Joint modelling of multivariate longitudinal and time-to-event data

Graeme L. Hickey

Department of Biostatistics, University of Liverpool, UK
graeme.hickey@liverpool.ac.uk

9th International Conference of the ERCIM WG on Computational and Methodological Statistics, Seville, Spain

Outline of talk

UNIVERSITY OF LIVERPOOL

Motivation for *multivariate* joint models

UNIVERSITY OF LIVERPOOL

- Clinical studies often repeatedly measure *multiple* biomarkers or other measurements **and** an event time
- Research has predominantly focused on a single event time and single measurement outcome
- Ignoring correlation leads to bias and reduced efficiency in estimation
- Harnessing all available information in a single model is advantageous and should lead to improved model predictions

What is the state of the field?

Hickey *et al. BMC Medical Research Methodology* (2016) 16:117
DOI 10.1186/s12874-016-0212-5

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

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

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

- A large number of models published over recent years incorporating different outcome types; distributions, multivariate event times; estimation approaches; association structures; disease areas; etc.
- Early adoption into clinical literature, but a lack of software!

Data

For each subject $i = 1, \ldots, n$, we observe

- $y_i = (y_{i1}^\top, \ldots, y_{iK}^\top)$ is the K-variate continuous outcome vector, where each y_{ik} denotes an $(n_{ik} \times 1)$-vector of observed longitudinal measurements for the k-th outcome type: $y_{ik} = (y_{i1k}, \ldots, y_{in_{ik}k})^\top$
- Observation times t_{ijk} for $j = 1, \ldots, n_{ik}$, which can differ between subjects and outcomes
- (T_i, δ_i), where $T_i = \min(T_i^*, C_i)$, where T_i^* is the true event time, C_i corresponds to a potential right-censoring time, and δ_i is the failure indicator equal to 1 if the failure is observed $(T_i^* \leq C_i)$ and 0 otherwise

Longitudinal sub-model

$$y_{ik}(t) = \mu_{ik}(t) + W_{1i}^{(k)}(t) + \varepsilon_{ik}(t),$$

where

- $\varepsilon_{ik}(t)$ is the model error term, which is i.i.d. $N(0, \sigma_k^2)$ and independent of $W_{1i}^{(k)}(t)$
- $\mu_{ik}(t) = x_{ik}^{\top}(t)\beta_k$ is the mean response
- $x_{ik}(t)$ is a p_k-vector of (possibly) time-varying covariates with corresponding fixed effect terms β_k
- $W_{1i}^{(k)}(t)$ is a zero-mean *latent* Gaussian process

Time-to-event sub-model

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ v_i^\top(t)\gamma_v + W_{2i}(t) \right\},$$

where

- $\lambda_0(\cdot)$ is an unspecified baseline hazard function
- $v_i(t)$ is a q-vector of (possibly) time-varying covariates with corresponding fixed effect terms γ_v
- $W_{2i}(t)$ is a zero-mean *latent* Gaussian process, independent of the censoring process

Association structure

Following Laird and Ware (1982):

$$W_{1i}^{(k)}(t) = z_{ik}^{\top}(t) b_{ik}$$

Association structure

Following Laird and Ware (1982):

$$W_{1i}^{(k)}(t) = z_{ik}^{\top}(t) b_{ik}$$

1. Within-subject correlation between longitudinal measurements: $b_{ik} \sim N(0, D_{kk})$

Association structure

Following Laird and Ware (1982):

$$W_{1i}^{(k)}(t) = z_{ik}^{\top}(t) b_{ik}$$

1. Within-subject correlation between longitudinal measurements: $b_{ik} \sim N(0, D_{kk})$
2. Between longitudinal outcomes correlation: $\text{cov}(b_{ik}, b_{il}) = D_{kl}$ for $k \neq l$

Association structure

Following Laird and Ware (1982):

$$W_{1i}^{(k)}(t) = z_{ik}^{\top}(t) b_{ik}$$

1. Within-subject correlation between longitudinal measurements: $b_{ik} \sim N(0, D_{kk})$
2. Between longitudinal outcomes correlation: $\text{cov}(b_{ik}, b_{il}) = D_{kl}$ for $k \neq l$
3. Correlation between sub-models[1]: $W_{2i}(t) = \sum_{k=1}^{K} \gamma_{yk} W_{1i}^{(k)}(t)$

[1] Extends model proposed Henderson et al. 2000, although many other $W_{2i}(t)$ specifications have been proposed in literature

Estimation

The estimation methodology mainly follows the 3 seminal works:

1. Wulfsohn, MS and Tsiatis, AA (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* 53(1), pp. 330–339
2. Henderson, R et al. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1(4), pp. 465–480
3. Lin, H et al. (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med* 21, pp. 2369–2382

Lin et al. (2002) is specific to multivariate longitudinal data

Likelihood

We can re-write the longitudinal sub-model as

$$y_i \mid b_i, \beta, \Sigma_i \sim N(X_i\beta + Z_i b_i, \Sigma_i), \text{ with } b_i \mid D \sim N(0, D),$$

where $\beta = (\beta_1^\top, \ldots, \beta_K^\top)$, and

$$X_i = \begin{pmatrix} X_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{iK} \end{pmatrix}, \quad D = \begin{pmatrix} D_{11} & \cdots & D_{1K} \\ \vdots & \ddots & \vdots \\ D_{1K}^\top & \cdots & D_{KK} \end{pmatrix}$$

$$Z_i = \begin{pmatrix} Z_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_{iK} \end{pmatrix}, \quad \Sigma_i = \begin{pmatrix} \sigma_1^2 I_{n_{i1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_K^2 I_{n_{iK}} \end{pmatrix}$$

