Computational Statistics and Data Analysis 187 (2023) 107819

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Fast estimation for generalised multivariate joint models using an approximate EM algorithm

James Murray *, Pete Philipson

School of Mathematics, Statistics and Physics, Newcastle University, United Kingdom

ARTICLE INFO

Article history:
Received 23 December 2022
Received in revised form 4 July 2023
Accepted 4 July 2023
Available online 13 July 2023

Keywords:
Generalised linear mixed models
Joint models
Survival analysis
Normal approximation
EM algorithm

ABSTRACT

Joint models for longitudinal and survival data have become an established tool for optimally handling scenarios when both types of data co-exist. Multivariate extensions to the classic univariate joint model have started to emerge but are typically restricted to the Gaussian case, deployed in a Bayesian framework or focused on dimension reduction. An approximate EM algorithm is utilised which circumvents the oft-lamented curse of dimensionality and offers a likelihood-based implementation which ought to appeal to clinicians and practitioners alike. The proposed method is validated in a pair of simulation studies, which demonstrate both its accuracy in parameter estimation and efficiency in terms of computational cost. Its clinical use is demonstrated via an application to primary billiary cirrhosis data. The proposed methodology for estimation of these joint models is available in `R` package `gmvjoint`.

1. Introduction

Joint models under the 'classic' framework of longitudinal outcomes and an event-time process are ubiquitous with classic treatise by Tsiatis and Davidian (2004) and Rizopoulos (2012b). Typically, the longitudinal responses are continuous and assumed to be Gaussian, though this may be for convenience of implementation rather than a true representation of the data. However, there are several scenarios where such assumptions would be unrealistic. For instance, it could be of importance to identify the status of a specific (binary) longitudinal outcome, which can vary over time, with its presence or absence associated with survival. Likewise, integer data (such as questionnaire scores) would likely better suit a count regression (sub-)model. In such cases, there is scant justification for the normality assumption. One could transform/standardise responses of interest such that the usual modelling assumptions are more likely to be satisfied. However, this may come at the cost of model interpretation, perhaps hindering their uptake in clinical use. Such examples have led to the recent emergence of joint models which accommodate longitudinal outcomes of varying type.

In circumstances where accommodation of discrete longitudinal outcomes is sought – such that a linear mixed model (LMM) is not appropriate – the corresponding sub-model is replaced by a suitable member of the exponential family and modelled by a generalised linear mixed model (GLMM).

In a joint modelling context, both Li et al. (2010) and Alam et al. (2021) utilise (partial) proportional odds models for a repeatedly measured response which is scored against a Likert-type scale, and He and Luo (2016) model cumulative probabilities of an ordinal outcome as part of a multivariate joint model. If indication of presence or absence of some response

* 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.2023.107819

during follow-up is of clinical importance, the sub-model for this longitudinal binary repeated measure is modelled by a logistic GLMM (Choi et al., 2015; Rustand et al., 2022b). Sunethra and Sooriyarachchi (2018, 2021) discuss joint modelling with a Poisson longitudinal process, and Zhu et al. (2018) utilise a zero-inflated Poisson (and generalised Poisson) GLMM in modelling daily cigarette count along with time to study dropout.

In literature, joint models with more than one longitudinal response in this aforementioned meta-model were largely restricted to methodological advances (Lin et al., 2002) in the first instance, with recent software developments over the past five years allowing for routine fitting of multivariate joint models (MVJM) with multiple longitudinal responses. These include likelihood-based inference facilitated by Monte Carlo EM (MCEM) in `R` package `joineRML` (Hickey et al., 2018); Markov Chain Monte Carlo (MCMC) methods in `JMbayes2` (Rizopoulos et al., 2021) and approximate Bayesian inference in `INLAjoint` (Rustand et al., 2022a). MVJMs are superior to multiple univariate fits as they take into account correlations between longitudinal responses of interest, thus obtaining correctly adjusted estimates for each response. Additionally, one obtains a single prediction using all possible information, rather than several from said multiple univariate fits, each of which is likely to overstate the importance of any association when treated in isolation.

When the longitudinal processes are assumed to come from different members of the exponential family – such that the multivariate longitudinal process is constructed by a mixture of GLMM types, i.e. continuous and binary – it is important to be able to facilitate joint models where each marker is modelled in its most natural way, rather than assuming normality for computational convenience. Even in light of software development in recent years, a common approach is to use dimension reduction techniques, such as functional principal components regression (Li and Luo, 2017; Li et al., 2021). Outside of this dimension reduction approach, MVJMs wholly constructed by LMMs are fit both by MCMC in Andrinopoulou et al. (2020) and Long and Mills (2018), and via maximum likelihood in Hickey et al. (2018) and Philipson et al. (2020). A multivariate GLMM specification is considered in both Rustand et al. (2022b) and Andrinopoulou and Rizopoulos (2016), with parameter estimation being performed using approximate Bayesian inference in the former, and by a fully Bayesian approach, with shrinkage priors, in the latter. Indeed, approaches within the Bayesian paradigm have become the preeminent method of inference for MVJMs with (at least one) GLMM sub-model.

Rather than seek dimension reduction directly – and remaining within the maximum likelihood framework – Bernhardt et al. (2015) proposed an approximate EM algorithm in the context of a multivariate joint model with a binary outcome in place of the usual survival one. A normal approximation of the distribution of the random effects (for each individual) conditional on the observed data was proposed. This has the effect of reducing the dimensionality of resulting integrals to be uniformally one, regardless of random effect complexity; thereby greatly improving computational efficiency. The authors previously extended this to the more traditional joint model specification of a survival sub-model (Murray and Philipson, 2022). Here, it was shown that the approximation suggested in Bernhardt et al. (2015) facilitated non-exponential increases in computation times as the dimension of random effects increased.

Hitherto, implementation of GLMM sub-models in a multivariate joint model setting has been predicated on the aforementioned Bayesian paradigm – to the best of the authors' knowledge – with no implementation via maximum likelihood, which is closer in spirit to well-established methods for analysing clinical data such as the Cox model and linear mixed effects model for survival and longitudinal data respectively. If MVJMs are to find their way into routine clinical use in a manner akin to Cox PH and LMMs it is important to at least have the option of a non-Bayesian implementation; this work attempts to fill this current void.

The rest of the paper is organised as follows: In Section 2 we establish our models and notation along with the likelihood and parameter estimation strategy. Comprehensive simulation studies are performed in Section 3, where we exhibit performance capabilities of the approximate EM algorithm in multiple simulated scenarios. We then present an application to primary billiary cirrhosis data in Section 4 before concluding with a discussion in Section 5.

2. Methods

2.1. Models and notation

For each subject $i = 1, \ldots, n$ we observe $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}{}^\top, \ldots, \boldsymbol{Y}_{iK}{}^\top)^\top$ where each $\boldsymbol{Y}_{ik}$ denotes the k^{th} longitudinal response vector of interest, for $k = 1, \ldots, K$. The k^{th} longitudinal response vector for subject i, $\boldsymbol{Y}_{ik} = (y_{i1k}, \ldots, y_{im_{ik}k})^\top$, is measured m_{ik} times, which can differ across subjects and 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 is the independent potential censoring time. We additionally introduce failure indicator Δ_i which takes value one if $T_i^* < C_i$ and zero otherwise.

We assume the conditional distribution of the k^{th} response $\boldsymbol{Y}_{ik}$ belongs to a member of the exponential family. We consider a generalised linear mixed model (GLMM) for each $\boldsymbol{Y}_{ik}$ with linear predictor $\boldsymbol{\eta}_{ik}$ i.e.

$$h_k\left(\mathbb{E}\left[\boldsymbol{Y}_{ik}|\boldsymbol{b}_{ik};\boldsymbol{\beta}_k\right]\right) = \boldsymbol{\eta}_{ik} = \mathrm{X}_{ik}(t)\,\boldsymbol{\beta}_k + \mathrm{Z}_{ik}(t)\,\boldsymbol{b}_{ik}, \tag{1}$$

where $h_k(\cdot)$ denotes the known link function imposed on the k^{th} response. Here, X_{ik} denotes the (possibly time-dependent) fixed effects design matrix associated with the k^{th} response for subject i, with corresponding p_k-vector of coefficients $\boldsymbol{\beta}_k$. Likewise, Z_{ik} denotes the (possibly time-dependent) random effects design matrix, and $\boldsymbol{b}_{ik}$ the q_k-vector of subject-specific random effects. These random effects are assumed to be multivariate normal with zero-mean and variance-covariance matrix D_k.

We then form the usual joint model by inducing an association between the expected values of the K responses and the hazard $\lambda_i(t)$ through the random effects $\boldsymbol{b}_{ik}$. The time-to-event sub-model is

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ \boldsymbol{S}_i^\top \boldsymbol{\zeta} + \sum_{k=1}^{K} \gamma_k \boldsymbol{W}_k(t)^\top \boldsymbol{b}_{ik} \right\}. \tag{2}$$

Here, $\boldsymbol{S}_i$ is the p_s-vector of baseline covariates of interest for subject i with corresponding p_s-vector of coefficients $\boldsymbol{\zeta}$ and $\lambda_0(t)$ is an unspecified baseline hazard. Parameter γ_k represents the strength of association between the random effects for the k^{th} longitudinal response in (1) and the hazard, with $\boldsymbol{W}_k(t)$ denoting the appropriate vector function of time corresponding to the random effects structure on the k^{th} longitudinal response in this general case. This could take the form of an intercept and slope, natural cubic splines and so on. Here we note that it is equally (if not more) popular in literature to define the nature of the association in (2) by the current value of the k^{th} linear predictor, rather than only the random effects; see Table 1 in Hickey et al. (2016). We opt for the shared random effects association structure out of preference alone; deviations away from some population mean trajectory being the driving force behind an observed association. Irrespective of the choice for the latent association, more complex association structures could include stationary Gaussian processes (Henderson et al., 2000; Martins, 2022), current-value-and-slope parameterisations (Rizopoulos and Ghosh, 2011; Rustand et al., 2022b) and variations thereof.

For any random effects specification, as the dimensionality of the random effects increases – whether through increased complexity and/or the number of longitudinal responses – we expect commensurate increases in computation time under current estimation approaches.

2.2. Likelihood

Assuming the survival process is conditionally independent of the longitudinal processes given the random effects, we define the observed data likelihood as

$$\prod_{i=1}^{n} \left\{ \int_{-\infty}^{\infty} \left[\prod_{k=1}^{K} f\left(\boldsymbol{Y}_{ik} | \boldsymbol{b}_{ik}; \boldsymbol{\beta}_k, \sigma_k\right) \right] f(T_i, \Delta_i | \boldsymbol{b}_i; \boldsymbol{\gamma}, \boldsymbol{\zeta}) f(\boldsymbol{b}_i | \mathrm{D}) d\boldsymbol{b}_i \right\}, \tag{3}$$

where $f(\boldsymbol{Y}_{ik}|\cdot)$ is an appropriate probability density function (pdf) belonging to the exponential family with dispersion/scale parameter σ_k for the k^{th} longitudinal response $\boldsymbol{Y}_{ik}$. Similarly $f(T_i, \Delta_i|\cdot)$ denotes the pdf for the survival response and $f(\boldsymbol{b}|\cdot)$ denotes that for the random effects. The form of the densities in (3) are provided in Appendix A. We introduce $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^\top, \ldots, \boldsymbol{b}_{iK}^\top)^\top$ as the collection of subject-specific random effects; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_K^\top)^\top$ the collection of fixed effects; $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_K)^\top$ the dispersion parameters and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)^\top$ is the vector of association parameters across the K responses. Additionally, D is the collection of variance-covariance matrices across the K responses, with its block diagonal elements constructed by D_k, and the covariance *between* responses on the off-diagonal. Subsequently, we define the parameter vector $\boldsymbol{\Omega} = \left(\mathrm{vech}(\mathrm{D})^\top, \boldsymbol{\beta}^\top, \boldsymbol{\sigma}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\zeta}^\top\right)^\top$, where $\mathrm{vech}(\cdot)$ denotes the half-vectorisation of its matrix argument, thus returning all unique elements. The baseline hazard $\lambda_0(\cdot)$ is not explicitly a member of $\boldsymbol{\Omega}$ since it is treated as a nuisance parameter.

