Joint modelling of longitudinal and time-to-event data with R

Graeme L. Hickey

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

5th July 2017
An MRC funded workshop
<table>
<thead>
<tr>
<th>Time</th>
<th>What</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>0900 – 0920</td>
<td>Introduction</td>
<td>G. L. Hickey</td>
</tr>
<tr>
<td>0920 – 0940</td>
<td>Univariate joint models</td>
<td>P. Philipson</td>
</tr>
<tr>
<td>0940 – 1000</td>
<td>Competing risks</td>
<td>R. Kolamunnage-Dona</td>
</tr>
<tr>
<td>1000 – 1025</td>
<td>Multivariate outcomes</td>
<td>G. L. Hickey</td>
</tr>
<tr>
<td>1025 – 1030</td>
<td>Summary</td>
<td>G. L. Hickey</td>
</tr>
<tr>
<td>1030 – 1045</td>
<td>Tea and coffee break</td>
<td></td>
</tr>
<tr>
<td>1045 – 1100</td>
<td>Overview of R packages</td>
<td>G. L. Hickey</td>
</tr>
<tr>
<td>1100 – 1230</td>
<td>Practical worksheet &amp; problems sheet</td>
<td></td>
</tr>
<tr>
<td>1230 – 1300</td>
<td>Lunch</td>
<td></td>
</tr>
</tbody>
</table>
References are given in the bibliography (slide 95 – onwards)

Time for questions during the break and lunch sessions\(^1\)

\(^1\)Please find myself, Ruwanthi, Pete, Maria, or Andrea
Introduction
Motivation

Many scientific investigations follow-up subjects repeatedly over time and generate both

**Longitudinal measurement data**
Repeated measurements of a response variable at a number of time points, e.g.
- biomarker measurements
- quality of life assessments

**Event history data**
Time(s) to (possibly recurrent or) terminating events, e.g.
- time to death
- time to study withdrawal / dropout
Schizophrenia trial data

- A placebo-controlled randomized clinical trial of drug treatments for schizophrenia (Henderson et al. 2000)
- In the original trial there were three treatment arms: placebo, standard, and novel compound (subdivided into 4 dosages)
- We will analyse a sample of 150 patients (50 patients per treatment arm)
Schizophrenia trial data

- The repeated measurement outcome was a questionnaire-based measure, the positive and negative symptom score (PANSS) — a measure of psychiatric disorder — taken at weeks 0, 1, 2, 4, 6 and 8 following randomization.

- The time-to-event outcome was defined as withdrawal from the study due to ‘inadequate response’.

- Failure to complete the trial protocol for other reasons, unrelated to the subject’s mental state, are regarded here as un informatively missing values, rather than as dropouts.
Schizophrenia trial data

Number of patients with missing PANSS at each measurement time

<table>
<thead>
<tr>
<th>Treatment</th>
<th>$T = 0$</th>
<th>$T = 1$</th>
<th>$T = 2$</th>
<th>$T = 4$</th>
<th>$T = 6$</th>
<th>$T = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>20</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>10</td>
<td>22</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>12</td>
<td>17</td>
<td>23</td>
</tr>
</tbody>
</table>
Depending of the focus of the research question, we might use either

- *Separate analyses* of one or more of the outcomes
- *Joint analysis* of the different outcomes
Separate analyses

- Is the average PANSS score trajectory different between treatment groups?
- Does patient dropout differ between treatment group?
Separate analyses: longitudinal outcomes
Separate analyses: longitudinal outcomes

Linear mixed effects model with treatment $X_i \in \{1, 2, 3\}$

$$Y_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + \beta_2 I(X_i = 2)t + \beta_3 I(X_i = 3)t + \varepsilon_i(t)$$

$$ (b_{i0}, b_{i1})^\top \sim N_2(0, D)$$

$$\varepsilon_i(t) \overset{i.i.d.}{\sim} N(0, \sigma^2)$$
Separate analyses: longitudinal outcomes

Linear mixed effects model with treatment $X_i \in \{1, 2, 3\}$

$$Y_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + \beta_2 I(X_i = 2)t + \beta_3 I(X_i = 3)t + \varepsilon_i(t)$$

$$(b_{i0}, b_{i1})^\top \sim N_2(0, D)$$

$$\varepsilon_i(t) \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

Fitting the model in R

Using `nlme::lme()`\(^2\), we get:

$\hat{\beta}_2 = -1.4$ (SE = 0.46) and $\hat{\beta}_3 = -2.1$ (SE = 0.45)

\(^2\)See Pinheiro and Bates (2000)
Separate analyses: time-to-event outcomes
Separate analyses: time-to-event outcomes

Cox proportional hazards model with treatment $X_i \in \{1, 2, 3\}$

$$
\lambda_i(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} = \lambda_0(t) \exp\{\gamma_2 I(X_i = 2) + \gamma_3 I(X_i = 3)\}
$$

with $\lambda_0(t)$ left unspecified
Cox proportional hazards model with treatment $X_i \in \{1, 2, 3\}$

$$\lambda_i(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt}$$

$$= \lambda_0(t) \exp\{\gamma_2 I(X_i = 2) + \gamma_3 I(X_i = 3)\}$$

with $\lambda_0(t)$ left unspecified

Fitting the model in R

Using `survival::coxph()`\(^3\) we get:

$\hat{\gamma}_2 = -0.59$ (SE = 0.29) and $\hat{\gamma}_3 = -0.99$ (SE = 0.33)

\(^3\)See Therneau and Grambsch (2000)
It is not clear whether the apparent decrease in PANSS profiles reflects

1. a genuine change over time, or
2. an artefact caused by selective drop-out, with patients with high (worse) PANSS values being less likely to complete the trial
In general, the primary focus for inference may be on:

1. **Improving inference for a repeated measurement outcome** subject to an informative dropout mechanism that is not of direct interest

2. **Improving inference for a time-to-event outcome**, whilst taking account of an intermittently and error-prone measured endogenous time-dependent variable

3. **Studying the relationship between the two correlated processes**

In the previous slide, our focus was of type 1
Joint analysis

Interest might also be about

1. **Prediction**\(^4\): given the measurement data for a patient up to time \( t \), what is their survivorship distribution at time \( s > t \)\

2. **Surrogacy**\(^5\): when the clinical event is rare or takes a long time to reach, we want to use more easily measured markers as a substitute, which can usually lead to substantial sample size reduction and shorter trial duration. In other words, is \( [T \mid X, Y] = [T \mid Y]? \)

---

\(^4\) e.g. Andrinopoulou et al. (2015)
\(^5\) e.g. Xu and Zeger (2001)
What methods can we use to answer these sorts of questions?

- **Separate analyses:** simply ignore the association
- **Naive models:** Cox regression with time-varying covariate(s) assuming last-value carried forward
- **Two-stage models:** the LMM is first fitted separately, and the best linear unbiased predictors (BLUPS) of the random effects extracted and included in a time-varying covariate survival model
- **Pattern mixture or selection models:** factorize the joint distribution as either \([Y, T] = [T] \times [Y | T] = [Y] \times [T | Y]\)
- **Landmarking:** refitting the survival model at sequential follow-up times conditional on subjects still in the risk set
- **Joint models:** focus of today’s workshop
Software implementations in R

- Advanced statistical methods have limited use if not readily and publicly available in software
- R is a free software environment for statistical computing and graphics
- It compiles and runs on a wide variety of UNIX platforms, Windows, and macOS
### Available packages in R

<table>
<thead>
<tr>
<th>Package</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>joineR</td>
<td>Fit models from Henderson et al. (2000) and Williamson et al. (2008)</td>
</tr>
<tr>
<td>joineRML</td>
<td>Extension of joineR to multivariate longitudinal data</td>
</tr>
<tr>
<td>JM</td>
<td>Fits array of univariate joint models with parametric, spline and semi-parametric survival models + flexible association structure + prediction tools</td>
</tr>
<tr>
<td>JMbayes</td>
<td>Bayesian version of JM + non-continuous outcomes + development version can model multivariate longitudinal data</td>
</tr>
<tr>
<td>JSM</td>
<td>Fits semi-parametric joint models with nonparametric multiplicative random effects and with transformation models</td>
</tr>
<tr>
<td>lcmm</td>
<td>Fits joint latent class mixed models</td>
</tr>
<tr>
<td>frailtypack</td>
<td>Joint models with or without recurrent events data</td>
</tr>
<tr>
<td>JMdesign</td>
<td>Power / sample size calculations</td>
</tr>
<tr>
<td>rstanarm</td>
<td>Fits range of models similar to JMbayes by exploiting Stan</td>
</tr>
<tr>
<td>JMboost</td>
<td>Implements a boosting algorithm for fitting joint models to potentially high-dimensional data</td>
</tr>
</tbody>
</table>
# Available packages in other software