Likelihood

The *observed* data likelihood is given by

$$\prod_{i=1}^{n}\left(\int_{-\infty}^{\infty} f(y_i \mid b_i,\theta) f(T_i,\delta_i \mid b_i,\theta) f(b_i \mid \theta) db_i\right)$$

where $\theta = (\beta^\top, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^\top, \gamma_y^\top)$

Likelihood

The *observed* data likelihood is given by

$$\prod_{i=1}^{n}\left(\int_{-\infty}^{\infty} f(y_i \mid b_i,\theta) f(T_i,\delta_i \mid b_i,\theta) f(b_i \mid \theta) db_i\right)$$

where $\theta = (\beta^\top, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^\top, \gamma_y^\top)$, and

$$f(y_i \mid b_i,\theta) = \left(\prod_{k=1}^{K}(2\pi)^{-\frac{n_{ik}}{2}}\right)|\Sigma_i|^{-\frac{1}{2}}$$
$$\exp\left\{-\frac{1}{2}(y_i - X_i\beta - Z_i b_i)^\top \Sigma_i^{-1}(y_i - X_i\beta - Z_i b_i)\right\}$$

Likelihood

The *observed* data likelihood is given by

$$\prod_{i=1}^{n}\left(\int_{-\infty}^{\infty} f(y_i \mid b_i,\theta) f(T_i,\delta_i \mid b_i,\theta) f(b_i \mid \theta) db_i\right)$$

where $\theta = (\beta^\top, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^\top, \gamma_y^\top)$, and

$$f(T_i,\delta_i \mid b_i;\theta) = \left[\lambda_0(T_i)\exp\left\{v_i^\top\gamma_v + W_{2i}(T_i,b_i)\right\}\right]^{\delta_i} \exp\left\{-\int_0^{T_i}\lambda_0(u)\exp\left\{v_i^\top\gamma_v + W_{2i}(u,b_i)\right\}du\right\}$$

Likelihood

The *observed* data likelihood is given by

$$\prod_{i=1}^{n}\left(\int_{-\infty}^{\infty} f(y_i \mid b_i,\theta) f(T_i,\delta_i \mid b_i,\theta) f(b_i \mid \theta) db_i\right)$$

where $\theta = (\beta^\top, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^\top, \gamma_y^\top)$, and

$$f(b_i \mid \theta) = (2\pi)^{-\frac{r}{2}} |D|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} b_i^\top D^{-1} b_i\right\},$$

with $r = \sum_{k=1}^{K} r_k$ is the total dimensionality of the random effects variance-covariance matrix.

EM algorithm (Dempster et al. 1977)

UNIVERSITY OF LIVERPOOL

E-step. At the m-th iteration, we compute the expected log-likelihood of the *complete* data conditional on the *observed* data and the current estimate of the parameters.

$$
\begin{aligned}
Q(\theta \mid \hat{\theta}^{(m)}) &= \sum_{i=1}^{n} \mathbb{E}\Big\{ \log f(y_i, T_i, \delta_i, b_i \mid \theta) \Big\}, \\
&= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \Big\{ \log f(y_i, T_i, \delta_i, b_i \mid \theta) \Big\} f(b_i \mid T_i, \delta_i, y_i; \hat{\theta}^{(m)}) db_i
\end{aligned}
$$

EM algorithm (Dempster et al. 1977)

UNIVERSITY OF LIVERPOOL

E-step. At the m-th iteration, we compute the expected log-likelihood of the *complete* data conditional on the *observed* data and the current estimate of the parameters.

$$\begin{aligned} Q(\theta \mid \hat{\theta}^{(m)}) &= \sum_{i=1}^{n} \mathbb{E}\Big\{ \log f(y_i, T_i, \delta_i, b_i \mid \theta) \Big\}, \\ &= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \Big\{ \log f(y_i, T_i, \delta_i, b_i \mid \theta) \Big\} f(b_i \mid T_i, \delta_i, y_i; \hat{\theta}^{(m)}) db_i \end{aligned}$$

M-step. We maximise $Q(\theta \mid \hat{\theta}^{(m)})$ with respect to θ. namely,

$$\hat{\theta}^{(m+1)} = \arg \max_{\theta} Q(\theta \mid \hat{\theta}^{(m)})$$

M-step: closed form estimators

$$
\begin{aligned}
\hat{\lambda}_0(t) &= \frac{\sum_{i=1}^{n} \delta_i I(T_i = t)}{\sum_{i=1}^{n} \mathbb{E}\left[\exp\left\{v_i^\top \gamma_v + W_{2i}(t, b_i)\right\}\right] I(T_i \geq t)} \\
\hat{\beta} &= \left(\sum_{i=1}^{n} X_i^\top X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^\top (y_i - Z_i \mathbb{E}[b_i])\right) \\
\hat{\sigma}_k^2 &= \frac{1}{\sum_{i=1}^{n} n_{ik}} \sum_{i=1}^{n} \Big\{ (y_{ik} - X_{ik}\beta_k)^\top (y_{ik} - X_{ik}\beta_k - 2Z_{ik}\mathbb{E}[b_{ik}]) \\
&\qquad + \text{trace}\left(Z_{ik}^\top Z_{ik} \mathbb{E}[b_{ik} b_{ik}^\top]\right) \Big\} \\
\hat{D} &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[b_i b_i^\top\right]
\end{aligned}
$$