2.3. Parameter estimation via an approximate EM algorithm

When faced with missing data in the form of unobserved random effects, the Expectation Maximisation (EM, Dempster et al., 1977) 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 taken with respect to the distribution of the unobserved random effects conditional on the observed data – and a series of maximum likelihood parameter updates that form the subsequent M-step. The complete data for subject i is $\{\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{b}_i\}$ where all are observed bar the random effects $\boldsymbol{b}_i$. As such, the complete data log-likelihood which forms the E-step is

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

The expectation $\mathbb{E}_i$ in (4) is taken with respect to the conditional distribution of the random effects on the observed data at a current set of parameter estimates $f(\boldsymbol{b}_i|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}})$. The M-step is formed by maximising the sum of n sets of expectations of the form $\mathbb{E}\left[g(\boldsymbol{b}_i)|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}}\right]$ in the preceding E-step, where $g(\boldsymbol{b}_i)$ denotes some function of the random effects, whose expectation is calculated with respect to the conditional distribution $f(\boldsymbol{b}_i|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}})$.

Bernhardt et al. (2015) proposed approximating the conditional distribution of random effects on the observed data at current parameter estimates to be (multivariate) normal. This then exploits that any linear combination of $\boldsymbol{b}_i$ would also be normal,

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

Here $\hat{\boldsymbol{b}}_i$ is the vector which maximises the complete data log-likelihood $\log f(\boldsymbol{b}_i, \boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}})$ at the current set of parameter estimates, with estimated variance

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

It was previously demonstrated by Rizopoulos (2012a) that $f(\boldsymbol{b}_i|\boldsymbol{Y}_i; \boldsymbol{\Omega})$ is approximately normal as $m_{ik} \to \infty$ within the confines of a classic univariate joint model. Subsequently, Bernhardt et al. (2015) extended this result for the 'full' conditional distribution shown in (5) for a multivariate Gaussian joint model. Away from joint models, Baghishani and Mohammadzadeh (2012) showed that the posterior distribution of random effects for a GLMM of any type are asymptotically normal.

In this work, we adopt the result from Baghishani and Mohammadzadeh (2012) and apply it to a multivariate GLMM joint model, demonstrating that the approximation utilised by Bernhardt et al. (2015) and Murray and Philipson (2022) holds in the more general case when $f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\Omega})$ is not restricted to being normally distributed. Justification for the approximation can be found in Supplementary Material I, where we argue by simulation that this normal approximation appears reasonable, even for lower periods of follow-up. The approximation (5) then allows for all necessary conditional expectations of the form $\mathbb{E}\left[g(\boldsymbol{b}_i)|\boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}}\right]$ to be taken with respect to a univariate normal distribution. These normal distributions are then evaluated in tandem with low-dimensional (adaptive) Gauss-Hermite quadrature (Bernhardt et al., 2015; Murray and Philipson, 2022), with details provided in Appendix B.

2.3.1. Starting values and convergence details

We detail the steps taken to fit the joint model proposed above, and subsequently investigated and presented in Sections 3 and 4, in similar spirit to previous work (Bernhardt et al., 2015; Murray and Philipson, 2022).

1. i. For $k = 1, \ldots, K$ obtain parameter estimates for the k^{th} longitudinal process by fitting a generalised linear mixed model using `glmmTMB` (Brooks et al., 2017). Here, one supplies an appropriate family for the k^{th} response along with fixed and random effects structure. This then returns the fixed effects $\boldsymbol{\beta}_k$, dispersion parameter σ_k (if applicable) and the covariance matrix for the random effects structure imposed, D_k.
   ii. Each of these K univariate fits return best linear unbiased predictors for the random effects for the k^{th} response, $\boldsymbol{b}_{ik}$.[1] We then employ these random effects as time-varying covariates in a Cox PH model (using e.g. `survival`, Therneau 2015) along with baseline covariates $\boldsymbol{S}_i$ which give rise to initial values for the time-to-event parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\zeta}$.
2. Maximise $\log f(\boldsymbol{b}_i, \boldsymbol{Y}_i, T_i, \Delta_i; \hat{\boldsymbol{\Omega}})$ with respect to $\boldsymbol{b}_i$ in order to obtain $\hat{\boldsymbol{b}}_i$ and $\hat{\Sigma}_i$. In practise this is done using optimisation function `optim` using the BFGS algorithm.
3. Use the approximation (5) in order to update parameter vector $\boldsymbol{\Omega}^{(m)} \to \boldsymbol{\Omega}^{(m+1)}$.
4. Cycle between steps 2. and 3. for a minimum of four iterations, then continue until the algorithm converges, which we deem to have occurred when, for the P parameters constructing $\boldsymbol{\Omega}$,

$$\begin{cases} \max\left(|\Omega_1^{(m+1)} - \Omega_1^{(m)}|, \ldots, |\Omega_P^{(m+1)} - \Omega_P^{(m)}|\right) & < \xi_1 \text{ if } |\Omega_x| < \upsilon \\ \max\left(\frac{|\Omega_1^{(m+1)} - \Omega_1^{(m)}|}{|\Omega_1^{(m)}| + \nu}, \ldots, \frac{|\Omega_P^{(m+1)} - \Omega_P^{(m)}|}{|\Omega_P^{(m)}| + \nu}\right) & < \xi_2 \text{ if } |\Omega_x| \geq \upsilon. \\ & \quad {\scriptstyle x=1,\ldots,P} \end{cases} \tag{6}$$

   Here, a two-pronged approach utilising both the absolute and relative difference convergence criteria is employed. If the x^{th} element of $\boldsymbol{\Omega}$ is deemed to be close to zero, the absolute difference criterion is used and otherwise the relative.

We elected to use `glmmTMB` as it allows for efficient generalised mixed model fitting for numerous families, and we found it to provide much more stable parameter estimates in a more timely manner when compared with competing packages. We opt for the convergence criterion shown above as the approach precludes issues one may encounter when considering solely one convergence criterion, whilst still affording accuracy in those parameters with estimates closer to zero; in particular those variance components on the diagonal of D that are bounded below by zero. In the relative difference criterion, ν is some small value added to the denominator to preclude numerical issues which might occur in the calculation here. In all presented simulations and analyses we use $\xi_1 = \nu = 10^{-3}$, $\upsilon = 0.1$; in simulations in Section 3 we employ $\xi_2 = 5 \times 10^{-3}$ and in the application in Section 4 $\xi_2 = 10^{-2}$.

[1] The covariance between $\boldsymbol{b}_e, \boldsymbol{b}_f \ \forall\ e \neq f$ is used to populate the off-block-diagonals in D.

2.3.2. Standard error calculation

Once the algorithm is deemed to have converged, we complete inference by obtaining standard errors for parameter estimates $\hat{\boldsymbol{\Omega}}$, which would allow for calculation of the usual Wald confidence intervals. We approximate the observed empirical information matrix at this set of parameter estimates (McLachlan and Krishnan, 2008; Hickey et al., 2018)

$$\tilde{\mathcal{I}}(\hat{\boldsymbol{\Omega}}) = \left\{ \sum_{i=1}^{n} s_i(\hat{\boldsymbol{\Omega}}) s_i^{\top}(\hat{\boldsymbol{\Omega}}) \right\} - n^{-1} S(\hat{\boldsymbol{\Omega}}) S^{\top}(\hat{\boldsymbol{\Omega}}), \tag{7}$$

where $s_i(\hat{\boldsymbol{\Omega}})$ denotes the gradient vector of the conditional expectation of the complete data (profile) log-likelihood function (4) evaluated at $\hat{\boldsymbol{\Omega}}$ (i.e. the score statistic) and $S(\hat{\boldsymbol{\Omega}}) = \sum_i s_i(\hat{\boldsymbol{\Omega}})$. The right hand side of (7) should equal zero at the MLEs $\hat{\boldsymbol{\Omega}}$ (McLachlan and Krishnan, 2008), however Bernhardt et al. (2015) note that these are not technically MLEs, and so it is included for completeness' sake. Other methods for obtaining the standard errors include bootstrapping, or obtaining the observed information matrix by the methodology outlined in Xu et al. (2014). We prefer the approach given by (7) due to it being comparatively efficient from a computational standpoint.

With methodology established, we undertake extensive simulation studies as well as an application. All simulation studies and applications in the following sections were executed on an Ubuntu Desktop with 4 GHz Intel core i7 with 8 GB RAM using `R` version 4.2.0. No high performance computing facility was used. `R` code used to generate the simulation studies as well as to fit the joint models constructed by (1) and (2) using the approximate EM algorithm is available in `R` package `gmvjoint` on CRAN.

3. Simulation

3.1. Scenario I: trivariate model

We first consider a simulation study with three longitudinal responses, which we model by Gaussian, Poisson and binomial GLMMs respectively. Across simulations, we fix sample size $n = 250$ and alter the maximal profile lengths $r = \{5, 10, 15\}$. We simulate failure times following the methodology outlined in Austin (2012), where we alter failure rates by controlling the shape and scale of a Gompertz baseline hazard. Independent censoring times were drawn from an exponential distribution with rate e^{-3}. If either the subject's failure or censoring time exceeded 5, their survival time was truncated at time 5.1 and treated as a censored observation. This process of censor-time simulation allows us to control the approximate proportion of failures occurring during follow-up $\omega = \{10\%, 30\%, 50\%\}$. We simulate and fit for each combination of r, ω two hundred times. For each $k = 1, \dots, 3$ the longitudinal response is measured at the regularly-spaced r-vector of possible times (hence $m_{ik} = m_i$), $\boldsymbol{t} = (0, \dots, 5)^{\top}$ (i.e. at $r = 15$ the vector of possible times is more dense), and truncated at T_i for each subject to produce the m_i-vector of measurement times for subject i, $\boldsymbol{t}_i$. For subject $i = 1, \dots, n$ we define y_{ijk} as

$$\begin{cases} y_{ij1} = (\beta_{10} + b_{i10}) + (\beta_{11} + b_{i11})\, t_{ij} + \beta_{12} x_{i1} + \beta_{13} x_{i2} + \varepsilon_{ij} \\ \log\left(\mathbb{E}\left[y_{ij2} | \boldsymbol{b}_{i2}\right]\right) = (\beta_{20} + b_{i20}) + (\beta_{21} + b_{i21})\, t_{ij} + \beta_{22} x_{i1} + \beta_{23} x_{i2} \\ \text{logit}\left(\mathbb{E}\left[y_{ij3} | \boldsymbol{b}_{i3}\right]\right) = (\beta_{30} + b_{i30}) + \beta_{31} t_{ij} + \beta_{32} x_{i1} + \beta_{33} x_{i2}. \end{cases}$$

Specifically, $\boldsymbol{Y}_{i1}$ is Gaussian, $\boldsymbol{Y}_{i2}$ Poisson and $\boldsymbol{Y}_{i3}$ binomial. Fixed-effect covariate x_{i1} is a standard normal deviate, and x_{i2} a single Bernoulli draw ($p = 0.5$). The random effects vector for subject i is $\boldsymbol{b}_i = (b_{i10}, b_{i11}, b_{i20}, b_{i21}, b_{i30})^{\top}$, with $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \mathrm{D})$. Notably then, the binomial response is simulated and fit under a random intercept specification only, whereas we include a random slope term if the response is either continuous or a count. We set $\text{diag}(\mathrm{D}) = (0.25, 0.09, 0.50, 0.10, 2.00)^{\top}$, and we induce correlation between random intercepts (and across responses) by setting $\mathrm{D}_{13} = \mathrm{D}_{15} = \mathrm{D}_{35} = 0.25$ and enforcing symmetry; the resulting correlation ρ_{ef} between the e^{th} and f^{th} random intercepts $\rho_{13} = 0.71$, $\rho_{15} = 0.35$, $\rho_{35} = 0.25$. Fixed effects were $\boldsymbol{\beta}_k = (2, -0.1, 0.1, -0.2)^{\top}$, $k = 1, 2$ and $\boldsymbol{\beta}_3 = (1, -1, 1, -1)^{\top}$. The Gaussian response is simulated with $\sigma_{\varepsilon}^2 = 0.16$. The survival sub-model is then

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ x_{i2}\zeta + \sum_{k=1}^{2} \gamma_k \left(b_{ik0} + b_{ik1} t\right) + \gamma_3 b_{i30} \right\},$$

with association parameters $\boldsymbol{\gamma} = (0.5, -0.5, 0.5)^{\top}$ and time-invariant survival parameter $\zeta = -0.2$.

