathcal{A}^T\mathcal{A}, \end{aligned}$$

giving

$$\hat{\sigma}^2_{\varepsilon_k} = \sum_{i=1}^{n} \sum_{\ell=1}^{\rho} w_\ell \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\hat{\boldsymbol{b}}_{ik} - \tau_{ik} v_\ell \right)^T \left(\boldsymbol{Y}_{ik} - X_{ik}\boldsymbol{\beta}_k - Z_{ik}\hat{\boldsymbol{b}}_{ik} - \tau_{ik} v_\ell \right) / \sum_{i=1}^{n} m_{ik}.$$

B.4. Update for $\lambda_0(u)$

Before considering the form of the update to the baseline hazard – and indeed all parameters associated with the time-to-event sub-model – we rewrite the log-likelihood for $\log f(T_i, \Delta_i|\cdot)$ given in Appendix A as

$$\log f(T_i, \Delta_i|\cdot) = \Delta_i \log \lambda_0(T_i) + \Delta_i \left[\boldsymbol{K}_i^T \boldsymbol{\eta} + \sum_{k=1}^{K} \gamma_k F_i \boldsymbol{b}_{ik}^T \right] - \lambda_0(\boldsymbol{u}_i) \exp\left\{ \boldsymbol{K}_i^T \boldsymbol{\eta} + \sum_{k=1}^{K} \gamma_k F_{u_i} \boldsymbol{b}_{ik} \right\}, \tag{B.1}$$

where we introduce $F_i = (1, T_i)$, $\boldsymbol{u}_i$ is a vector of failure times survived by subject i, up to and including the subject's failure time, and F_{u_i} the design matrix on these failure times.
We now consider the expectation on the part of $\log f(T_i, \Delta_i|\cdot)$ containing the baseline hazard

$$\begin{aligned} E_i\left[\log f(T_i, \Delta_i|\cdot)\right] &\underset{\lambda_0(\cdot)}{\propto} E_i\left[\Delta_i \log \lambda_0(T_i) - \lambda_0(\boldsymbol{u}_i) \exp\left\{ \boldsymbol{K}_i^T \boldsymbol{\eta} + \sum_{k=1}^{K} \gamma_k F_{u_i} \boldsymbol{b}_{ik} \right\} \right] \\ &= \Delta_i \log \lambda_0(T_i) - \lambda_0(\boldsymbol{u}_i) E_i\left[\exp\left\{ \boldsymbol{K}_i^T \boldsymbol{\eta} + \sum_{k=1}^{K} \gamma_k F_{u_i} \boldsymbol{b}_{ik} \right\} \right]. \end{aligned}$$

We make use of normal approximation (5)

$$\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^{K}\gamma_k F_{u_i}\boldsymbol{b}_{ik} \overset{\text{appx.}}{\sim} N\left(\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^{K}\gamma_k F_{u_i}\hat{\boldsymbol{b}}_{ik}, \sum_{k=1}^{K}\gamma_k^2 F_{u_i}\hat{\Sigma}_{ik}F_{u_i}^T\right) = N(\mu_i, \tau_i^2). \tag{B.2}$$

Thus,

$$\tilde{E}_i[\ell_{\lambda_0}] = \Delta_i \log\lambda_0(T_i) - \lambda_0(\boldsymbol{u}_i)\sum_{\ell=1}^{\rho} w_\ell \exp\{\mu_i + \tau_i v_\ell\}.$$

Taking score equations, we can quite trivially form the update for $\lambda_0(u)$

$$\hat{\lambda}_0(u) = \frac{\sum_{i=1}^{n}\Delta_i I(T_i = u)}{\sum_{i=1}^{n}\sum_{\ell=1}^{\rho} w_\ell \exp\{\mu_i + \tau_i v_\ell\} I(T_i \geq u)},$$

where $I(\cdot)$ is the indicator function.