M-step: non-closed form estimators

There is no closed form update for $\gamma = (\gamma_v^\top, \gamma_y^\top)$, so use a one-step Newton-Raphson iteration

$$\hat{\gamma}^{(m+1)} = \hat{\gamma}^{(m)} + I\left(\hat{\gamma}^{(m)}\right)^{-1} S\left(\hat{\gamma}^{(m)}\right),$$

where

$$\begin{aligned} S(\gamma) &= \sum_{i=1}^{n} \left[\delta_i \mathbb{E}\left[\tilde{v}_i(T_i)\right] - \int_0^{T_i} \lambda_0(u) \mathbb{E}\left[\tilde{v}_i(u) \exp\{\tilde{v}_i^\top(u)\gamma\}\right] du \right] \\ I(\gamma) &= -\frac{\partial}{\partial \gamma} S(\gamma) \end{aligned}$$

with $\tilde{v}_i(t) = \left(v_i^\top, z_{i1}^\top(t) b_{i1}, \ldots, z_{iK}^\top(t) b_{iK}\right)$ a $(q+K)$–vector

M-step: non-closed form estimators

- Calculation of $I(\gamma)$ is the computational bottleneck of the estimation algorithm
- computation time $\mathcal{O}(DJ^2)$ (D = number of MC samples; J = number of unique failure times)
- Accounts for 76% of algorithm time in typical example problem
- **Possible solution**: use a Gauss-Newton-like approximation for $I(\gamma)$?

E-step

$$\mathbb{E}\left[h(b_i) \mid T_i, \delta_i, y_i; \hat{\theta}\right] = \frac{\int_{-\infty}^{\infty} h(b_i) f(b_i \mid y_i; \hat{\theta}) f(T_i, \delta_i \mid b_i; \hat{\theta}) db_i}{\int_{-\infty}^{\infty} f(b_i \mid y_i; \hat{\theta}) f(T_i, \delta_i \mid b_i; \hat{\theta}) db_i},$$

where

$$\begin{aligned} h(\cdot) &= \text{any known fuction,} \\ b_i \mid y_i, \theta &\sim N\left(A_i \left\{Z_i^\top \Sigma_i^{-1}(y_i - X_i\beta)\right\}, A_i\right), \text{ and} \\ A_i &= \left(Z_i^\top \Sigma_i^{-1} Z_i + D^{-1}\right)^{-1} \end{aligned}$$

Monte Carlo E-step

Expectations might be unruly if $r = \dim(b_i)$ is large, so use Monte Carlo integration $\Rightarrow$ Monte Carlo Expectation-Maximization (MCEM) algorithm (Wei and Tanner 1990)

$$\mathbb{E}\left[h(b_i) \mid T_i, \delta_i, y_i; \hat{\theta}\right] \approx \frac{\frac{1}{N}\sum_{d=1}^{N} h\left(b_i^{(d)}\right) f\left(T_i, \delta_i \mid b_i^{(d)}; \hat{\theta}\right)}{\frac{1}{N}\sum_{d=1}^{N} f\left(T_i, \delta_i \mid b_i^{(d)}; \hat{\theta}\right)}$$

where $b_i^{(1)}, b_i^{(2)}, \ldots, b_i^{(D)}$ are a random sample from $b_i \mid y_i, \theta$

Monte Carlo E-step

As proposed by Henderson et al. (2000), we use antithetic simulation for variance reduction instead of directly sampling from the MVN distribution for $b_i \mid y_i; \hat{\theta}$:

Sample $\Omega \sim N(0, I_r)$ and obtain the *pairs*

$$A_i \left\{ Z_i^\top \Sigma_i^{-1} (y_i - X_i \beta) \right\} \pm C_i \Omega,$$

where C_i is the Cholesky decomposition of A_i such that $C_i C_i^\top = A_i$

Negative correlation between the pairs $\Rightarrow$ smaller variance in the sample means than would be obtained from N independent simulations

Convergence

In standard EM, convergence usually declared at $(m+1)$-th iteration if one of the following criteria satisfied

- Relative change: $\Delta_{\text{rel}}^{(m+1)} = \max\left\{\frac{|\hat{\theta}^{(m+1)} - \hat{\theta}^{(m)}|}{|\hat{\theta}^{(m)}| + \epsilon_1}\right\} < \epsilon_0$
- Absolute change: $\Delta_{\text{abs}}^{(m+1)} = \max\left\{|\hat{\theta}^{(m+1)} - \hat{\theta}^{(m)}|\right\} < \epsilon_2$

for some choice of ϵ_0, ϵ_1, and ϵ_2

Convergence

UNIVERSITY OF LIVERPOOL

In MCEM framework, there are 2 complications to account for

Convergence

In MCEM framework, there are 2 complications to account for

1. spurious convergence declared due to random chance

Convergence

In MCEM framework, there are 2 complications to account for

1. spurious convergence declared due to random chance

   $\Rightarrow$ **Solution**: require convergence for 3 iterations in succession

Convergence

In MCEM framework, there are 2 complications to account for

1. spurious convergence declared due to random chance
   $\Rightarrow$ **Solution**: require convergence for 3 iterations in succession
2. estimators swamped by Monte Carlo error, thus precluding convergence