<table>
<thead>
<tr>
<th>Package</th>
<th>Platform</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>stjm</td>
<td>Stata</td>
<td>Fits array of univariate (+ multivariate in future) joint models with parametric survival models + flexible association structure + prediction tools</td>
</tr>
<tr>
<td>JMFit</td>
<td>SAS</td>
<td>Fits simple random-effects parameterisation models</td>
</tr>
<tr>
<td>%JM</td>
<td>SAS</td>
<td>Similar to R package JM</td>
</tr>
<tr>
<td>jmxtstcox</td>
<td>Stata</td>
<td>Fits semi-parametric joint models with shared random intercepts only</td>
</tr>
</tbody>
</table>

+ many application-specific user-written scripts (WinBUGS, C, etc.) published as appendices in literature
Learning objectives

1. Recognise what problems joint models can be used for
2. Understand how joint models can be fitted (in frequentist framework)
3. Fit joint models in R independently
4. Interpret joint models for inference

Everything discussed today stems from our own research and actual clinical questions using real clinical data
Univariate joint models
A brief history

- Originally motivated by AIDS research in which a biomarker such as CD4 lymphocyte count is determined intermittently and its relationship with time to seroconversion or death is of interest.
- Seminal articles by Wulfsohn and Tsiatis (1997) and Henderson et al. (2000) contributed to a rapidly growing research field.
- As of 2015: > 200 methodological papers and > 60 clinical applications of joint models.
Conceptual principle
Conceptual principle

- Covariate data
- Time-to-event outcomes data
- Frailties
Conceptual principle

- Covariate data
- Time-to-event outcomes data
- Longitudinal outcomes data
- Random effects
- Frailties

\[ \beta \]
\[ \gamma_y \]
Conceptual principle

Covariate data → \( \beta \) → Longitudinal outcomes data → Frailties

Covariate data → \( \gamma_y \) → Time-to-event outcomes data → Random effects

\( \beta \) and \( \gamma \) represent the parameters or coefficients in the model, where \( \beta \) connects the covariate data to the longitudinal outcomes, and \( \gamma \) connects the longitudinal outcomes to the frailties. The model integrates the longitudinal outcomes, time-to-event outcomes, and frailties, which are crucial for understanding the underlying processes in complex data structures.
Data

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

- An $n_i$-vector of observed **longitudinal measurements** for the longitudinal outcome: $y_i = (y_{i1}, \ldots, y_{in_i})^T$

- **Observation times** $t_{ij}$ for $j = 1, \ldots, n_i$, which can differ between subjects

- $(T_i, \delta_i)$, 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 $\delta_i$ is the failure indicator equal to 1 if the failure is observed ($T_i^* \leq C_i$) and 0 otherwise
Longitudinal outcome sub-model

\[ y_i(t) = \mu_i(t) + W_{1i}(t) + \varepsilon_i(t), \]

where

- \( \varepsilon_i(t) \) is the model error term, which is i.i.d. \( N(0, \sigma^2) \) and independent of \( W_{1i}(t) \)
- \( \mu_i(t) = x_i^\top(t) \beta \) is the mean response
- \( x_i(t) \) is a \( p \)-vector of (possibly) time-varying covariates with corresponding fixed effect terms \( \beta \)
- \( W_{1i}(t) \) is a zero-mean latent Gaussian process
Time to event outcome 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 $\gamma_v$
- $W_{2i}(t)$ is a zero-mean latent Gaussian process, independent of the censoring process
Association structure

**$W_1(t)$**

Following Laird and Ware (1982):

$$W_{1i}(t) = z_i^T(t)b_i,$$

with $b_i \sim N(0, D)$
Association structure

\[ W_1(t) \]