The parameter estimates for the survival parameters $\boldsymbol{\gamma}$ and ζ are presented in Fig. 1, with detailed estimation capabilities for all parameters given in Supplementary Material II. The time-invariant survival coefficient ζ is well-estimated with minimal bias, and decreasing variability in these estimates as ω increases (i.e. more information is present). The time-varying survival parameters $\boldsymbol{\gamma}$ are generally well-estimated. Generally speaking, we note a reduction in bias as r increases from 5 time-points. However, we note routine underestimation in γ_3, though we do not expect this to improve with longer profiles, since it is attached to the random intercept only.

Table 1
Median [IQR] elapsed time in seconds for simulations considered in simulation scenario I. The proportion who fail is denoted by ω and the maximal length of the longitudinal profiles by r.

	$r = 5$	$r = 10$	$r = 15$
$\omega = 10\%$	4.628 [4.125, 5.237]	3.217 [2.896, 3.608]	3.175 [2.828, 3.631]
$\omega = 30\%$	5.868 [5.167, 6.916]	4.110 [3.720, 4.577]	3.964 [3.620, 4.383]
$\omega = 50\%$	7.422 [6.520, 8.624]	5.237 [4.771, 5.801]	4.819 [4.380, 5.412]

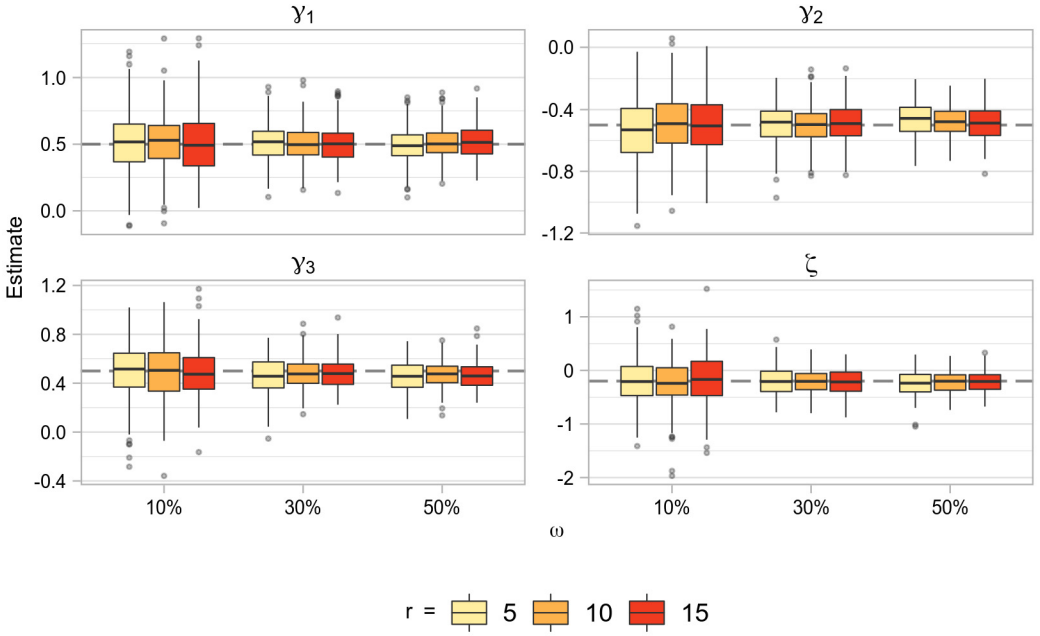

Fig. 1. Estimates for survival parameters $\boldsymbol{\gamma}$ and ζ for differing maximal profile lengths r and failure rates ω. The black dashed line indicates the true parameter value. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

Beyond these survival parameters, we observe broadly good coverage for components of vech (D), apart from D_{55} – attached to the binary sub-model – which routinely has poorest coverage. There does appear to be some systematic underestimation of these variance/covariance components; this bias decreasing as r increases. The fixed effects $\boldsymbol{\beta}$ also appear to enjoy reduction in bias as the profile length increases, with the intercept attached to the binary sub-model suffering the poorest coverage across simulations, with all other components having coverage around the nominal 0.95. Finally we observe the average estimated standard error generally gets closer to the empirical standard deviation as r increases, this suggesting that a longer profile allows the approach to more accurately capture the variability surrounding these parameter estimates. The time taken for the approximate EM algorithm to converge calculation of standard errors is given in Table 1, where we note a longer profile reduces this computation time, and a larger proportion of failures increases it.

3.2. Scenario II: five-variate model

With performative capabilities laid out in a trivariate setting – across differing simulation scenarios – in the previous section, we seek to offer an idea of performance of the algorithm for both a larger sample size $n = 500$ and a greater number of longitudinal responses $K = 5$. Here, we fix the failure rate to approximately 30% and the follow-up period to ten time-points for five hundred simulations, with follow-up period and survival times generated in the same way outlined in Simulation Scenario I. For the two Gaussian and two count responses, $k = 1, \ldots, 4$ we define:

$$\mathbb{E}\left[y_{ijk}|\boldsymbol{b}_{ik}\right] = h_k^{-1}\left((\beta_{k0} + b_{ik0}) + (\beta_{k1} + b_{ik1})\, t_j + \beta_{k2} x_{i1} + \beta_{k3} x_{i2}\right),$$

and for the binary response ($k = 5$),

$$\text{logit}\left(\mathbb{E}\left[y_{ij5}|\boldsymbol{b}_{i5}\right]\right) = (\beta_{50} + b_{i50}) + \beta_{51} t_j + \beta_{52} x_{i1} + \beta_{53} x_{i2}.$$

Here, we set x_{i1} as a standard normal deviate and x_{i2} a Bernoulli draw ($p = 0.5$), and $h_k(\cdot)$ takes the form of the usual link function in each case. We set the fixed effects $\boldsymbol{\beta}_e = (2, -0.1, 0.1, -0.2)^\top$, $e = 1, 3, 4$, $\boldsymbol{\beta}_2 = (-2, 0.1, -0.1, 0.2)^\top$ and $\boldsymbol{\beta}_5 = (1, -1, 1, -1)^\top$. The Gaussian responses are simulated with variance $\sigma^2_{\varepsilon_k} = 0.16$, $k = 1, 2$. The random ef-

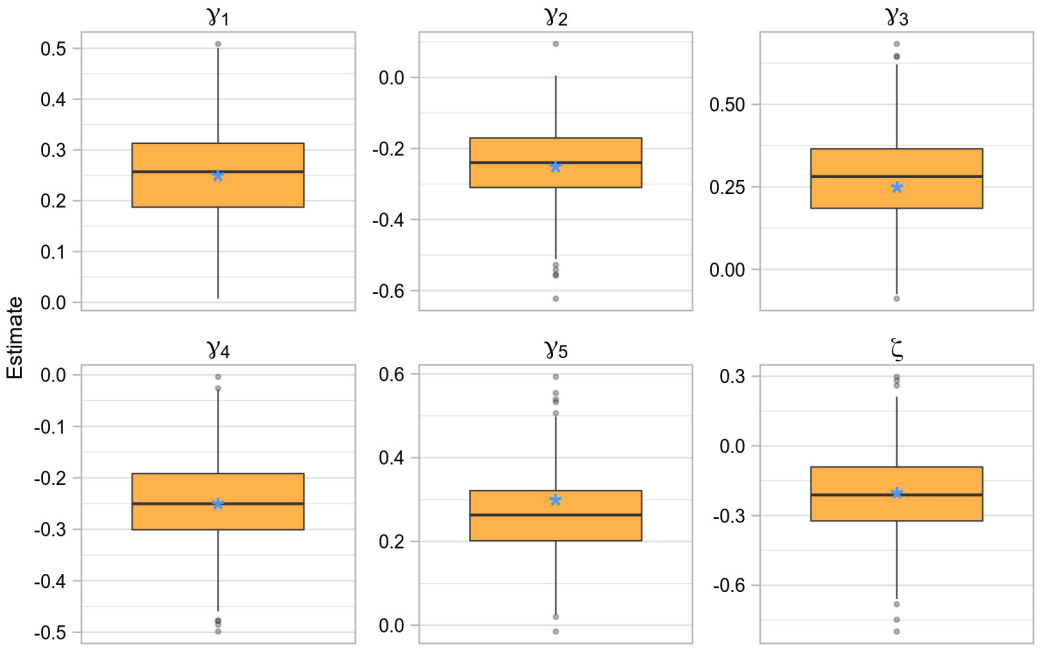

Fig. 2. Estimates for survival parameters $\boldsymbol{\gamma}$ and ζ for the five-variate simulation. A blue asterisk (*) denotes the true value for each parameter.

fects for subject i now have dimension nine, $\boldsymbol{b}_i = (b_{i10}, b_{i11}, b_{i20}, b_{i21}, b_{i30}, b_{i31}, b_{i40}, b_{i41}, b_{i50})^\top$, with $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \mathrm{D})$. As was the case in Scenario I, the binomial response is simulated and fit with a random intercept only. We set $\mathrm{diag}(D) = (0.25, 0.09, 0.30, 0.06, 0.20, 0.04, 0.50, 0.09, 2.00)^\top$ and induce light-to-moderate correlation across responses by setting $\mathrm{D}_{ef} = 0.125,\ e, f = 1, 3, 5, 7, 9,\ e \neq f$ with resulting correlations $\rho_{13} = 0.46,\ \rho_{15} = 0.56,\ \rho_{17} = 0.35,\ \rho_{19} = 0.18,\ \rho_{35} = 0.51,\ \rho_{37} = 0.32,\ \rho_{39} = 0.16,\ \rho_{57} = 0.40,\ \rho_{59} = 0.20,\ \rho_{79} = 0.13$. We then define our hazard as

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ x_{i2}\zeta + \sum_{k=1}^{4} \gamma_k (b_{ik0} + b_{ik1} t) + \gamma_5 b_{i50} \right\},$$

where we set association parameters $\boldsymbol{\gamma} = (0.25, -0.25, 0.25, -0.25, 0.30)^\top$ and $\zeta = -0.2$.

The estimates for the survival parameters $\boldsymbol{\gamma}$ and ζ are presented in Fig. 2 where we note good estimation of all five association parameters as well as the time-invariant ζ. We do note underestimation of γ_5, and overestimation for γ_3 with percentage biases of (-)12.0% and 11.6%, respectively. The median [IQR] elapsed time for the approximate EM algorithm to converge and standard errors calculated was 20.684 [19.358, 22.491] seconds; the escalation in computing times that hampers traditional quadrature approaches is diluted under the proposed approach. Empirical mean (SD) of all 500 sets of model fits are given in Table 2 along with the average estimated standard error, average bias, mean squared error (MSE) and 95% coverage probabilities (CP).

Generally speaking, we do not observe deterioration in estimation capabilities when considering two additional longitudinal responses, though the estimates (and/or their estimated standard errors) could be deemed conservative. As was the case with the trivariate scenario, we note the fixed effect intercept, as well as the variance of the random intercept for the binomial sub-model suffers the poorest coverage. The rest of vech (D) appears to be well estimated, but we once more note systematic *underestimation* of these terms; something that could improve by alterations to convergence criterion (6).