Convergence

In MCEM framework, there are 2 complications to account for

1. spurious convergence declared due to random chance

   $\Rightarrow$ **Solution**: require convergence for 3 iterations in succession

2. estimators swamped by Monte Carlo error, thus precluding convergence

   $\Rightarrow$ **Solution**: increase Monte Carlo size N as algorithm moves closer towards maximizer

Convergence

- Using large N when far from maximizer = computationally inefficient
- Using small N when close to maximizer = unlikely to detect convergence

Solution (proposed by Ripatti et al. 2002): after a 'burn-in' phase, calculate the *coefficient of variation* statistic

$$\text{cv}(\Delta_{\text{rel}}^{(m+1)}) = \frac{\text{sd}(\Delta_{\text{rel}}^{(m-1)}, \Delta_{\text{rel}}^{(m)}, \Delta_{\text{rel}}^{(m+1)})}{\text{mean}(\Delta_{\text{rel}}^{(m-1)}, \Delta_{\text{rel}}^{(m)}, \Delta_{\text{rel}}^{(m+1)})},$$

and increase N to $N + \lfloor N/\delta \rfloor$ if $\text{cv}(\Delta_{\text{rel}}^{(m+1)}) > \text{cv}(\Delta_{\text{rel}}^{(m)})$ for some small positive integer δ

Standard error estimation

UNIVERSITY OF LIVERPOOL

There are two approaches available:

Standard error estimation

There are two approaches available:

1. Bootstrap estimator

Hsieh et al. (2006) demonstrated that the profile likelihood approach in the EM algorithm leads to underestimation in the SEs, so recommended bootstrapping:

1. sample n subjects with replacement and re-label with indices $i' = 1, \ldots, n$
2. re-fit the model to the bootstrap-sampled dataset
3. repeat steps 1 and 2 B-times, for each iteration extracting the model parameter estimates for $(\beta^\top, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \gamma_v^\top, \gamma_y^\top)$
4. calculate SEs of B sets of estimates

Standard error estimation

There are two approaches available:

2. Empirical information matrix approximation

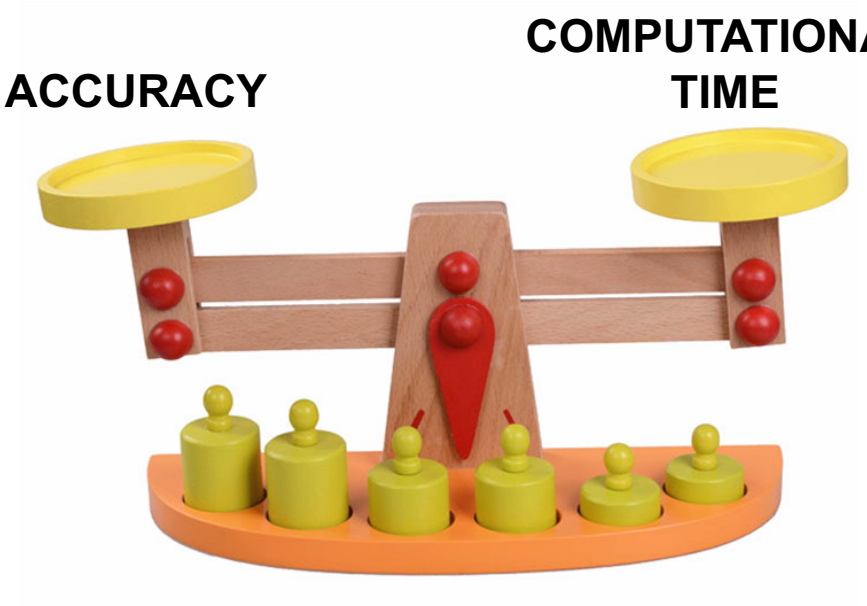
Following McLachlan and Krishnan (2008), $\text{SE}(\theta) \approx I_e^{-1/2}(\hat{\theta})$, where

$$I_e(\theta) = \sum_{i=1}^{n} s_i(\theta) s_i^{\top}(\theta) - \frac{1}{n} S(\theta) S^{\top}(\theta),$$

$S(\theta) = \sum_{i=1}^{n} s_i(\theta)$ is the score vector

NB. SEs only calculated for $\theta_{-\lambda_0(t)}$, as profile likelihood arguments are used

Standard error estimation



Bootstrap *versus* approximate information matrix

Simulation study set-up

UNIVERSITY OF LIVERPOOL

- 200 simulations of $n = 250$ / 500 patients
- Planned measurement of 2 biomarkers at 0, 1, 2, 3, 4, and 5 years; mean $= 4.2$ measurements
- Random-intercepts and random slopes simulated from $N_4(0, D)$
- Followed until 6-years with event time simulated from Gompertz PH model with shape $= 0.25$ and scale $= \exp(-3.5) \Rightarrow$ event rate $\approx 46\%$ at 5-years
- Independent censoring time from exponential distribution with scale $= \exp(-3) \Rightarrow \approx 19\%$ censored before end of follow-up
- 1 $N(0, 1)$ continuous covariate, and 1 Bernoulli(0.5) binary covariate