Following Laird and Ware (1982):

\[ W_{1i}(t) = z_i^T(t)b_i, \]

with \( b_i \sim N(0, D) \)

\[ W_2(t) \]

Following Henderson et al. (2000)

\[ W_{2i}(t) = \gamma_y W_{1i}(t), \]

with \( \gamma_y \) a scalar parameter for estimation
Henderson et al. (2000) proposed several extensions:

1. $W_1(t) = z_i^\top(t)b_i + V_{1i}(t)$, where $V_{1i}(t)$ is a stationary Gaussian process with zero mean, variance $\sigma^2_{V_1}$, and correlation function $r_1(u) = \text{cov} \{V_{1i}(t), V_{1i}(t - u)\} / \sigma^2_{V_1}$

2. $W_2(t) = \gamma_{y1}^\top b_i + \gamma_{y2} W_{1i}(t) + \theta_i$, where $\gamma_y = (\gamma_{y1}, \gamma_{y2})^\top$ is a vector of association parameters, and $\theta_i$ is a frailty term independent from the longitudinal data
Association structure: alternatives

Many other proposals for association structures in the literature:

- **Current value parameterisation:** \( W_{2i}(t) = \gamma_y \{ \mu_i(t) + W_{1i}(t) \} \)
- **Random effects parameterisation:** \( W_{2i}(t) = \gamma_{y1}^\top b_i \)
- **Bivariate distribution:** \( (W_{1i}, W_{2i}) \sim N(0, \Omega) \)
- **Random-slopes parameterisation:**
  \[
  W_{2i}(t) = \gamma_{y1} \{ \mu_i(t) + W_{1i}(t) \} + \gamma_{y2} \frac{\partial}{\partial t} \{ \mu_i(t) + W_{1i}(t) \}
  \]
- \( \ldots \)
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 $\theta = (\beta^T, \text{vech}(D), \sigma^2, \lambda_0(t), \gamma_v^T, \gamma_y)$
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 $\theta = (\beta^T, \text{vech}(D), \sigma^2, \lambda_0(t), \gamma_v^T, \gamma_y)$, and

$$f(y_i \mid b_i, \theta) = (2\pi)^{-\frac{n_i}{2}} \sigma^{-n_i} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - X_i\beta - Z_i b_i)^T (y_i - X_i\beta - Z_i b_i) \right\}$$
The observed data likelihood is given by

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

where \(\theta = (\beta^T, \text{vech}(D), \sigma^2, \lambda_0(t), \gamma_v^T, \gamma_y)\), and

\[
f(T_i, \delta_i | b_i; \theta) = \left[ \lambda_0(T_i) \exp \left\{ \gamma_v^T \gamma_v + W_{2i}(T_i, b_i) \right\} \right]^\delta_i \\
\exp \left\{ - \int_0^{T_i} \lambda_0(u) \exp \left\{ \gamma_v^T \gamma_v + W_{2i}(u, b_i) \right\} \, du \right\}
\]
The *observed* data likelihood is given by

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

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

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

with \( r = \text{dim}(b_i) \)
Multiple approaches have been considered over the years:

- Markov chain Monte Carlo (MCMC)
- Direct likelihood maximisation (e.g. Newton-methods)
- Generalised estimating equations
- EM algorithm (treating the random effects as missing data)
- ...
EM algorithm (Dempster et al. 1977)

**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.

\[
Q(\theta \mid \hat{\theta}^{(m)}) = \sum_{i=1}^{n} \mathbb{E}\left\{ \log f(y_i, T_i, \delta_i, b_i \mid \theta) \right\},
\]

\[
= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left\{ \log f(y_i, T_i, \delta_i, b_i \mid \theta) \right\} f(b_i \mid T_i, \delta_i, y_i; \hat{\theta}^{(m)}) db_i
\]
**EM algorithm (Dempster et al. 1977)**

**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.

$$Q(\theta | \hat{\theta}^{(m)}) = \sum_{i=1}^{n} \mathbb{E}\left\{ \log f(y_i, T_i, \delta_i, b_i | \theta) \right\},$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left\{ \log f(y_i, T_i, \delta_i, b_i | \theta) \right\} f(b_i | T_i, \delta_i, y_i; \hat{\theta}^{(m)}) db_i$$