B.5. Update for $(\boldsymbol{\gamma}, \boldsymbol{\eta})$

Both the fixed effects for the covariates in the survival sub-model, $\boldsymbol{\eta}$ and the association parameters, $\boldsymbol{\gamma}$ appear only in the survival sub-model, the form of which is given in (B.1). We know from the previous section

$$\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}] = \Delta_i\left[\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^{K}\gamma_k F_i\boldsymbol{b}_{ik}\right] - \sum_{\ell=1}^{\rho} w_\ell\lambda_0(\boldsymbol{u}_i)\exp\{\mu_i + \tau_i v_\ell\},$$

wherein μ_i and τ_i are given by (B.2). The presence of exponents indicates the update will not be in closed form. We therefore undertake a one-step Newton Raphson iteration to update $\Gamma = (\boldsymbol{\gamma}, \boldsymbol{\eta})$ at each M-step. The update is of the form

$$\hat{\Gamma} = \Gamma + S(\Gamma)/I(\Gamma),$$

where $S(\gamma)$ is a vector of length $K + P_s$ and $I(\Gamma)$ a symmetric square information matrix of dimension $(K + P_s) \times (K + P_s)$, with P_s denoting the number of covariates in the survival model to estimate.
Starting with the score $S(\Gamma)$, we calculate

$$\frac{\partial\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]}{\partial\gamma_k} = \Delta_i F_i^T\hat{\boldsymbol{b}}_{ik} - \sum_{\ell=1}^{\rho} w_\ell\xi^T F_{u_i}\hat{\boldsymbol{b}}_{ik} + \gamma_k w_\ell v_\ell\left[\xi\odot\tau_i\right]^T\xi,$$

$$\frac{\partial\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]}{\partial\boldsymbol{\eta}} = \Delta_i\boldsymbol{K}_i - \boldsymbol{K}_i^T\sum_{\ell=1}^{\rho} w_\ell\xi,$$

where $\odot$ denotes element-wise multiplication and we define
$\xi = \lambda_0(\boldsymbol{u}_i)\odot\exp\left\{\boldsymbol{K}_i^T\boldsymbol{\eta} + \sum_{k=1}^{K}\gamma_k F_{u_i}\boldsymbol{b}_{ik} + v_\ell\tau_i\right\}$.
Thus, we form the score vector $S(\Gamma) = \left(\frac{\partial\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]}{\partial\gamma_1}, \ldots, \frac{\partial\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]}{\partial\gamma_K}, \frac{\partial\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]}{\partial\boldsymbol{\eta}}\right)$.

We wish to populate the information matrix where we let $f = \tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]$.

$$I(\Gamma) = -\begin{bmatrix} \frac{\partial^2 f}{\partial\gamma_1^2} & \cdots & \frac{\partial^2 f}{\partial\gamma_1\gamma_K} & \frac{\partial^2 f}{\partial\gamma_1\boldsymbol{\eta}} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial\gamma_K\gamma_1} & \cdots & \frac{\partial^2 f}{\partial\gamma_K^2} & \frac{\partial^2 f}{\partial\gamma_K\boldsymbol{\eta}} \\ \frac{\partial^2 f}{\partial\gamma_1\boldsymbol{\eta}}^T & \cdots & \frac{\partial^2 f}{\partial\gamma_K\boldsymbol{\eta}}^T & \frac{\partial^2 f}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T} \end{bmatrix},$$

where we calculate the following second derivatives

$$-\frac{\partial f}{\partial\gamma_k^2} = \sum_{\ell=1}^{\rho} w_\ell\hat{\boldsymbol{b}}_{ik}{}^T F_{u_i}^T\left[\xi\odot F_{u_i}\hat{\boldsymbol{b}}_{ik}\right] + \gamma_k v_\ell w_\ell\hat{\boldsymbol{b}}_{ik}{}^T F_{u_i}^T\left[\xi\odot\tau_i^*\odot\tilde{\tau}_{ik}\right] + v_\ell w_\ell\left[\xi\odot\tau_i\right]^T\xi +$$
$$2\gamma_k v_\ell w_\ell\left[\xi\odot\tau_i\odot\xi\right]^T F_{u_i}\hat{\boldsymbol{b}}_{ik} + 2w_\ell\gamma_k^2 v_\ell^2\left[\xi\odot\tau_i\odot\xi\odot\tau_i^*\right]^T\tilde{\tau}_{ik} + \gamma_k^2 v_\ell w_\ell\left[\xi\odot\xi\tau_i^2\right]^T\tilde{\tau}_{ik};$$