Results $n = 250$

	True value	Mean estimate	Empirical SE	Mean SE	Bias	MSE	Coverage[2]
			Longitudinal sub-model 1				
$\texttt{(Intercept)}_1$	0.0000	0.0002	0.0605	0.0582	0.0002	0.0037	0.9350
$\texttt{time}_1$	1.0000	0.9982	0.0187	0.0197	-0.0018	0.0004	0.9750
$\texttt{ctsxl}_1$	1.0000	0.9964	0.0381	0.0416	-0.0036	0.0015	0.9600
$\texttt{binxl}_1$	1.0000	1.0005	0.0810	0.0821	0.0005	0.0066	0.9350
			Longitudinal sub-model 2				
$\texttt{(Intercept)}_2$	0.0000	0.0033	0.0554	0.0577	0.0033	0.0031	0.9550
$\texttt{time}_2$	-1.0000	-0.9996	0.0173	0.0191	0.0004	0.0003	0.9850
$\texttt{ctsxl}_2$	0.0000	-0.0004	0.0409	0.0415	-0.0004	0.0017	0.9450
$\texttt{binxl}_2$	0.5000	0.4975	0.0801	0.0815	-0.0025	0.0064	0.9500
			Time-to-event sub-model				
ctsx	0.0000	-0.0034	0.1188	0.1173	-0.0034	0.0141	0.9350
binx	1.0000	1.0228	0.2387	0.2301	0.0228	0.0575	0.9400
γ_1	-0.5000	-0.5243	0.1348	0.1540	-0.0243	0.0188	0.9800
γ_2	1.0000	1.0109	0.1585	0.1675	0.0109	0.0253	0.9650

[2] Mean SEs and coverage calculated using empirical information approximation

Results $n = 500$

	True value	Mean estimate	Empirical SE	Mean SE	Bias	MSE	Coverage[3]
			Longitudinal sub-model 1				
$\texttt{(Intercept)}_1$	0.0000	-0.0022	0.0376	0.0402	-0.0022	0.0014	0.9600
$\texttt{time}_1$	1.0000	1.0001	0.0129	0.0137	0.0001	0.0002	0.9750
$\texttt{ctsxl}_1$	1.0000	0.9959	0.0243	0.0285	-0.0041	0.0006	0.9700
$\texttt{binxl}_1$	1.0000	1.0045	0.0527	0.0564	0.0045	0.0028	0.9600
			Longitudinal sub-model 2				
$\texttt{(Intercept)}_2$	0.0000	0.0017	0.0352	0.0400	0.0017	0.0012	0.9600
$\texttt{time}_2$	-1.0000	-0.9992	0.0135	0.0131	0.0008	0.0002	0.9350
$\texttt{ctsxl}_2$	0.0000	0.0013	0.0269	0.0284	0.0013	0.0007	0.9500
$\texttt{binxl}_2$	0.5000	0.4973	0.0526	0.0563	-0.0027	0.0028	0.9750
			Time-to-event sub-model				
ctsx	0.0000	0.0104	0.0791	0.0789	0.0104	0.0064	0.9550
binx	1.0000	0.9952	0.1627	0.1571	-0.0048	0.0265	0.9300
γ_1	-0.5000	-0.4976	0.0987	0.1006	0.0024	0.0098	0.9700
γ_2	1.0000	1.0061	0.1045	0.1091	0.0061	0.0109	0.9500

[3] Mean SEs and coverage calculated using empirical information approximation

joineRML

- An R package is now available for fitting this model: joineRML
- Currently on GitHub (due for CRAN submission shortly): github.com/graemeleehickey/joineRML
- Complements existing R package for univariate joint models: joineR (available on CRAN)

Example code

UNIVERSITY OF LIVERPOOL

```
library(joineRML)
data(pbc2)

fit.pbc <- mjoint(
  formLongFixed = list("bilirubin" = log.b ~ year + drug,
                       "albumin" = log.a ~ year),
  formLongRandom = list("bilirubin" = ~ year | id,
                        "albumin" = ~ 1 | id),
  formSurv = Surv(years, status2) ~ age + drug,
  data = pbc2,
  timeVar = "year",
  control = list(convCrit = "sas", tol0 = 0.002, tol2 = 0.002),
  inits = list(gamma = gamma.inits),
  verbose = TRUE)

summary(fit.pbc)
```

Alternative options

- Pre-2016: none!
- 2016-onwards (all still at development stage):
  - `stjm`: a new extension to the Stata package[4] written by Michael Crowther
  - `rstanjm`: a new R package[5] that utilises the Bayesian package Stan written by Sam Brilleman
  - `JMbayes`: a new extension[6] to the R package written by Dimitris Rizopoulos

[4] Crowther MJ. *Joint Statistical Meeting.* Seattle; 2015.