4. Application: primary billiary cirrhosis

We now undertake application to the oft-used primary billiary cirrhosis (PBC) data. PBC is a chronic liver disease which affects the bile ducts of the liver, complications of which can ultimately lead to death. The longitudinal profile of numerous biomarkers were observed for 312 patients at the Mayo Clinic between 1974 and 1984 with patients assigned to either the active (D-penicillamine, $n = 154$ (50.6%)) or placebo treatment arm (Murtaugh et al., 1994).

The presence of many longitudinal biomarkers of clinical interest as well as an event-time has lead to the PBC data becoming a widely used example in literature. For instance, it appears in both univariate joint modelling contexts (Crowther et al., 2013) as well as multivariate using only continuous (i.e. Gaussian) responses (Albert and Shih, 2010; Hickey et al., 2018; Philipson et al., 2020) as well as multivariate GLMM specifications (Andrinopoulou and Rizopoulos, 2016; Rustand et al., 2022b).

Table 2
Parameter estimates for five-variate simulation scenario. 'Emp. Mean (SD)' denotes the average estimated value with the standard deviation of parameter estimates and Mean SE the mean standard error calculated for each parameter from each model fit. Coverage probabilities are calculated from $\hat{\boldsymbol{\Omega}} \pm 1.96\text{SE}(\hat{\boldsymbol{\Omega}})$. The median [IQR] total computation time (e.g. including time taken to obtain initial estimates etc.) was 27.164 [25.859, 28.991] seconds.

Parameter	Emp. Mean (SD)	Mean SE	Bias	MSE	CP
$D_{11} = 0.250$	0.248 (0.020)	0.022	-0.002	0.000	0.962
$D_{31} = 0.125$	0.125 (0.015)	0.018	0.000	0.000	0.974
$D_{51} = 0.125$	0.123 (0.014)	0.016	-0.002	0.000	0.960
$D_{71} = 0.125$	0.123 (0.019)	0.022	-0.002	0.000	0.976
$D_{91} = 0.125$	0.117 (0.043)	0.049	-0.008	0.002	0.970
$D_{22} = 0.090$	0.090 (0.007)	0.008	0.000	0.000	0.968
$D_{33} = 0.300$	0.299 (0.022)	0.026	-0.001	0.000	0.956
$D_{53} = 0.125$	0.125 (0.014)	0.017	0.000	0.000	0.970
$D_{73} = 0.125$	0.123 (0.021)	0.023	-0.002	0.000	0.964
$D_{93} = 0.125$	0.119 (0.047)	0.053	-0.006	0.002	0.966
$D_{44} = 0.060$	0.060 (0.005)	0.005	0.000	0.000	0.968
$D_{55} = 0.200$	0.198 (0.017)	0.019	-0.002	0.000	0.970
$D_{75} = 0.125$	0.122 (0.017)	0.020	-0.003	0.000	0.980
$D_{95} = 0.125$	0.117 (0.039)	0.045	-0.008	0.002	0.974
$D_{66} = 0.040$	0.039 (0.003)	0.004	-0.001	0.000	0.942
$D_{77} = 0.500$	0.484 (0.034)	0.041	-0.016	0.001	0.942
$D_{97} = 0.125$	0.119 (0.059)	0.065	-0.006	0.004	0.962
$D_{88} = 0.090$	0.087 (0.007)	0.008	-0.003	0.000	0.942
$D_{99} = 2.000$	1.696 (0.210)	0.267	-0.304	0.137	0.784
$\beta_{10} = 2.000$	1.990 (0.035)	0.039	-0.010	0.001	0.966
$\beta_{11} = -0.100$	-0.100 (0.015)	0.017	0.000	0.000	0.972
$\beta_{12} = 0.100$	0.100 (0.026)	0.028	0.000	0.001	0.960
$\beta_{13} = -0.200$	-0.200 (0.051)	0.054	0.000	0.003	0.964
$\beta_{20} = -2.000$	-2.007 (0.038)	0.042	-0.007	0.001	0.962
$\beta_{21} = 0.100$	0.103 (0.012)	0.014	0.003	0.000	0.976
$\beta_{22} = -0.100$	-0.098 (0.027)	0.030	0.002	0.001	0.960
$\beta_{23} = 0.200$	0.200 (0.053)	0.058	0.000	0.003	0.970
$\beta_{30} = 2.000$	1.998 (0.032)	0.035	-0.002	0.001	0.966
$\beta_{31} = -0.100$	-0.101 (0.011)	0.012	-0.001	0.000	0.974
$\beta_{32} = 0.100$	0.101 (0.022)	0.025	0.001	0.001	0.970
$\beta_{33} = -0.200$	-0.198 (0.046)	0.049	0.002	0.002	0.962
$\beta_{40} = 2.000$	2.004 (0.049)	0.051	0.004	0.002	0.946
$\beta_{41} = -0.100$	-0.097 (0.016)	0.017	0.003	0.000	0.952
$\beta_{42} = 0.100$	0.100 (0.033)	0.037	0.000	0.001	0.978
$\beta_{43} = -0.200$	-0.205 (0.070)	0.072	-0.005	0.005	0.958
$\beta_{50} = 1.000$	0.891 (0.137)	0.129	-0.109	0.031	0.840
$\beta_{51} = -1.000$	-0.984 (0.042)	0.048	0.016	0.002	0.958
$\beta_{52} = 1.000$	1.034 (0.086)	0.091	0.034	0.009	0.956
$\beta_{53} = -1.000$	-1.035 (0.170)	0.162	-0.035	0.030	0.938
$\sigma_1^2 = 0.160$	0.160 (0.004)	0.004	0.000	0.000	0.982
$\sigma_2^2 = 0.160$	0.160 (0.004)	0.004	0.000	0.000	0.974
$\gamma_1 = 0.250$	0.254 (0.094)	0.099	0.004	0.009	0.940
$\gamma_2 = -0.250$	-0.243 (0.107)	0.114	0.007	0.012	0.960
$\gamma_3 = 0.250$	0.279 (0.137)	0.148	0.029	0.020	0.960
$\gamma_4 = -0.250$	-0.248 (0.085)	0.092	0.002	0.007	0.960
$\gamma_5 = 0.300$	0.264 (0.093)	0.099	-0.036	0.010	0.946
$\zeta = -0.200$	-0.210 (0.175)	0.184	-0.010	0.031	0.962

Nine longitudinal biomarkers exist with varying degrees of completeness in the data. Of these, we consider four to be Gaussian: Log serum bilirubin; log serum aspartate aminotransferase ('AST'); serum albumin and prothrombin time. Three are binary markers which indicate presence of: Enlarged liver (hepatomegaly); accumulation of fluid in abdomen (ascites) and malformed blood vessels in skin ('spiders'). Finally, platelet count and alkaline phosphatase are treated as count biomarkers.

4.1. Modelling approach

We begin by examining the longitudinal profiles of the biomarkers we consider to be Gaussian or counts. We visually appraise them in Fig. 3 and fit candidate GLMMs with different specifications of time. Throughout this process, we monitor model fit and simplicity. We choose between linear, quadratic, or cubic spline time (t) specifications, assuming that recipiency of the study drug interacts with the time specification in all cases. After this relatively informal sub-model selection process, we proceed with cubic splines for platelet count, a quadratic time specification for log(Serum bilirubin), and

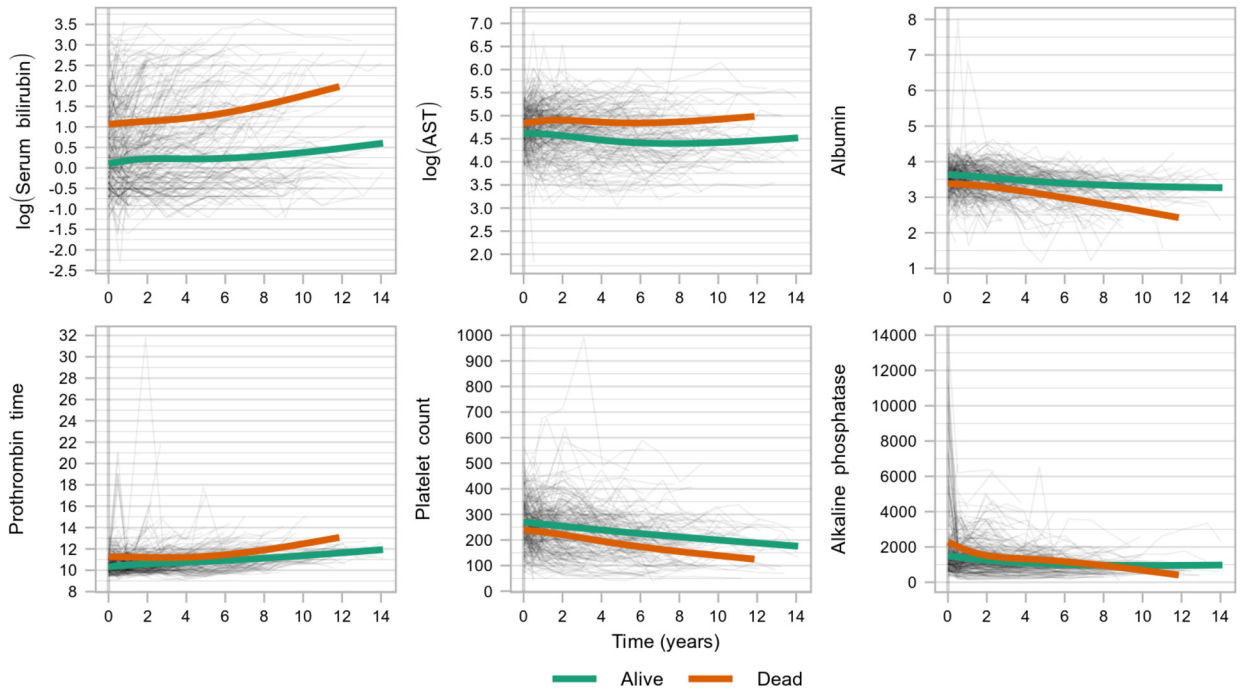

Fig. 3. Longitudinal trajectories for the chosen (non-binary) biomarkers in the application to PBC data. Grey lines show individual trajectories and the overlaid green and orange lines shows a smoothed (LOESS) curve for those who survived and those who died, respectively.

a linear one for the remaining biomarkers: albumin, log(AST), prothrombin, and alkaline. The random effects structure for Gaussian and count responses mirror the fixed effect time specification chosen. For binary biomarkers, we solely consider an intercept-only random effect specification to facilitate model fit. Ascites is not further considered due to its low prevalence among subjects (approximately 33% of patients with at least one incidence and approximately 9% overall prevalence), which leads to problems with model convergence. The analysis set consists of the subset of data with no missing measurements in any of the remaining eight biomarkers listed; the sample size (n) remains unchanged.