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

$$\hat{\theta}^{(m+1)} = \arg \max_{\theta} Q(\theta | \hat{\theta}^{(m)})$$
M-step: closed form estimators

\[
\hat{\lambda}_0(t) = \frac{\sum_{i=1}^{n} \delta_i I(T_i = t)}{\sum_{i=1}^{n} \mathbb{E} \left[ \exp \left\{ \mathbf{v}_i^\top \gamma_v + W_{2i}(t, b_i) \right\} \right] I(T_i \geq t)}
\]

\[
\hat{\beta} = \left( \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^{n} \mathbf{X}_i^\top (y_i - Z_i \mathbb{E}[b_i]) \right)
\]

\[
\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \left\{ (y_i - X_i \hat{\beta})^\top (y_i - X_i \hat{\beta} - 2Z_i \mathbb{E}[b_i]) \right. + \left. \text{trace} \left( Z_i^\top Z_i \mathbb{E}[b_i b_i^\top] \right) \right\}
\]

\[
\hat{D} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ b_i b_i^\top \right]
\]
There is no closed form update for \( \gamma = (\gamma_v^\top, \gamma_y)^\top \), so use a one-step Newton-Raphson iteration

\[
\hat{\gamma}^{(m+1)} = \hat{\gamma}^{(m)} + I(\hat{\gamma}^{(m)})^{-1} S(\hat{\gamma}^{(m)}), \text{ where}
\]

\[
S(\gamma) = \sum_{i=1}^{n} \left[ \delta_i \mathbb{E}[\tilde{v}_i(T_i)] - \int_0^{T_i} \lambda_0(u) \mathbb{E}[\tilde{v}_i(u) \exp{\{\tilde{v}_i^\top(u)\gamma}\}] \, du \right]
\]

\[
I(\gamma) = -\frac{\partial}{\partial \gamma} S(\gamma)
\]

with \( \tilde{v}_i(t) = (v_i^\top, z_i^\top(t)b_i)^\top \) a \((q + 1)\)-vector\(^6\)

\(^6\)Generalises to a \((q + |\gamma_y|)\)-vector if \(\gamma_y\) is a vector
E-step

Need to calculate several multi-dimensional expectations of the form

$$\mathbb{E}[h(b_i)] = \int \int \cdots \int_{-\infty}^{\infty} h(b_i) f(b_i \mid T_i, \delta_i, y_i, \hat{\theta}) \, db_i$$
E-step

Need to calculate several multi-dimensional expectations of the form

\[ \mathbb{E} [h(b_i)] = \int \cdots \int_{-\infty}^{\infty} h(b_i) f(b_i \mid T_i, \delta_i, y_i, \hat{\theta}) \, db_i \]
E-step

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

where

\[
h(\cdot) = \text{any known function},
\]

\[
b_i \mid y_i, \theta \sim N \left( A_i \left\{ \sigma^{-2} Z_i^\top (y_i - X_i \beta) \right\}, A_i \right), \quad \text{and}
\]

\[
A_i = \left( \sigma^{-2} Z_i^\top Z_i + D^{-1} \right)^{-1}
\]
E-step

- Approach used by Wulfsohn and Tsiatis (1997) and joineR software is Gauss-Hermite quadrature:

\[
\int_{-\infty}^{\infty} e^{-\phi^2} f(\phi) d\phi \approx \sum_{j=1}^{p} w_j f(x_j),
\]

where \( x_j \) \((j = 1, \ldots, p)\) are tabulated abscissa values for \( \phi \), with corresponding weights \( w_j \)

- Several studies have indicated \( p = 3 \) or \( p = 4 \) yields reasonable accuracy

- More accurate approximations have been developed in recent years, e.g. adaptive-GH quadrature
Standard errors

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

- Conceptually simple:
  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 $(\beta^T, \text{vech}(D), \sigma^2, \gamma_v^T, \gamma_y)$
  4. calculate SEs of $B$ sets of estimates

- Notice no SEs calculated for $\lambda_0(t)$, but that’s generally not of inferential interest
We can implement all of this in the R package joineR, with the general recipe:

1. Create a jointdata object using the joineR::jointdata() function
2. Fit a joint model using the joineR::joint() function
3. Calculate SEs using joineR::jointSE() function

Each piece can be used in separate auxiliary functions along the way
Recall question from slide 16:

**Question**

Does the decrease in PANSS profiles reflects a genuine change over time or an artefact caused by selective dropout

⇒ Let’s fit a joint model between longitudinal PANSS score and time-to-dropout
Schizophrenia trial data

Longitudinal sub-model

\[ Y_i(t) = \beta_0 + \beta_1 t + \sum_{k=2}^{3} \beta_k I(X_i = k)t + W_{1i}(t) + \varepsilon_i(t) \]

\[ W_{1i}(t) = b_{i0} + b_{i1} t \]

\[ (b_{i0}, b_{i1})^\top \sim N_2(0, D) \]

\[ \varepsilon_i(t) \sim N(0, \sigma^2) \]

Time-to-event sub-model

\[ \lambda_i(t) = \lambda_0(t) \exp \left\{ \sum_{k=2}^{3} \gamma_k I(X_i = k) + W_{2i}(t) \right\} \]

\[ W_{2i}(t) = \gamma_y W_{1i}(t) \]
Example code

```r
# Set-up the data in a jointdata object
data(mental)
mental.unbalanced <- to.unbalanced(mental, id.col = 1, 
    times = c(0, 1, 2, 4, 6, 8), 
    Y.col = 2:7, other.col = 8:11)
names(mental.unbalanced)[3] <- "Y"
mental.long <- mental.unbalanced[, 1:3]
mental.surv <- UniqueVariables(mental.unbalanced, 6:7, id.col = 1)
mental.baseline <- UniqueVariables(mental.unbalanced, 4, id.col = 1)
mental.jd <- jointdata(
    mental.long, mental.surv, 
    mental.baseline, 
    id.col = "id", time.col = "time")

# Fit the joint model
model.joint <- joint(
    mental.jd, Y ~ time + time:treat, 
    Surv(surv.time, cens.ind) ~ treat, 
    model = "intslope")

# Bootstrap SEs
model.jointSE <- jointSE(model.joint, 100)
```
# Schizophrenia trial data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Separate</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>53.85</td>
<td>0.88</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.85</td>
<td>0.35</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-1.42</td>
<td>0.46</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-2.09</td>
<td>0.45</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.59</td>
<td>0.29</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.99</td>
<td>0.33</td>
</tr>
<tr>
<td>$\gamma_y$</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Log-likelihood: -2873.3, -2840.9
Extension I: Competing risks events
Motivation

- Joint models generally consider survival data that allows for one event with a single mode of failure and an assumption of independent censoring.

- When there are several reasons why the failure event can occur, or some informative censoring occurs, it is known as competing risks.
SANAD trial data

- SANAD (Standard And New Antiepileptic Drugs) was a non-blinded randomized control trial recruiting patients with epilepsy to test anti-epilepsy drugs (AEDs).
- Patients were randomized to one of 5 antiepilepsy drugs (AEDs).
- Subgroup analysis \( n = 605 \) patients) comparing 2 drugs: LTG (newer drug) vs. CBZ (standard).
Primary outcome was time to treatment failure, defined as the time to withdrawal of a randomised drug or addition of another/switch to an alternative AED

Patients decided to switch to an alternative AED because of

- Inadequate seizure control (ISC; $n = 94$, 15%)
- Unacceptable adverse effects (UAE; $n = 120$, 20%)
Clinical question

**Question**

Is LTG superior to CBZ in terms of (a) seizure control and (b) tolerability?
Rationale for competing events

- Should we just analyse composite time-to-treatment failure for any reason?
Rationale for competing events

- Should we just analyse composite time-to-treatment failure for any reason?
- **No:** assumes reasons of failure are of equal importance (which may not be the case):
  - Loss of driving license due to continued seizures
  - Side effects such as nausea, dizziness or rash
Original results

LTG is significantly more tolerable than CBZ

![Graph showing the probability over years for LTG and CBZ](image-url)
Original results

LTG is significantly more tolerable than CBZ

LTG is similar to CBZ in terms of seizure control

UAE

\[ \text{Probability} = f(\text{Years}, \text{LTG, CBZ}) \]