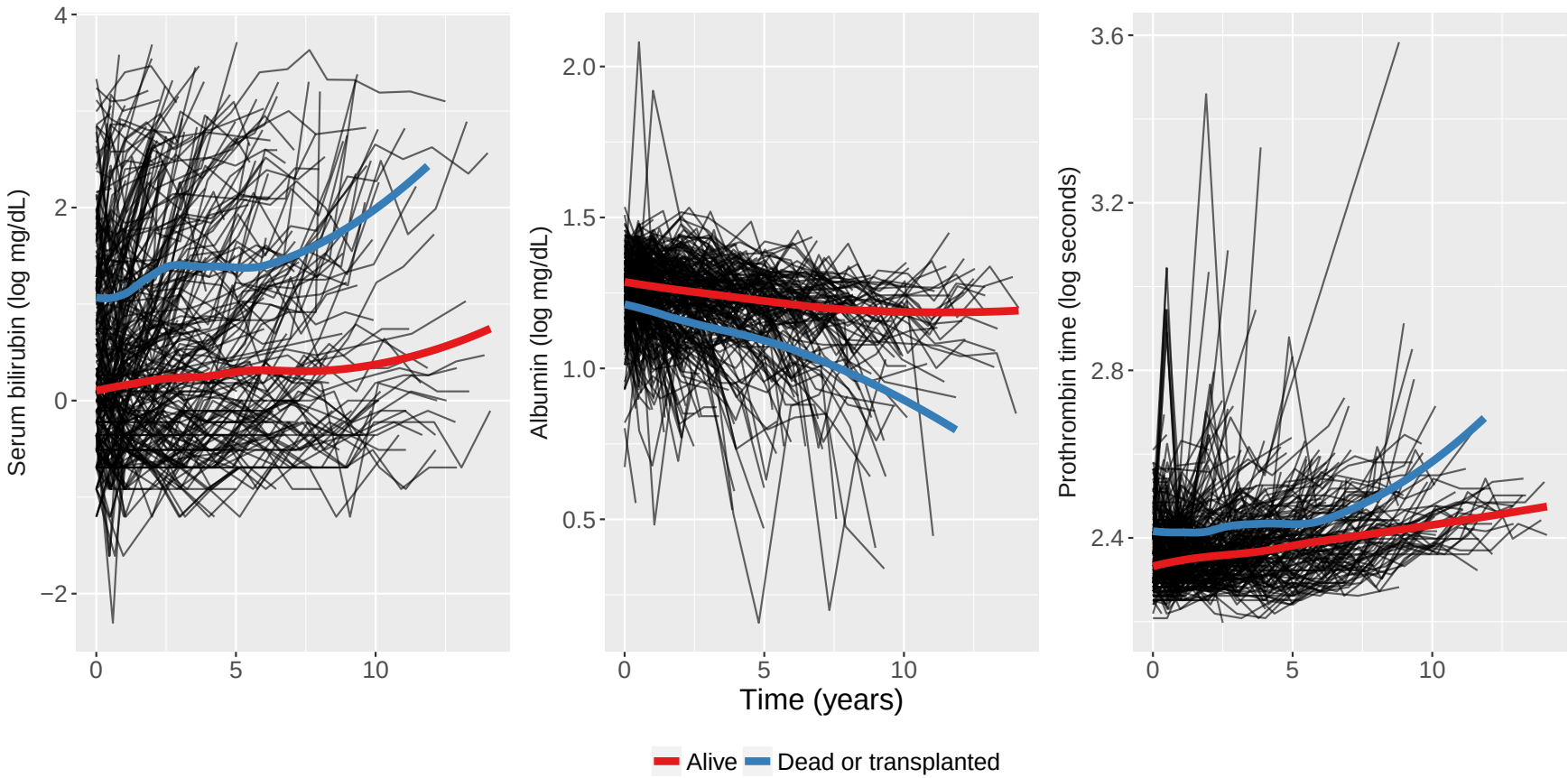
[5] `github.com/sambrilleman/rstanjm`

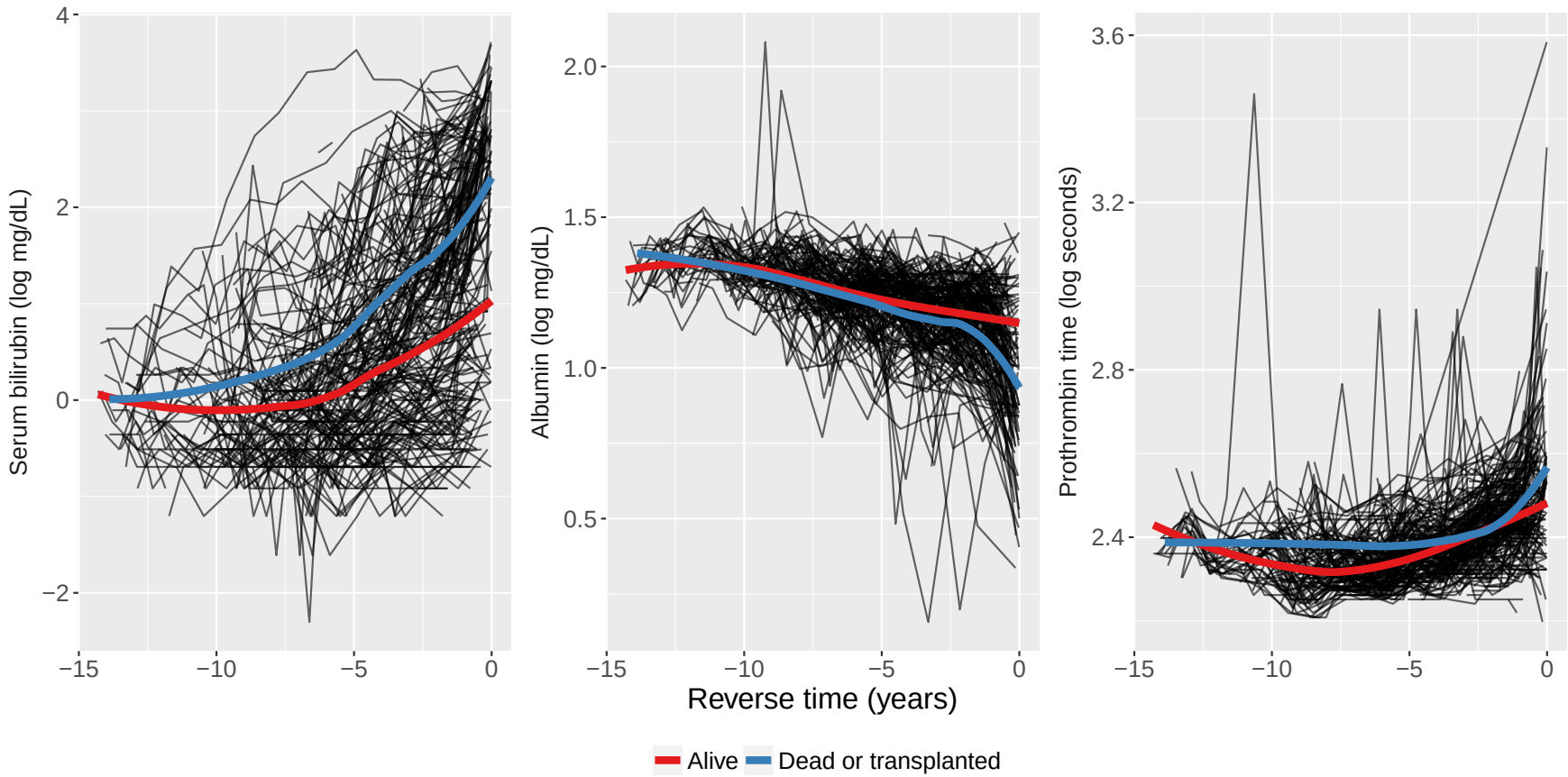
[6] `github.com/drizopoulos/JMbayes`

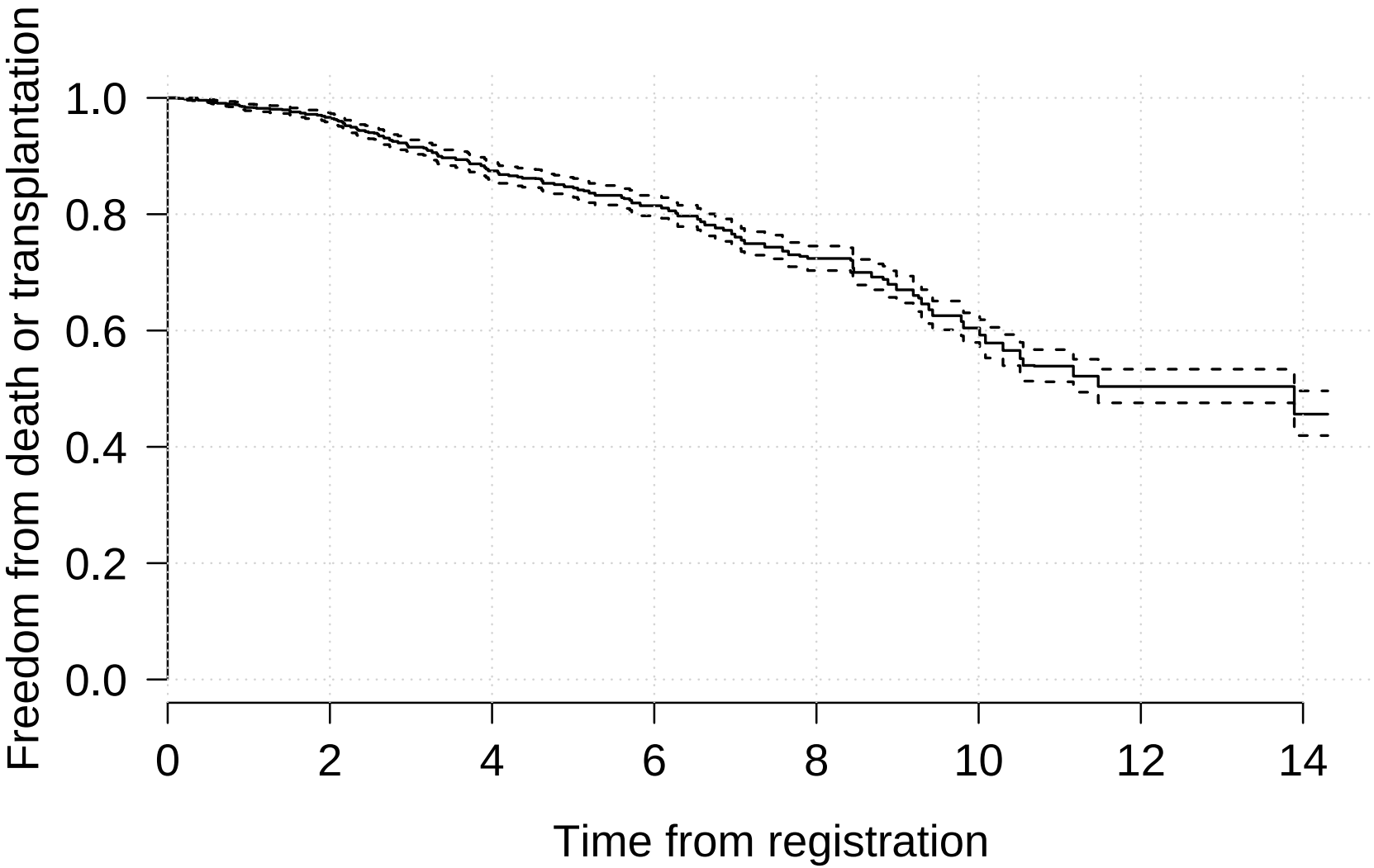
The Mayo Clinic PBC data

UNIVERSITY OF LIVERPOOL

- Primary biliary cirrhosis (PBC) is a chronic liver disease characterized by inflammatory destruction of the small bile ducts, which eventually leads to cirrhosis of the liver (Murtaugh et al. 1994)
- Trial conducted between 1974 and 1984 randomized 312 patients to either placebo or D-penicillamine
- Multiple biomarkers repeatedly measured at intermittent times:
  1. serum bilirunbin (mg/dl)
  2. serum albumin (mg/dl)
  3. prothrombin time (seconds)
- Time to death or transplantation (competing risks)