Next, we fit three multivariate joint models on the candidate continuous ($K = 4$), count ($K = 2$) and binary ($K = 2$) responses separately. The bivariate joint model for the binary biomarkers implied only hepatomegaly was associated with mortality. In the bivariate model for the count biomarkers, we concluded that only platelet count was significantly associated. Finally, for the four-variate Gaussian model, we conclude that all continuous biomarkers besides prothrombin are significantly associated with mortality; log(AST) being significant at the 10% level. These family-specific joint model fits are presented in Supplementary Material II. We consider then a five-variate joint model:

$$
\begin{aligned}
\log(\text{Serum bilirubin}) &= (\beta_{10} + b_{i10}) + \beta_{11}x_i + (\beta_{12} + b_{i11})t \\
&\quad + (\beta_{13} + b_{i12})t^2 + \beta_{14}x_i t + \beta_{15}x_i t^2 \\
&\quad + \varepsilon_{i1}(t), \\
\text{Albumin} &= (\beta_{20} + b_{i20}) + \beta_{21}x_i + (\beta_{22} + b_{i21})t \\
&\quad + \beta_{23}x_i t + \varepsilon_{i2}(t), \\
\log(\text{AST}) &= (\beta_{30} + b_{i30}) + \beta_{31}x_i + (\beta_{32} + b_{i31})t \\
&\quad + \beta_{33}x_i t + \varepsilon_{i3}(t), \\
\log(\mathbb{E}\left[\text{Platelets}|\boldsymbol{b}_{i4}\right]) &= (\beta_{40} + b_{i40}) + \beta_{41}x_i + (\beta_{42} + b_{i41})N_1(t) \\
&\quad + (\beta_{43} + b_{i42})N_2(t) + (\beta_{44} + b_{i43})N_3(t) \\
&\quad + \beta_{45}x_i N_1(t) + \beta_{46}x_i N_2(t) + \beta_{47}x_i N_3(t) \\
&\quad + \varepsilon_{i4}(t), \\
\text{logit}(\mathbb{E}\left[\text{Hepatomegaly}|b_{i50}\right]) &= (\beta_{50} + b_{i50}) + \beta_{51}x_i + \beta_{52}t \\
&\quad + \beta_{53}x_i t, \\
\lambda_i(t) &= \lambda_0(t)\exp\left\{x_i \times \zeta + \sum_{k=1}^{5}\gamma_k \boldsymbol{W}_k(t)^\top \boldsymbol{b}_{ik}\right\}.
\end{aligned}
\tag{8}
$$

The multivariate joint model (8) then consists of three Gaussian, one count and one binary longitudinal responses. $N_1(t), \ldots, N_3(t)$ denotes the set of natural cubic splines with internal knots at tertiles of follow-up and covariate x_i takes value one if the subject received D-penicillamine. In each case, the random effects $\boldsymbol{b}_{ik}$ are shared with the hazard $\lambda_i(t)$ through function of time $\boldsymbol{W}_k(t)$. We recognise the model-building process undertaken here could have been performed in a variety of ways; the focus here is to loosely justify the choice of biomarkers used in this application without deploying a formal model selection approach. One alternative approach would be to fit a 'saturated' model containing all longitudinal parameters at the outset. Results from such an approach are presented in Supplementary Material II.

4.2. Results

Parameter estimates from the joint model (8) are presented in Table 3. Examining first the linear predictors for the five biomarkers, we note that recipiency of the study drug is not significantly associated with mortality, and decreases both the incidence log-odds of hepatomegaly by (point estimate [95% CI]) -0.648 $[-1.230,\ -0.066]$ as well as levels of log(AST) by -0.149 $[-0.277,\ -0.021]$ units on average; the study drug did not hold significant interaction with any time specification we chose. We found that the levels of log(Serum bilirubin) significantly increased over follow-up (0.164 [0.076, 0.252]), as did hepatomegaly (0.131 [0.081, 0.182]). Levels of platelets decreased over time, along with albumin (-0.091, $[-0.110,\ -0.071]$).

The results for association parameters $\boldsymbol{\gamma}$ allow us to infer that a one unit increase away from the average trajectory in log(Serum bilirubin) significantly increases the log-odds of mortality by 1.195 [0.723, 1.668]; the same true for a one unit decrease in albumin levels (-2.334 $[-3.922,\ -0.766]$); and platelets, which was borderline significant at the 10% level (-0.502, $[-1.103,\ 0.099]$). Both log(AST) and hepatomegaly were not found to be associated with mortality; attenuating due to presence of log(Serum bilirubin), albumin and platelet count, with these biomarkers explaining the purported association we observed at the model-building stage.

Additionally in Table 3 we present the results from an analogous model fit carried out using conventional software `JMbayes2` (Rizopoulos et al., 2021). This MCMC fit was carried out using one chain, with 11,000 iterations after 1,000 iterations of burn-in. We note at the outset that the parameterisation (2) is different across approaches, with `JMbayes2` using the *current value* of the linear predictor in place of the shared random effects, and a smoothed baseline hazard.

We note broadly good agreement in both sign and magnitude for the parameter estimates of fixed effects and elements of vech (D) between the two methods. Some (inferential) differences do exist, such as the Bayesian approach finding levels of log (AST) to significantly decrease over time. In general, we expect MCMC to be the 'gold standard', due to inherent differences in inferential approaches; we discuss this further in Section 5.

5. Discussion

We have presented methodology of an approximate EM algorithm for multivariate joint models with a (potential mixture of) GLMM sub-models and an event time of interest, which we modelled by the usual Cox proportional hazards sub-model. This was achieved by the use of an approximation (5) previously applied to multivariate joint models of wholly Gaussian longitudinal responses and a binary (Bernhardt et al., 2015), or survival sub-model (Murray and Philipson, 2022). We posit that the accuracy and computational efficiency of this approximate method – exhibited through extensive simulation studies in Section 3 – make this approximate EM algorithm an attractive alternative to existing methods for fitting MVJMs with at least one GLMM sub-model: This approach being the natural extension of likelihood-based inference on the more orthodox MVJMs with LMM sub-models.

We simulated under multiple scenarios. In Section 3.1 we altered the failure rate and profile length for a mixture of GLMM sub-models, and in Section 3.2 we considered a five-variate scenario. We note under certain circumstances the algorithm (and/or the standard errors produced by (7)) produced slightly conservative coverage. Given the low computation time, we argue that the trade-off here is a reasonable one. We additionally noted poorer coverage of elements attached to binary responses, which may be related to comparatively less information being available.

We also undertook an application to primary billiary cirrhosis data in Section 4. Here, we considered multiple family-specific joint models with more than one response, and fit a five-variate model to the data with three different families. The estimates obtained from the proposed approximate method were compared to those from an analogous model fit using MCMC techniques (via `JMbayes2`). We observed generally strong agreement between the proposed method and this established 'gold standard'. However, a few discrepancies were noted. It is generally expected that parameter estimates, along with their associated uncertainty, obtained through MCMC methods surpass those derived from maximum likelihood estimation: The sampling technique employed in MCMC explores a wider range of the parameter space, potentially revealing complex interactions that may go unnoticed by the MLE approach; additionally, the symmetric distribution of errors assumed by MLE may be unrealistic when compared to the true posterior distribution sampled by MCMC. Results here may have agreed to a larger extent by e.g. using a different choice of priors, or analogously defining an unspecified baseline hazard in the Bayesian fit for greater model parity.

To that end, if more accurate estimates, or a better handle on uncertainty in the estimates were required then, as recommended by Bernhardt et al. (2015), the algorithm we present could be used to quickly obtain starting values for some MCMC scheme, thereby reducing overall computation time. We would also perhaps recommend that the algorithm is

Table 3
Parameter estimates (SE) for application to PBC data. 'Hepato.': Hepatomegaly. Total computation time for the approximate EM algorithm was 42.211 seconds. The time-invariant survival parameter ζ is not associated with a specific response, as such it is reported separately. Parameter estimates (SD) from `JMbayes2` are additionally reported ('CrI': Credible interval). Computation time for the MCMC scheme in `JMbayes2` fit was 489.440 seconds, with an additional 205.897 seconds elapsed in obtention of its initial conditions. We note that `JMbayes2` reports the *current value* of the linear predictor for each response, so estimates for $\boldsymbol{\gamma}$ are *not directly* comparable. The variance-covariance matrix D_k is reported for each of the responses in the form $D_{k,ef}$ where k denotes the longitudinal response, and e, f the random effect indices.

		Approximate EM		`JMbayes2`	
	Parameter	Estimate (SE)	95% CI	Mean (SD)	95% CrI
log(Serum bilirubin)	$D_{1,00}$	1.005 (0.150)	[0.712, 1.299]	0.991 (0.129)	[0.764, 1.281]
	$D_{1,10}$	0.051 (0.034)	[-0.015, 0.118]	0.035 (0.021)	[-0.002, 0.077]
	$D_{1,20}$	-0.002 (0.005)	[-0.011, 0.007]	0.001 (0.002)	[-0.004, 0.005]
	$D_{1,11}$	0.086 (0.015)	[0.057, 0.116]	0.069 (0.010)	[0.052, 0.094]
	$D_{1,21}$	-0.006 (0.001)	[-0.009, -0.003]	-0.005 (0.001)	[-0.007, -0.003]
	$D_{1,22}$	0.001 (0.000)	[0.000, 0.001]	0.000 (0.000)	[0.000, 0.001]
	β_{10}	0.548 (0.101)	[0.350, 0.747]	0.583 (0.082)	[0.424, 0.743]
	β_{11}	-0.139 (0.142)	[-0.418, 0.139]	-0.132 (0.115)	[-0.361, 0.092]
	β_{12}	0.164 (0.045)	[0.076, 0.252]	0.162 (0.029)	[0.107, 0.220]
	β_{13}	-0.002 (0.005)	[-0.013, 0.008]	0.001 (0.003)	[-0.006, 0.007]
	β_{14}	-0.031 (0.057)	[-0.143, 0.081]	-0.023 (0.040)	[-0.100, 0.055]
	β_{15}	0.004 (0.007)	[-0.010, 0.018]	0.003 (0.004)	[-0.005, 0.012]
	σ_1^2	0.085 (0.003)	[0.079, 0.090]	0.088 (0.006)	[0.081, 0.095]
	γ_1	1.195 (0.241)	[0.723, 1.668]	1.279 (0.184)	[0.932, 1.654]
Albumin	$D_{2,00}$	0.130 (0.018)	[0.096, 0.165]	0.125 (0.013)	[0.101, 0.150]
	$D_{2,10}$	0.000 (0.003)	[-0.007, 0.007]	0.001 (0.002)	[-0.003, 0.005]
	$D_{2,11}$	0.003 (0.001)	[0.001, 0.005]	0.003 (0.001)	[0.002, 0.004]
	β_{20}	3.557 (0.043)	[3.473, 3.641]	3.542 (0.033)	[3.478, 3.606]
	β_{21}	0.006 (0.060)	[-0.112, 0.125]	-0.001 (0.046)	[-0.090, 0.087]
	β_{22}	-0.091 (0.010)	[-0.110, -0.071]	-0.099 (0.008)	[-0.114, -0.084]
	β_{23}	0.007 (0.013)	[-0.019, 0.033]	0.008 (0.010)	[-0.010, 0.027]
	σ_2^2	0.097 (0.002)	[0.093, 0.102]	0.098 (0.006)	[0.091, 0.106]
	γ_2	-2.344 (0.805)	[-3.922, -0.766]	-2.674 (0.451)	[-3.610, -1.788]
log(AST)	$D_{3,00}$	0.181 (0.026)	[0.130, 0.231]	0.181 (0.017)	[0.150, 0.218]
	$D_{3,10}$	0.003 (0.004)	[-0.006, 0.011]	0.002 (0.002)	[-0.002, 0.006]
	$D_{3,11}$	0.004 (0.001)	[0.002, 0.006]	0.004 (0.001)	[0.002, 0.006]
	β_{30}	4.768 (0.048)	[4.675, 4.862]	4.783 (0.037)	[4.711, 4.857]
	β_{31}	-0.149 (0.065)	[-0.277, -0.021]	-0.157 (0.051)	[-0.259, -0.057]
	β_{32}	0.014 (0.012)	[-0.010, 0.037]	0.020 (0.009)	[0.003, 0.038]
	β_{33}	-0.013 (0.014)	[-0.041, 0.016]	-0.011 (0.010)	[-0.032, 0.009]
	σ_3^2	0.075 (0.002)	[0.072, 0.078]	0.075 (0.005)	[0.070, 0.082]
	γ_3	-0.684 (0.530)	[-1.723, 0.355]	-0.680 (0.365)	[-1.416, 0.025]
Platelet count	$D_{4,00}$	0.142 (0.016)	[0.110, 0.174]	0.148 (0.012)	[0.126, 0.173]
	$D_{4,10}$	-0.037 (0.057)	[-0.148, 0.075]	-0.035 (0.032)	[-0.101, 0.024]
	$D_{4,20}$	-0.074 (0.126)	[-0.321, 0.173]	-0.073 (0.047)	[-0.157, 0.026]
	$D_{4,30}$	-0.044 (0.239)	[-0.512, 0.424]	-0.037 (0.088)	[-0.197, 0.145]
	$D_{4,11}$	1.353 (0.184)	[0.991, 1.714]	1.208 (0.158)	[0.918, 1.533]
	$D_{4,21}$	-1.874 (0.425)	[-2.708, -1.040]	-1.336 (0.177)	[-1.673, -0.990]
	$D_{4,31}$	-3.785 (0.772)	[-5.297, -2.272]	-2.877 (0.344)	[-3.563, -2.245]
	$D_{4,22}$	5.024 (1.119)	[2.831, 7.217]	4.006 (0.460)	[3.056, 4.853]
	$D_{4,32}$	8.654 (1.911)	[4.908, 12.399]	6.824 (0.807)	[5.109, 8.265]
	$D_{4,33}$	16.350 (3.369)	[9.746, 22.953]	13.349 (1.483)	[10.234, 15.988]
	β_{40}	5.521 (0.044)	[5.435, 5.608]	5.522 (0.031)	[5.462, 5.583]
	β_{41}	-0.058 (0.058)	[-0.172, 0.056]	-0.061 (0.044)	[-0.147, 0.025]
	β_{42}	-0.182 (0.180)	[-0.534, 0.170]	-0.171 (0.117)	[-0.401, 0.063]
	β_{43}	-0.980 (0.417)	[-1.797, -0.162]	-1.197 (0.212)	[-1.610, -0.801]
	β_{44}	-1.347 (0.770)	[-2.856, 0.162]	-1.714 (0.399)	[-2.482, -0.966]
	β_{45}	0.190 (0.236)	[-0.273, 0.652]	0.104 (0.149)	[-0.187, 0.400]
	β_{46}	-0.405 (0.559)	[-1.500, 0.691]	-0.115 (0.273)	[-0.630, 0.438]
	β_{47}	-0.501 (1.012)	[-2.484, 1.483]	0.059 (0.502)	[-0.899, 1.067]
	γ_4	-0.502 (0.307)	[-1.103, 0.099]	-0.729 (0.219)	[-1.168, -0.318]
Hepato.	$D_{5,00}$	4.293 (0.953)	[2.426, 6.161]	6.898 (1.046)	[5.019, 9.178]
	β_{50}	0.285 (0.213)	[-0.133, 0.704]	0.508 (0.261)	[0.001, 1.031]
	β_{51}	-0.648 (0.297)	[-1.230, -0.066]	-0.703 (0.368)	[-1.428, 0.020]
	β_{52}	0.131 (0.026)	[0.081, 0.182]	0.166 (0.035)	[0.100, 0.234]
	β_{53}	0.041 (0.035)	[-0.027, 0.109]	0.029 (0.049)	[-0.068, 0.123]
	γ_5	0.027 (0.156)	[-0.280, 0.333]	0.007 (0.072)	[-0.142, 0.145]
	ζ	-0.256 (0.374)	[-0.990, 0.478]	-0.222 (0.256)	[-0.730, 0.277]