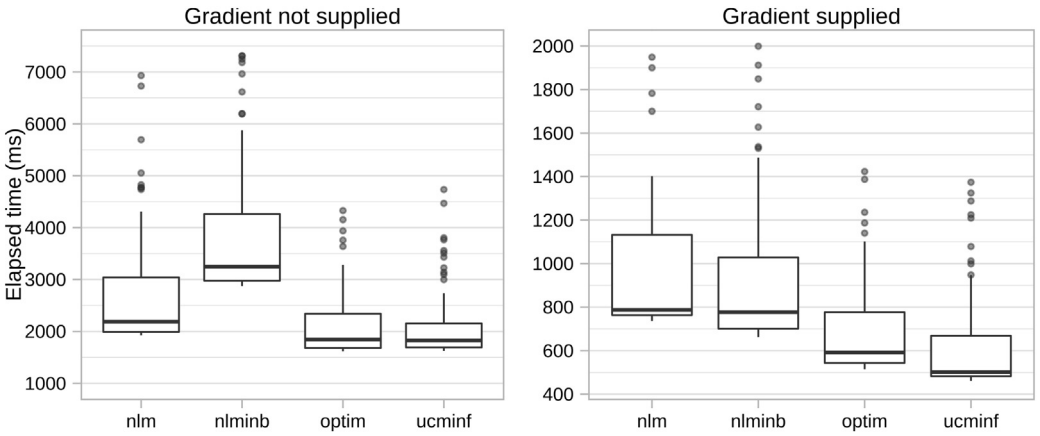

Fig. C.1. Benchmarks for one hundred maximisations of the complete data log-likelihood for simulated sample of size $n = 250$. `optim` was fit using the 'BFGS' method.

$$
\begin{aligned}
-\frac{\partial f}{\partial \gamma_p \partial \gamma_q} &= \sum_{\ell=1}^{\rho} w_\ell \hat{\boldsymbol{b}}_{ip}{}^T F_{u_i} \left[\xi \odot F_{u_i} \hat{\boldsymbol{b}}_{iq}\right] + \gamma_q v_\ell w_\ell \hat{\boldsymbol{b}}_{ip}{}^T F_{u_i}^T \left[\xi \odot \tau_i^* \odot \tau_{iq}\right] + 2\gamma_p v_\ell w_\ell \left[\xi \odot \tau_i \odot \xi\right]^T F_{u_i} \hat{\boldsymbol{b}}_{iq} + \\
&\quad 2\gamma_p \gamma_q w_\ell v_\ell^2 \left[\xi \odot \tau_i \odot \xi \odot \tau_i^*\right]^T \tau_{iq} + \gamma_p \gamma_q v_\ell w_\ell \left[\xi \odot \xi \odot \tau_i^*\right]^T \tau_{iq}; \\
-\frac{\partial^2 f}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} &= \sum_{\ell=1}^{\rho} w_\ell \left[\text{diag}(\xi) \boldsymbol{K}_i\right]^T \boldsymbol{K}_i; \\
-\frac{\partial f}{\partial \gamma_k \partial \boldsymbol{\eta}} &= \sum_{\ell=1}^{\rho} w_\ell \boldsymbol{K}_i^T \left[\left(F_{u_i} \hat{\boldsymbol{b}}_{ik}\right) \odot \xi\right] + 2\gamma_k v_\ell w_\ell \boldsymbol{K}_i^T \left[\xi \odot \tau_i \odot \xi\right].
\end{aligned}
$$

Here we have additionally included $\tilde{\tau}_{ik} = \text{diag}(F_{u_i}^T \hat{\Sigma}_{ik} F_{u_i})$ as well as $\tau_i^* = \left(\sum_{k=1}^K \gamma_k^2 F_{u_i}^T \hat{\Sigma}_{ik} F_{u_i}\right)^{-\frac{1}{2}}$. The above analytical implementation can be used, or numerical methods on $\tilde{E}_i[\ell_{(\boldsymbol{\gamma},\boldsymbol{\eta})}]$ to obtain the score vector and subsequently the Hessian matrix.