Serum bilirubin (log mg/dL)
4
2
0
−2
0
5
10
Albumin (log mg/dL)
2.0
1.5
1.0
0.5
0
5
10
Time (years)
Prothrombin time (log seconds)
3.6
3.2
2.8
2.4
0
5
10
Alive
Dead or transplanted


Serum bilirubin (log mg/dL)
Albumin (log mg/dL)
Prothrombin time (log seconds)
Reverse time (years)
Alive
Dead or transplanted


Freedom from death or transplantation
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
14
Time from registration

Proposed model

Longitudinal sub-model

$$
\begin{aligned}
\log(\texttt{bil}) &= (\beta_{0,1} + b_{0i,1}) + (\beta_{1,1} + b_{1i,1})\texttt{year} + \varepsilon_{ij1}, \\
\log(\texttt{alb}) &= (\beta_{0,2} + b_{0i,2}) + (\beta_{1,2} + b_{1i,2})\texttt{year} + \varepsilon_{ij2}, \\
\log(\texttt{pro}) &= (\beta_{0,3} + b_{0i,3}) + (\beta_{1,3} + b_{1i,3})\texttt{year} + \beta_{2,3}(\texttt{year}/10)^2 + \varepsilon_{ij3}, \\
b_i &\sim N_6(0, D), \text{ and } \varepsilon_{ijk} \sim N(0, \sigma_k^2) \text{ for } k = 1, 2, 3
\end{aligned}
$$

Time-to-event sub-model

$$
\begin{aligned}
\lambda_i(t) &= \lambda_0(t) \exp\{\gamma_v \text{age} + W_{2i}(t)\} \\
W_{2i}(t) &= \gamma_{\texttt{bil}}(b_{0i,1} + b_{1i,1}t) + \gamma_{\texttt{alb}}(b_{0i,2} + b_{1i,2}t) + \gamma_{\texttt{pro}}(b_{0i,3} + b_{1i,3}t)
\end{aligned}
$$

Results

Longitudinal sub-model

Biomarker		**Estimate**	**SE**	***P***
log(bilirubin)	`(Intercept)`	0.4841	0.0536	< 0.0001
	`year`	0.2008	0.0131	< 0.0001
log(albumin)	`(Intercept)`	1.2620	0.0074	< 0.0001
	`year`	-0.0382	0.0021	< 0.0001
log(prothrombin)	`(Intercept)`	2.3695	0.0060	< 0.0001
	`year`	0.0100	0.0027	0.0002
	`I((year/10)^2)`	0.2428	0.0287	< 0.0001

Results

Time-to-event sub-model[7]

	Estimate	**SE**	***P***
age	0.0462	0.0089	< 0.0001
$\gamma_{\texttt{bil}}$	0.9862	0.1381	< 0.0001
$\gamma_{\texttt{alb}}$	-4.6996	1.0007	< 0.0001
$\gamma_{\texttt{pro}}$	3.0901	1.7779	0.0822

[7] γ parameters were initialized at their separate univariate joint model estimates

Future research

- Develop `joineRML` package to be faster and more accurate
- Extend to include competing risks and recurrent events; e.g. Williamson et al. (2008)
- Incorporate model diagnostics; e.g. residuals

Acknowledgements

- **Project investigators:**
  - Dr. Ruwanthi Kolamunnage-Dona (University of Liverpool)
  - Dr. Pete Philipson (University of Northumbria)
  - Dr. Andrea Jorgensen (University of Liverpool)
- **Statistical collaborators:**
  - Prof. Robin Henderson (University of Newcastle)
  - Prof. Paula Williamson (University of Liverpool)
- **Funding:** Medical Research Council (MR/M013227/1)

References I


Hickey, GL et al. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Meth* 16(1), pp. 1–15.


Laird, NM and Ware, JH (1982). Random-effects models for longitudinal data. *Biometrics* 38(4), pp. 963–74.


Henderson, R et al. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1(4), pp. 465–480.


Wulfsohn, MS and Tsiatis, AA (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* 53(1), pp. 330–339.


Lin, H et al. (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med* 21, pp. 2369–2382.


Dempster, AP et al. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B* 39(1), pp. 1–38.

References II


Wei, GC and Tanner, MA (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J Am Stat Assoc* 85(411), pp. 699–704.


Ripatti, S et al. (2002). Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. *Life Dat Anal* 8(2002), pp. 349–360.


Hsieh, F et al. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* 62(4), pp. 1037–1043.


McLachlan, GJ and Krishnan, T (2008). The EM Algorithm and Extensions. Second. Wiley-Interscience.


Murtaugh, Paul A et al. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology* 20(1), pp. 126–134.


Williamson, Paula R et al. (2008). Joint modelling of longitudinal and competing risks data. *Statistics in Medicine* 27, pp. 6426–6438.