slightly better suited to scenarios with a lower failure rate and/or longer period of follow-up, although performance is by no means poor if these are not met.

Although it is certainly beneficial to be able to routinely fit joint models with flexibility around both the longitudinal specification as well as the response's underlying distribution, we have considered only three here. Indeed, future work entails moving further toward greater flexibility in the latter here. For instance, it is unlikely that the responses we considered to be counts would necessarily be well-represented by the usual Poisson regression model, as they would likely not exhibit equidispersion. To this end, more emphasis could be placed on implementation of non-standard GLMM sub-models, with the generalised Poisson (Zamani and Ismail, 2012); negative binomial; and Gamma distributions available in package `gmvjoint`.

We have solely considered a univariate event time, which we modelled by the usual Cox proportional hazards model. Identification of a specific event, such as survival beyond a pre-specified time (Bernhardt et al., 2015) or simply drop-out, may also be related to some longitudinal process, leading to the survival sub-model (2) being replaced by a logistic model for this binary outcome. Further examples include presence of orthostatic hypotension (Hwang et al., 2011, 2015); successful pregnancy (Horrocks and van Den Heuvel, 2009) and diagnosis of a depressive disorder (Li et al., 2015). Furthermore, a joint model with a competing risks survival sub-model (Williamson et al., 2008; Rustand et al., 2022b) would be more useful in circumstances where patients can experience multiple events of interest. Additionally, joint frailty copula models could be considered (Emura et al., 2017, 2022; Peng et al., 2018).

Acknowledgements

This research was funded by an Engineering and Physical Sciences Research Council 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 three reviewers for their helpful comments, which substantially improved and refined the manuscript.

Appendix A. Likelihood

The observed data likelihood for subject i is defined in (3). Since we seek to obtain $\hat{\boldsymbol{\Omega}}$ by maximum likelihood, we present the complete data log-likelihoods (ℓ) required by the E-step in (4) i.e.

$$\ell_i = \left\{ \sum_{k=1}^{K} \log f\left(\boldsymbol{Y}_{ik} | \boldsymbol{b}_{ik}; \boldsymbol{\Omega}\right) \right\} + \log f\left(T_i, \Delta_i | \boldsymbol{b}_i; \boldsymbol{\Omega}\right) + \log f\left(\boldsymbol{b}_i | \boldsymbol{\Omega}\right).$$

Here, log-likelihood of the event time process (i.e. Cox PH sub-model) is

$$\begin{aligned} \log f\left(T_i, \Delta_i | \boldsymbol{b}_i; \boldsymbol{\Omega}\right) = \Delta_i \log \lambda_0\left(T_i\right) + \Delta_i \left[\boldsymbol{s}_i^\top \boldsymbol{\zeta} + \sum_{k=1}^{K} \gamma_k \boldsymbol{W}_k(T_i)^\top \boldsymbol{b}_{ik} \right] \\ - \int_0^{T_i} \lambda_0\left(u\right) \exp\left\{ \boldsymbol{s}_i^\top \boldsymbol{\zeta} + \sum_{k=1}^{K} \gamma_k \boldsymbol{W}_k(u)^\top \boldsymbol{b}_{ik} \right\} du, \end{aligned} \tag{A.1}$$

and the log-likelihood of the random effects is given by

$$\log f\left(\boldsymbol{b}_i | \boldsymbol{\Omega}\right) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathrm{D}| - \frac{1}{2} \boldsymbol{b}_i^\top \mathrm{D}^{-1} \boldsymbol{b}_i,$$

where $q = \sum_{k=1}^{K} q_k$ and $|\mathrm{M}|$ denotes the determinant of matrix argument M. The log-likelihood of the k^{th} longitudinal process obviously depends on the model chosen for it. Here, we have considered three families: Gaussian, Poisson and binomial. Irregardless of family, we have linear predictor $\boldsymbol{\eta}_{ik} = \mathrm{X}_{ik}\boldsymbol{\beta}_k + \mathrm{Z}_{ik}\boldsymbol{b}_{ik}$. For the Gaussian we set

$$\log f\left(\boldsymbol{Y}_{ik} | \boldsymbol{\Omega}\right) = -\frac{m_{ik}}{2} \log 2\pi - \frac{1}{2} \log |\mathrm{V}_{ik}| - \frac{1}{2} \left(\boldsymbol{Y}_{ik} - \boldsymbol{\eta}_{ik}\right)^\top \mathrm{V}_{ik}^{-1} \left(\boldsymbol{Y}_{ik} - \boldsymbol{\eta}_{ik}\right), \tag{A.2}$$

where m_{ik} is the number of observations for subject i for the k^{th} response and $\mathrm{V}_{ik} = \sigma^2_{\varepsilon_k} \mathbb{I}_{m_{ik}}$. Here, we introduce $\sigma^2_{\varepsilon_k}$ the residual variance and $\mathbb{I}_x$ an $x \times x$ identity matrix. When $\boldsymbol{Y}_{ik}$ is modelled by a Poisson GLMM,

$$\log f\left(\boldsymbol{Y}_{ik} | \boldsymbol{\Omega}\right) = \sum_{j=1}^{m_{ik}} \left[Y_{ikj} \eta_{ikj} - \exp\left\{\eta_{ikj}\right\} + \log\left(Y_{ikj}!\right) \right]. \tag{A.3}$$

Finally, the binomial log-likelihood is

$$\log f\left(\boldsymbol{Y}_{ik} | \boldsymbol{\Omega}\right) = \sum_{j=1}^{m_{ik}} \left[Y_{ikj} \log\left(\frac{\exp\left\{\eta_{ikj}\right\}}{\exp\left\{\eta_{ikj}\right\} + 1} \right) + \left(1 - Y_{ikj}\right) \log\left(1 - \frac{\exp\left\{\eta_{ikj}\right\}}{\exp\left\{\eta_{ikj}\right\} + 1} \right) \right]. \tag{A.4}$$

Appendix B. M-step details

We seek to elucidate the underlying algorithm utilised to update parameters at, say, iteration (m) to iteration $(m+1)$.

B.1. Update for D

We have

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

Since D is symmetric the partial derivative w.r.t. its inverse is

$$\begin{aligned}\frac{\partial \mathbb{E}_i\left[\log f\left(\boldsymbol{b}_i|\mathrm{D}\right)\right]}{\partial \mathrm{D}^{-1}} &= \frac{1}{2}\mathrm{D} - \frac{1}{2}\mathbb{E}_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^\top\right],\\ \implies \hat{\mathrm{D}} &= \mathbb{E}_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^\top\right].\end{aligned}$$

Next, we utilise the normal approximation (5) i.e., $\mathbb{E}_i\left[\boldsymbol{b}_i\right] = \hat{\boldsymbol{b}}_i$ and $\mathrm{Var}\left[\boldsymbol{b}_i\right] = \hat{\Sigma}_i$. This leads to

$$\begin{aligned}\mathrm{Var}\left[\boldsymbol{b}_i\right] = \hat{\Sigma}_i &= \mathbb{E}_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^\top\right] - \mathbb{E}_i\left[\boldsymbol{b}_i\right]\mathbb{E}_i\left[\boldsymbol{b}_i\right]^\top\\ \implies \mathbb{E}_i\left[\boldsymbol{b}_i\boldsymbol{b}_i^\top\right] &= \hat{\Sigma}_i + \mathbb{E}_i\left[\boldsymbol{b}_i\right]\mathbb{E}_i\left[\boldsymbol{b}_i\right]^\top\\ &= \hat{\Sigma}_i + \hat{\boldsymbol{b}}_i\hat{\boldsymbol{b}}_i^\top.\end{aligned}$$

Finally then, we obtain

$$\hat{\mathrm{D}} = \frac{\sum_{i=1}^{n}\left(\hat{\Sigma}_i + \hat{\boldsymbol{b}}_i\hat{\boldsymbol{b}}_i^\top\right)}{n}.$$

B.2. Update for fixed effects $\boldsymbol{\beta}$

Considering now the updates to the vector of fixed effects $\boldsymbol{\beta}$. For the case when the chosen family is non-Gaussian, we employ a one-step Newton-Raphson iteration to update $\boldsymbol{\beta}$ at iteration (m) to iteration $(m+1)$

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \left[\sum_{i=1}^{n}\mathrm{H}_i\left(\boldsymbol{\beta}^{(m)}\right)\right]^{-1}\left[\sum_{i=1}^{n}s_i\left(\boldsymbol{\beta}^{(m)}\right)\right]$$

where $s_i(\boldsymbol{\beta})$ is the gradient vector of the conditional expectation of the requisite complete data log-likelihood with respect to $\boldsymbol{\beta}$ and $\mathrm{H}_i(\boldsymbol{\beta})$ the matrix of second derivatives for subject i. We now illustrate the procedure to find $s_i(\boldsymbol{\beta})$ for each sub-model considered here in turn. We drop subscript k for notational convenience.