Appendix C. Choice of optimiser

One has a choice between a number of optimisation functions when using `R`. Included with base `R` are the functions `optim`, `nlm` and `nlminb`. A plethora of pre-existing packages are available which perform non-linear optimisation required in the E-step exist for use with `R`.[2] One such package is `ucminf` (Nielsen and Mortensen, 2016) and was considered due to syntactic similarities with `optim`. Fig. C.1 shows these four optimising functions benchmarked against one another for maximising $\log f(\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{b}_i; \boldsymbol{\Omega})$ w.r.t $\boldsymbol{b}_i \ \forall \ i = 1, \ldots, 250$ from some simulated trivariate data (i.e. dimensionality of random effects $q = 6$) one hundred times. We used `R` package `microbenchmark` (Mersmann, 2019) for this benchmarking. One can additionally supply these optimisers with a gradient function, $\frac{\partial \log f(\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{b}_i; \boldsymbol{\Omega})}{\partial \boldsymbol{b}_i}$. Irregardless of the optimiser, it is clear – and unsurprising – that providing the optimiser with a gradient function greatly improves computational efficiency under default control parameters. Furthermore, one can alter convergence criteria for the optimisers which improves efficiency further. For all model fits and simulations in the main paper, a relative difference of 0.001 was used in this maximisation step using `ucminf`

Appendix D. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2022.107438.

References

Alam, K., Maity, A., Sinha, S.K., Rizopoulos, D., Sattar, A., 2021. Joint modeling of longitudinal continuous, longitudinal ordinal, and time-to-event outcomes. Lifetime Data Anal. 27, 64–90. https://doi.org/10.1007/s10985-020-09511-3.

[2] https://cran.r-project.org/web/views/Optimization.html.