ISC

\[ \text{Probability} = g(\text{Years}, \text{LTG, CBZ}) \]
Drug titration

- It was suggested that different titration rates may have been to the disadvantage of standard drug CBZ.
- AED titrated more quickly brings benefits in terms of seizure control but be more likely to cause adverse effects.
Drug titration

- It was suggested that different titration rates may have been to the disadvantage of standard drug CBZ.
- AED titrated more quickly brings benefits in terms of seizure control but be more likely to cause adverse effects.
Revised clinical questions

Questions

1. Is drug titration associated with time to treatment failure for each cause?

2. After adjusting for drug titration is LTG still superior to CBZ in terms of UAE and ISC?
For each subject $i = 1, \ldots, n$, we observe

- An $n_i$-vector of observed longitudinal measurements for the longitudinal outcome: $y_i = (y_{i1}, \ldots, y_{in_i})^\top$
- Observation times $t_{ij}$ for $j = 1, \ldots, n_i$, which can differ between subjects
- $(T_i, \delta_i)$, where $T_i = \min(T_i^*, C_i)$, where $T_i^*$ is the true event time, $C_i$ corresponds to a potential non-informative right-censoring time, and $\delta_i$ is the failure indicator equal to $g$ ($g = 1, \ldots, G$) if the failure is observed ($T_i^* \leq C_i$) and due to cause $g$, and 0 otherwise
Model

- **Longitudinal sub-model**: same as for the univariate joint model
- **Time-to-event sub-model**: cause-specific hazards model
Time-to-event sub-model

Cause-specific hazards model with sub-model for cause \( g \) \((g = 1, \ldots, G)\):

\[
\lambda_{i}^{(g)}(t) = \lambda_{0}^{(g)}(t) \exp \left\{ v^{\top} \gamma^{(g)} + W_{2i}^{(g)}(t) \right\},
\]

where

- \( \lambda_{0}^{(g)}(\cdot) \) \((g = 1, \ldots, G)\) are unspecified baseline hazard functions
- \( v_{i} \) is a \( q \)-vector of baseline covariates with corresponding fixed effect terms \( \gamma_{v}^{(g)} \) \((g = 1, \ldots, G)\)
- \( W_{2i}^{(g)}(t) \) \((g = 1, \ldots, G)\) are zero-mean latent Gaussian processes
Association structure

Following Williamson et al. (2008)

\[ W_{2i}^{(g)}(t) = \gamma_y^{(g)} W_{1i}(t), \text{ for } g = 1, \ldots, G \]

with \( \gamma_y^{(g)} \) a scalar parameter for estimation capturing the association between the \( g \)-th cause-specific hazard function and \( W_1(t) \)
EM algorithm as per before, with exception that we now estimate G-pairs of \( (\lambda_0(g)(t), \gamma_y(g)) \) during the M-step

Standard errors estimated using same bootstrap method
Software

- We can fit these models in `joineR` using the same general work-flow as for the single failure type joint model.
- `joineR::joint()` automatically detects presence of competing risks\(^7\)
- Currently limited to 2 failure types

\(^7\)Requires failure types to be coded as 0 (censored), 1 (first failure type), 2 (second failure type).
Epilepsy data

Longitudinal sub-model: with $X_i = 1$ (LTG) vs. $X_i = 0$ (CBZ)

$$Y_i(t) = \beta_0 + \beta_1 t + \beta_2 X_i + \beta_3 X_i \times t + W_{1i}(t) + \epsilon_i(t)$$

$$W_{1i} = b_{i0} + b_{i1} t$$

$$(b_{i0}, b_{i1})^\top \sim N_2(0, D)$$

$$\epsilon_i(t) \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

Time-to-event sub-model: with $g \in \{\text{UAE, ISC}\}$

$$\lambda_i^{(g)}(t) = \lambda_0^{(g)}(t) \exp \left\{ \gamma_v^{(g)} X_i + W_{2i}^{(g)}(t) \right\}$$

$$W_{2i}^{(g)}(t) = \gamma_y^{(g)} W_{1i}(t)$$
# Setup joint data

data(epileptic)

epileptic$interaction <- with(epileptic, time * (treat == "LTG"))

longitudinal <- epileptic[, c(1:3, 13)]

survival <- UniqueVariables(epileptic, c(4, 6), "id")

baseline <- UniqueVariables(epileptic, "treat", "id")

data <- jointdata(longitudinal = longitudinal,
                    survival = survival,
                    baseline = baseline,
                    id.col = "id", time.col = "time")