B.2.1. Update for $\boldsymbol{\beta}$: Gaussian

We first set the log-likelihood w.r.t. $\boldsymbol{\beta}$,

$$\ell_i\left(\boldsymbol{\beta}\right) \underset{\boldsymbol{\beta}}{\propto} -\frac{1}{2}\left(\boldsymbol{Y}_i - \boldsymbol{\eta}_i\right)^\top \mathrm{V}_i^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\eta}_i\right).$$

This has expected value

$$\begin{aligned}\mathbb{E}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right] &= -\frac{1}{2}\mathbb{E}_i\left[\left(\boldsymbol{Y}_i - \boldsymbol{\eta}_i\right)^\top \mathrm{V}_i^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\eta}_i\right)\right],\\ &= -\frac{1}{2}\mathrm{Tr}\left\{\mathrm{V}_i^{-1}\mathbb{E}_i\left[\left(\boldsymbol{Y}_i - \boldsymbol{\eta}_i\right)\left(\boldsymbol{Y}_i - \boldsymbol{\eta}_i\right)^\top\right]\right\},\end{aligned}$$

which we can appraise using the approximation (5). We set

$$\boldsymbol{\eta}_i = \mathrm{X}_i\boldsymbol{\beta} + \mathrm{Z}_i\boldsymbol{b}_i \sim N\left(\mathrm{X}_i\boldsymbol{\beta} + \mathrm{Z}_i\hat{\boldsymbol{b}}_i, \mathrm{Z}_i\hat{\Sigma}_i\mathrm{Z}_i^\top\right) = N\left(\boldsymbol{\mu}_i, \mathrm{A}_i\right), \tag{B.1}$$

where A_i is the variance-covariance matrix of the approximated (multivariate) normal distribution. Subsequently, we define $\boldsymbol{\tau}_i = \mathrm{diag}\left(\mathrm{A}_i\right)^{\frac{1}{2}}$ and write the approximated expectation

$$\tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right] = -\frac{1}{2}\sum_{l=1}^{\varrho} w_l \mathrm{Tr}\left\{\mathrm{V}_i^{-1}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i - \boldsymbol{\tau}_i v_l\right)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i - \boldsymbol{\tau}_i v_l\right)^\top\right\},$$

where $w_l, v_l,\ l = 1, \ldots, \varrho$ are the Gauss-Hermite weights and abscissae, respectively; $\sum_{l=1}^{\varrho} w_l = 1,\ \sum_{l=1}^{\varrho} v_l = 0$. In practise these weights and abscissae are found using `gauss.quad.prob` from `R` package `statmod` (Smyth, 2005).

The derivative of this approximate expectation with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial \tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right]}{\partial \boldsymbol{\beta}} = \mathrm{X}_i^\top \mathrm{V}_i^{-1}\sum_{l=1}^{\varrho} w_l\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i - \boldsymbol{\tau}_i v_l\right)$$

where we note since $\sum_{l=1}^{\varrho} w_l = 1,\ \sum_{l=1}^{\varrho} v_l = 0$ the derivative is evaluated at $\hat{\boldsymbol{b}}_i$ only. Equating the above derivative to zero and solving for $\boldsymbol{\beta}$ for all $i = 1, \ldots, n$ we obtain the closed-form update

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathrm{X}_i^\top \mathrm{V}_i^{-1} \mathrm{X}_i\right)^{-1}\left(\sum_{i=1}^{n} \mathrm{X}_i^\top\left[\boldsymbol{Y}_i - \mathrm{Z}_i\hat{\boldsymbol{b}}_i\right]\right).$$

B.2.2. Update for $\boldsymbol{\beta}$: Poisson

The log-likelihood of (A.3) w.r.t. $\boldsymbol{\beta}$ is

$$\ell_i\left(\boldsymbol{\beta}\right) \underset{\boldsymbol{\beta}}{\propto} \boldsymbol{Y}_i^\top \boldsymbol{\eta}_i - \mathbf{1}^\top \exp\left\{\boldsymbol{\eta}_i\right\}$$

where $\mathbf{1}$ is an appropriately-dimensioned vector of ones, with expectation

$$\mathbb{E}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right] = \boldsymbol{Y}_i^\top \mathbb{E}_i\left[\boldsymbol{\eta}_i\right] - \mathbf{1}^\top \mathbb{E}_i\left[\exp\left\{\boldsymbol{\eta}_i\right\}\right].$$

We make use of the approximation (5) in the same way as in (B.1). Since the term $\exp\left\{\boldsymbol{\mu}_i\right\}$ is then lognormal, its expectation is simply $\exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i^2/2\right\}$, with $\boldsymbol{\tau}_i^2$ the diagonal of A_i (i.e. the variances). The approximate expectation is simply

$$\tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right] = \boldsymbol{Y}_i^\top \boldsymbol{\mu}_i - \mathbf{1}^\top \exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i^2/2\right\}.$$

The score for $\boldsymbol{\beta}$ is then given by

$$\frac{\partial \tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right]}{\partial \boldsymbol{\beta}} = \mathrm{X}_i^\top\left(\boldsymbol{Y}_i - \exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i^2/2\right\}\right).$$

B.2.3. Update for $\boldsymbol{\beta}$: binomial

The log-likelihood of (A.3) w.r.t. $\boldsymbol{\beta}$ is

$$\begin{aligned}\ell_i\left(\boldsymbol{\beta}\right) &= \boldsymbol{Y}_i^\top \log\left(\frac{\exp\left\{\boldsymbol{\eta}_i\right\}}{\exp\left\{\boldsymbol{\eta}_i\right\} + \mathbf{1}}\right) + \left(\mathbf{1} - \boldsymbol{Y}_i\right)^\top \log\left(\mathbf{1} - \frac{\exp\left\{\boldsymbol{\eta}_i\right\}}{\exp\left\{\boldsymbol{\eta}_i\right\} + \mathbf{1}}\right), \\ &= \boldsymbol{Y}_i^\top \boldsymbol{\eta}_i - \mathbf{1}^\top \log\left(\exp\left\{\boldsymbol{\eta}_i\right\} + \mathbf{1}\right),\end{aligned}$$

with expectation

$$\mathbb{E}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right] = \boldsymbol{Y}_i^\top \mathbb{E}_i\left[\mathrm{X}_i\boldsymbol{\beta} + \mathrm{Z}_i\boldsymbol{b}_i\right] - \mathbf{1}^\top \mathbb{E}_i\left[\log\left(\exp\left\{\mathrm{X}_i\boldsymbol{\beta} + \mathrm{Z}_i\boldsymbol{b}_i\right\} + \mathbf{1}\right)\right].$$

Once again we utilise approximation (5) identically to (B.1). Subsequently, we can rewrite then the above expectation as

$$\tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right] = \boldsymbol{Y}_i^\top \boldsymbol{\mu}_i - \mathbf{1}^\top \sum_{l=1}^{\varrho} w_l \log\left(\exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i v_l\right\} + \mathbf{1}\right).$$

Then the first derivative with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial \tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\beta}\right)\right]}{\partial \boldsymbol{\beta}} = \mathrm{X}_i^\top\left(\boldsymbol{Y}_i - \sum_{l=1}^{\varrho} w_l\left(\frac{\exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i v_l\right\}}{\exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i v_l\right\} + \mathbf{1}}\right)\right).$$

B.3. Update for dispersion parameter σ_ε^2

The log-likelihood (A.2) pertaining to σ_ε^2 – the residual variance of the normal distribution – for subject i at time-point j is

$$\ell_i\left(\sigma_\varepsilon^2\right) \underset{\sigma_\varepsilon^2}{\propto} -\frac{1}{2}\log\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\left(Y_{ij}-\eta_{ij}\right)^2.$$

This has expected value

$$\mathbb{E}_i\left[\ell_i\left(\sigma_\varepsilon^2\right)\right] = -\frac{1}{2}\log\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\mathbb{E}_i\left[\left(Y_{ij}-\eta_{ij}\right)^2\right].$$

Utilising the normal approximation (5) we set

$$\mathrm{X}_{ij}\boldsymbol{\beta} + \mathrm{Z}_{ij}\boldsymbol{b}_i \sim N\left(\mathrm{X}_{ij}\boldsymbol{\beta} + \mathrm{Z}_{ij}\hat{\boldsymbol{b}}_i, \mathrm{Z}_{ij}\hat{\Sigma}_i{\mathrm{Z}_{ij}}^\top\right) = N\left(\mu_{ij}, \tau_{ij}^2\right),$$

subsequently forming the approximate expectation

$$\tilde{\mathbb{E}}_i\left[\ell_i\left(\sigma_\varepsilon^2\right)\right] = -\frac{1}{2}\log\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{l=1}^{\mathcal{Q}} w_l\left(Y_{ij}-\mu_{ij}-\tau_{ij}\nu_l\right)^2,$$

with derivative w.r.t. σ_ε^2

$$\frac{\partial\tilde{\mathbb{E}}_i\left[\ell_i\left(\sigma_\varepsilon^2\right)\right]}{\partial\sigma_\varepsilon^2} = -\frac{1}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4}\sum_{l=1}^{\mathcal{Q}} w_l\left(Y_{ij}-\mu_{ij}-\tau_{ij}\nu_l\right)^2.$$

Equating to zero and solving for σ_ε^2 across all m_{ik} observed timepoints for all n subjects, we obtain

$$\hat{\sigma}_\varepsilon^2 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{l=1}^{\mathcal{Q}} w_l\left(Y_{ij}-\mu_{ij}-\tau_{ij}\nu_l\right)^2 \Big/ \sum_{i=1}^{n} m_{ik}.$$

B.4. Update for λ_0

We rewrite the log-likelihood for the event-time process shown in (A.1) to eschew the integrand, instead introducing matrices and vectors which contain all required information and functions of time. We now write

$$\begin{aligned}\log f\left(T_i, \Delta_i|\boldsymbol{b}_i;\boldsymbol{\Omega}\right) = \Delta_i\log\lambda_0(T_i) + \Delta_i\left[\boldsymbol{S}_i^\top\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\boldsymbol{F}_{ik}^\top\boldsymbol{b}_{ik}\right] - \\ \lambda_0\left(\boldsymbol{u}_i\right)^\top\exp\left\{\mathrm{S}_i\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\mathrm{F}_{\boldsymbol{u}_ik}\boldsymbol{b}_{ik}\right\},\end{aligned} \tag{B.2}$$

where we introduce multiple items for notational convenience and brevity. The vector of failure times survived by subject i is denoted by $\boldsymbol{u}_i$. Vector $\boldsymbol{F}_{ik}$ denotes $\boldsymbol{W}_k(t)$ evaluated at T_i; if response k is to be modelled by an intercept-and-slope random effects specification then $\boldsymbol{F}_{ik} = (1, T_i)^\top$. The matrix $\mathrm{F}_{\boldsymbol{u}_ik}$ is defined in a similar spirit, except its *rows* are determined by $\boldsymbol{W}_k(t)$ evaluated at each element of $\boldsymbol{u}_i$. For convenience's sake we have also introduced S_i, simply representing a $\mathrm{len}\left(\boldsymbol{u}_i\right)\times p_s$ matrix, whose rows are replicates of $\boldsymbol{S}_i$.