Andrinopoulou, E.R., Rizopoulos, D., 2016. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. Stat. Med. 35, 4813–4823. https://doi.org/10.1002/sim.7027.
Baghishani, H., Mohammadzadeh, M., 2012. Asymptotic normality of posterior distributions for generalized linear mixed models. J. Multivar. Anal. 111, 66–77. https://doi.org/10.1016/j.jmva.2012.05.003.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.
Bernhardt, P.W., Zhang, D., Wang, H.J., 2015. A fast em algorithm for fitting joint models of a binary response to multiple longitudinal covariates subject to detection limits. Comput. Stat. Data Anal. 85, 37–53. https://doi.org/10.1016/j.csda.2014.11.011.
Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc., Ser. B, Methodol. 34, 187–220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.
Dai, H., Pan, J., 2018. Joint modelling of survival and longitudinal data with informative observation times: joint modelling. Scand. J. Stat. 45. https://doi.org/10.1111/sjos.12314.
Dil, E., Karasoy, D., 2020. Joint modeling of a longitudinal measurement and parametric survival data with application to primary biliary cirrhosis study. Pak. J. Stat. Oper. Res. 16, 295–304. https://doi.org/10.18187/pjsor.v16i2.3131.
Emura, T., Nakatochi, M., Murotani, K., Rondeau, V., 2017. A joint frailty-copula model between tumour progression and death for meta-analysis. Stat. Methods Med. Res. 26, 2649–2666. https://doi.org/10.1177/0962280215604510.
Furgal, A.K., Sen, A., Taylor, J.M., 2019. Review and comparison of computational approaches for joint longitudinal and time-to-event models. Int. Stat. Rev. 87, 393–418. https://doi.org/10.1111/insr.12322.
He, B., Luo, S., 2016. Joint modeling of multivariate longitudinal measurements and survival data with applications to parkinson's disease. Stat. Methods Med. Res. 25, 1346–1358. https://doi.org/10.1177/0962280213480877.
Henderson, R., Diggle, P., Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. Biostatistics 4, 465–480. https://doi.org/10.1093/biostatistics/1.4.465.
Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., 2016. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med. Res. Methodol. 117. https://doi.org/10.1186/s12874-016-0212-5.
Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., 2018a. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. BMC Med. Res. Methodol. 50. https://doi.org/10.1186/s12874-018-0502-1.
Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., 2018b. Joint models of longitudinal and time-to-event data with more than one event time outcome: a review. Int. J. Biostat. 14, 20170047. https://doi.org/10.1515/ijb-2017-0047.
Hwang, Y.T., Tsai, H.Y., Chang, Y.J., Kuo, H.C., Wang, C.C., 2011. The joint model of the logistic model and linear random effect model — an application to predict orthostatic hypertension for subacute stroke patients. Comput. Stat. Data Anal. 55, 914–923. https://doi.org/10.1016/j.csda.2010.07.024.
Ibrahim, J.G., Chu, H., Chen, L.M., 2010. Basic concepts and methods for joint models of longitudinal and survival data. J. Clin. Oncol. 28, 2796–2801. https://doi.org/10.1200/JCO.2009.25.0654.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974. https://doi.org/10.2307/2529876.
Li, K., Chan, W., Doody, R.S., Quinn, J., Luo, S., 2017. Prediction of conversion to alzheimer's disease with longitudinal measures and time-to-event data. J. Alzheimer's Dis. 58, 361–371. https://doi.org/10.3233/JAD-161201.
Li, K., Luo, S., 2017. Functional joint model for longitudinal and time-to-event data: an application to alzheimer's disease. Stat. Med. 36, 3560–3572. https://doi.org/10.1002/sim.7381.
Li, N., Elashoff, R.M., Li, G., Saver, J., 2010. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the ninds rt-pa stroke trial. Stat. Med. 29, 546–557. https://doi.org/10.1002/sim.3798.
Lin, H., McCulloch, C.E., Mayne, S.T., 2002. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. Stat. Med. 21, 2369–2382. https://doi.org/10.1002/sim.1179.
McCrink, L.M., Marshall, A.H., Cairns, K.J., 2013. Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. Int. Stat. Rev. 81, 249–269. https://doi.org/10.1111/insr.12018.
McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions, 2nd ed. Wiley-Interscience.
Mersmann, O., 2019. microbenchmark: accurate timing functions. https://CRAN.R-project.org/package=microbenchmark. r package version 1.4-7.
Murtaugh, P.A., Dickson, E.R., Van Dam, G.M., Malinchoc, M., Grambsch, P.M., Langworthy, A.L., Gips, C.H., 1994. Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. Hepatology 20, 126–134. https://doi.org/10.1002/hep.1840200120.
Nielsen, H.B., Mortensen, S.B., 2016. ucminf: general-purpose unconstrained non-linear optimization. https://CRAN.R-project.org/package=ucminf. r package version 1.1-4.
Peng, M., Xiang, L., Wang, S., 2018. Semiparametric regression analysis of clustered survival data with semi-competing risks. Comput. Stat. Data Anal. 124, 53–70. https://doi.org/10.1016/j.csda.2018.02.003.
Philipson, P., Hickey, G.L., Crowther, M.J., Kolamunnage-Dona, R., 2020. Faster monte carlo estimation of joint models for time-to-event and multivariate longitudinal data. Comput. Stat. Data Anal. 151, 107–110. https://doi.org/10.1016/j.csda.2020.107010.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2021. nlme: linear and nonlinear mixed effects models. https://CRAN.R-project.org/package=nlme. r package version 3.1-152.
Rizopoulos, D., 2012. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. Comput. Stat. Data Anal. 56, 491–501. https://doi.org/10.1016/j.csda.2011.09.007.
Rizopoulos, D., 2016. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using mcmc. J. Stat. Softw. 72, 1–45. https://doi.org/10.18637/jss.v072.i07.
Rizopoulos, D., Verbeke, G., Lesaffre, E., 2009. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. J. R. Stat. Soc., Ser. B, Stat. Methodol. 71, 637–654. https://doi.org/10.1111/j.1467-9868.2008.00704.x.
Smyth, G.K., 2005. Numerical integration. In: Encyclopedia of Biostatistics, pp. 3088–3095.
Sofeu, C.L., Emura, T., Rondeau, V., 2021. A joint frailty-copula model for meta-analytic validation of failure time surrogate endpoints in clinical trials. Biom. J. 63, 423–446. https://doi.org/10.1002/bimj.201900306.
Song, X., Davidian, M., Tsiatis, A.A., 2002. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3, 511–528. https://doi.org/10.1093/biostatistics/3.4.511.
Therneau, T.M., 2015. A package for survival analysis in S. https://CRAN.R-project.org/package=survival. version 2.38.
Wang, Y., Ibrahim, J.G., Zhu, H., 2020. Partial least squares for functional joint models with applications to the alzheimer's disease neuroimaging initiative study. Biometrics 76, 1109–1119. https://doi.org/10.1111/biom.13219.
Williamson, P.R., Kolamunnage-Done, R., Philipson, P., Marson, A.G., 2008. Joint modeling of longitudinal and competing risks data. Stat. Med. 27, 6426–6438. https://doi.org/10.1002/sim.3451.
Wulfsohn, M.S., Tsiatis, A.A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics 53, 330–339. https://doi.org/10.2307/2533118.
Zhu, H., DeSantis, S.M., Luo, S., 2018. Joint modeling of longitudinal zero-inflated count and time-to-event data: a bayesian perspective. Stat. Methods Med. Res. 27, 1258–1270. https://doi.org/10.1177/0962280216659312.