With our rewritten log-likelihood established above, we can consider the conditional expectation on $\lambda_0(\cdot)$:

$$\begin{aligned}\ell_i(\lambda_0) &\underset{\lambda_0}{\propto} \Delta_i\log\lambda_0(T_i) - \lambda_0\left(\boldsymbol{u}_i\right)^\top\exp\left\{\mathrm{S}_i\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\mathrm{F}_{\boldsymbol{u}_ik}\boldsymbol{b}_{ik}\right\} \\ \mathbb{E}_i\left[\ell_i(\lambda_0)\right] &= \Delta_i\log\lambda_0(T_i) - \lambda_0(\boldsymbol{u}_i)^\top\mathbb{E}_i\left[\exp\left\{\mathrm{S}_i\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\mathrm{F}_{\boldsymbol{u}_ik}\boldsymbol{b}_{ik}\right\}\right].\end{aligned}$$

To evaluate the expectation $\mathbb{E}_i\left[\exp\left\{\mathrm{S}_i\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\mathrm{F}_{\boldsymbol{u}_ik}\boldsymbol{b}_{ik}\right\}\right]$, we make use of normal approximation (5)

$$\mathrm{S}_i\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\mathrm{F}_{\boldsymbol{u}_ik}\boldsymbol{b}_{ik} \overset{\text{appx.}}{\sim} N\left(\mathrm{S}_i\boldsymbol{\zeta} + \sum_{k=1}^{K}\gamma_k\mathrm{F}_{\boldsymbol{u}_ik}\hat{\boldsymbol{b}}_{ik}, \mathrm{Q}\hat{\Sigma}_i\mathrm{Q}^\top\right) = N(\boldsymbol{\mu}_i, \mathrm{A}_i), \tag{B.3}$$

with A_i previously defined and the matrix $\mathrm{Q} = \mathrm{F}_{\boldsymbol{u}_i} \mathrm{diag}\left(\boldsymbol{\gamma}^*\right)$, where $\boldsymbol{\gamma}^*$ is the vector with q_k replicates of $\gamma_k \ \forall \ k = 1, \ldots, K$ and $\mathrm{F}_{\boldsymbol{u}_i}$ is the horizontal concatenation of matrices $\mathrm{F}_{\boldsymbol{u}_i 1}, \ldots, \mathrm{F}_{\boldsymbol{u}_i K}$. We can then approximate the expectation above as

$$\tilde{\mathbb{E}}_i \left[\ell_i(\lambda_0)\right] = \Delta_i \log \lambda_0(T_i) - \lambda_0(\boldsymbol{u}_i)^\top \sum_{l=1}^{\mathcal{Q}} w_l \exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i v_l\right\}.$$

Now, taking derivative with respect to $\lambda_0(\cdot)$, we can trivially form the update for $\hat{\lambda}_0(\cdot)$

$$\hat{\lambda}_0(u) = \frac{\sum_{i=1}^n \Delta_i I\left(T_i = u\right)}{\sum_{i=1}^n \sum_{l=1}^{\mathcal{Q}} w_l \exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i v_l\right\} I\left(T_i \geq u\right)}, \tag{B.4}$$

where $I(\cdot)$ is the indicator function. The estimate for $\lambda_0(\cdot)$ is updated at each EM iteration, but is not monitored for convergence of the algorithm.

B.5. Update for survival parameters $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$

Utilising the rewritten survival log-likelihood (B.2), we can calculate the expectation on the survival pair $\boldsymbol{\Phi} = \left(\boldsymbol{\zeta}^\top, \boldsymbol{\gamma}^\top\right)^\top$:

$$\ell_i\left(\boldsymbol{\Phi}\right) \underset{\boldsymbol{\Phi}}{\propto} \Delta_i \left\{\boldsymbol{S}_i^\top \boldsymbol{\zeta} + \sum_{k=1}^K \gamma_k \boldsymbol{F}_{ik}^\top \boldsymbol{b}_{ik}\right\} - \lambda_0\left(\boldsymbol{u}_i\right)^\top \exp\left\{\mathrm{S}_i \boldsymbol{\zeta} + \sum_{k=1}^K \gamma_k \mathrm{F}_{\boldsymbol{u}_i k} \boldsymbol{b}_{ik}\right\},$$

$$\mathbb{E}_i\left[\ell_i\left(\boldsymbol{\Phi}\right)\right] = \Delta_i \left\{\boldsymbol{S}_i^\top \boldsymbol{\zeta} + \sum_{k=1}^K \gamma_k \boldsymbol{F}_{ik}^\top \mathbb{E}_i\left[\boldsymbol{b}_{ik}\right]\right\} - \lambda_0(\boldsymbol{u}_i)^\top \mathbb{E}_i\left[\exp\left\{\mathrm{S}_i \boldsymbol{\zeta} + \sum_{k=1}^K \gamma_k \mathrm{F}_{\boldsymbol{u}_i k} \boldsymbol{b}_{ik}\right\}\right].$$

Here, we take $\mathbb{E}_i\left[\boldsymbol{b}_{ik}\right] = \hat{\boldsymbol{b}}_{ik}$ and the approximation used to evaluate $\mathbb{E}_i\left[\exp\left\{\mathrm{S}_i \boldsymbol{\zeta} + \sum_{k=1}^K \gamma_k \mathrm{F}_{\boldsymbol{u}_i k} \boldsymbol{b}_{ik}\right\}\right]$ is the same as (B.3). Thus the approximate expectation is

$$\tilde{\mathbb{E}}_i\left[\ell_i\left(\boldsymbol{\Phi}\right)\right] = \Delta_i \left\{\boldsymbol{S}_i^\top \boldsymbol{\zeta} + \sum_{k=1}^K \gamma_k \boldsymbol{F}_{ik}^\top \hat{\boldsymbol{b}}_{ik}\right\} - \lambda_0(\boldsymbol{u}_i)^\top \sum_{l=1}^{\mathcal{Q}} w_l \exp\left\{\boldsymbol{\mu}_i + \boldsymbol{\tau}_i v_l\right\}, \tag{B.5}$$

wherein we substitute $\lambda_0(\cdot)$ with $\hat{\lambda}_0(\cdot)$ from (B.4). We note a closed-form solution for $\hat{\boldsymbol{\Phi}}$ doesn't exist. Thus, we opt for a one-step Newton-Raphson iteration to update $\boldsymbol{\Phi}$ at iteration (m) to iteration $(m+1)$

$$\boldsymbol{\Phi}^{(m+1)} = \boldsymbol{\Phi}^{(m)} - \left[\sum_{i=1}^n \mathrm{H}_i\left(\boldsymbol{\Phi}^{(m)}\right)\right]^{-1} \left[\sum_{i=1}^n s_i\left(\boldsymbol{\Phi}^{(m)}\right)\right],$$

where $s_i\left(\boldsymbol{\Phi}\right)$ is the gradient vector of (B.5) and H_i the matrix of second derivatives w.r.t. elements of $\boldsymbol{\Phi}$. In practise $s_i\left(\boldsymbol{\Phi}\right)$ and $\mathrm{H}_i\left(\boldsymbol{\Phi}\right), \ \forall \ i = 1, \ldots, n$ are calculated by central differencing.

Appendix C. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2023.107819.

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.

Albert, P.S., Shih, J.H., 2010. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. Ann. Appl. Stat. 4, 1517–1532. https://doi.org/10.1214/10-AOAS339.

Andrinopoulou, E.R., Clancy, J.P., Szczesniak, R., 2020. Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. BMC Polm. Med. 20, 1–11. https://doi.org/10.1186/s12890-020-1177-z.

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.

Austin, P.C., 2012. Generating survival times to simulate cox proportional hazards models with time-varying covariates. Stat. Med. 31, 3946–3958.

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.

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.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M., Bolker, B.M., 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 9, 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html.

Choi, J., Cai, J., Zeng, D., Olshan, A.F., 2015. Joint analysis of survival time and longitudinal categorical outcomes. Stat. Biosci. 7, 19–47. https://doi.org/10.1007/s12561-013-9091-z.

Crowther, M.J., Abrams, K.R., Lambert, P.C., 2013. Joint modeling of longitudinal and survival data. Stata J. 13, 165–184. https://doi.org/10.1177/1536867X1301300112.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. B 39, 1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
Emura, T., Michimae, H., Matsui, S., 2022. Dynamic risk prediction via a joint frailty-copula model and IPD meta-analysis: building web applications. Entropy 24, 589. https://doi.org/10.3390/e24050589.
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.
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., 2018. 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.
Horrocks, J., van Den Heuvel, M.J., 2009. Prediction of pregnancy: a joint model for longitudinal and binary data. Bayesian Anal. 4, 523–538. https://doi.org/10.1214/09-BA419.
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.
Hwang, Y.T., Wang, C.C., Wang, C.H., Tseng, Y.K., Chang, Y.J., 2015. Joint model of multiple longitudinal measures and a binary outcome: an application to predict orthostatic hypertension for subacute stroke patients. Biom. J. 57, 661–675. https://doi.org/10.1002/bimj.201400044.
Li, C., Xiao, L., Luo, S., 2021. Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's disease. Biometrics. https://doi.org/10.1111/biom.13427.
Li, D., Wang, X., Song, S., Zhang, N., Dey, D.K., 2015. Flexible link functions in a joint model of binary and longitudinal data. Stat 4, 320–330. https://doi.org/10.1002/sta4.98.
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.
Long, J.D., Mills, J.A., 2018. Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. BMC Med. Res. Methodol. 18, 1–15. https://doi.org/10.1186/s12874-018-0592-9.
Martins, R., 2022. A flexible link for joint modelling longitudinal and survival data accounting for individual longitudinal heterogeneity. Stat. Methods Appl. 31, 41–61. https://doi.org/10.1007/s10260-021-00566-6.
McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions, 2nd ed. Wiley-Interscience.
Murray, J., Philipson, P., 2022. A fast approximate EM algorithm for joint models of survival and multivariate longitudinal data. Comput. Stat. Data Anal. 170, 107438. https://doi.org/10.1016/j.csda.2022.107438.
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.
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.
Rizopoulos, D., 2012a. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Comput. Stat. Data Anal. 56, 491–501. https://doi.org/10.1016/j.csda.2011.09.007.
Rizopoulos, D., 2012b. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. CRC Press.
Rizopoulos, D., Ghosh, P., 2011. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat. Med. 30, 1366–1380. https://doi.org/10.1002/sim.4205.
Rizopoulos, D., Papageorgiou, G., Miranda Afonso, P., 2021. JMbayes2: extended joint models for longitudinal and time-to-event data. https://CRAN.R-project.org/package=JMbayes2. r package version 0.1-8.
Rustand, D., Krainski, E.T., Rue, H., van Niekerk, J., 2022a. INLAjoint: multivariate joint modeling for longitudinal and time-to-event outcomes with INLA. https://github.com/DenisRustand/INLAjoint.
Rustand, D., van Niekerk, J., Krainski, E.T., Rue, H., Proust-Lima, C., 2022b. Fast and flexible inference approach for joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations. https://arxiv.org/abs/2203.06256.
Smyth, G.K., 2005. Numerical integration. In: Encyclopedia of Biostatistics, pp. 3088–3095.
Sunethra, A., Sooriyarachchi, M., 2021. A novel method for joint modeling of survival data and count data for both simple randomized and cluster randomized data. Commun. Stat., Theory Methods 50, 4180–4202. https://doi.org/10.1080/03610926.2020.1713366.
Sunethra, A., Sooriyarachchi, R., 2018. A joint model for exponential survival data and Poisson count data. Am. J. Appl. Math. Stat. 6, 72–79. https://doi.org/10.12691/ajams-6-2-6.
Therneau, T.M., 2015. A package for survival analysis in S. https://CRAN.R-project.org/package=survival. Version 2.38.
Tsiatis, A.A., Davidian, M., 2004. Joint modeling of longitudinal and time-to-event data: an overview. Stat. Sin. 14, 809–834.
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.
Xu, C., Baines, P.D., Wang, J.L., 2014. Standard error estimation using the EM algorithm for the joint modeling of survival and longitudinal data. Biostatistics 15, 731–744. https://doi.org/10.1093/biostatistics/kxu015.
Zamani, H., Ismail, N., 2012. Functional form for the generalized Poisson regression model. Commun. Stat., Theory Methods 41, 3666–3675. https://doi.org/10.1080/03610926.2011.564742.
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.