# Fit joint model

fit2 <- joint(data = data,
               long.formula = dose ~ time + treat + interaction,
               surv.formula = Surv(with.time, with.status2) ~ treat,
               longsep = FALSE, survsep = FALSE)
### Epilepsy data

<table>
<thead>
<tr>
<th>( \beta ) or ( \gamma )</th>
<th>Description</th>
<th>Est.</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>(Intercept)</td>
<td>1.9730</td>
<td>0.0512</td>
<td>(1.8701, 2.0803)</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>time</td>
<td>0.0004</td>
<td>0.0001</td>
<td>(0.0002, 0.0005)</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>LTG</td>
<td>-0.1381</td>
<td>0.0720</td>
<td>(-0.2908, -0.0307)</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>time ( \times ) LTG</td>
<td>0.0006</td>
<td>0.0001</td>
<td>(0.0003, 0.0009)</td>
</tr>
<tr>
<td>( \gamma_v^{(UAE)} )</td>
<td>LTG</td>
<td>-0.6602</td>
<td>0.2008</td>
<td>(-1.1063, -0.2338)</td>
</tr>
<tr>
<td>( \gamma_y^{(UAE)} )</td>
<td>Titration</td>
<td>-0.9259</td>
<td>0.2600</td>
<td>(-1.4707, -0.4510)</td>
</tr>
<tr>
<td>( \gamma_v^{(ISC)} )</td>
<td>LTG</td>
<td>0.0272</td>
<td>0.1811</td>
<td>(-0.3940, 0.3306)</td>
</tr>
<tr>
<td>( \gamma_y^{(ISC)} )</td>
<td>Titration</td>
<td>0.5894</td>
<td>0.0873</td>
<td>(0.4031, 0.7513)</td>
</tr>
</tbody>
</table>
Conclusion

Is LTG superior to CBZ after adjusting for titration in terms of UAE? ISC?

- If LTG is titrated at the same rate as CBZ, the beneficial effect of LTG on UAE would still be evident
- LTG and CBZ still appear to provide similar seizure control
Extension II: Multiple longitudinal outcomes
Motivation

- Clinical studies often repeatedly measure *multiple* biomarkers or other measurements and an event time.
- Research has predominantly focused on single measurement outcomes.
- Harnessing all available information in a single model is advantageous and should lead to improved estimation and predictions.
The Mayo Clinic PBC data

- 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)

- We analyse a subset of 154 patients randomized to placebo

- Multiple biomarkers repeatedly measured at intermittent times:
  1. serum bilirubin (mg/dl)
  2. serum albumin (mg/dl)
  3. prothrombin time (seconds)

- Patients drop out if they die ($n_{\text{die}} = 69, 44.8\%$)
**NB.** transformations chosen according to Box-Cox transformations
What is the state of the field?

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

- $y_i = (y_{i1}^T, \ldots, y_{iK}^T)$ 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:
  
  $y_{ik} = (y_{i1k}, \ldots, y_{in_{ik}k})^T$

- Observation times $t_{ijk}$ for $j = 1, \ldots, n_{ik}$, which can differ between subjects and outcomes

- $(T_i, \delta_i)$, 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 $\delta_i$ is the failure indicator equal to 1 if the failure is observed ($T_i^* \leq C_i$) and 0 otherwise
Longitudinal sub-model

For the \( k \)-th outcome \((k = 1, \ldots, K)\)

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

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 \( \beta_k \)
- \( W_{1i}^{(k)}(t) \) 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 \( \gamma_v \)
- \( W_{2i}(t) \) is a zero-mean latent Gaussian process, independent of the censoring process.
Extending Laird and Ware (1982) to the multivariate case:

\[ W_{1i}^{(k)}(t) = z_{ik}(t)b_{ik} \]
Association structure

Extending Laird and Ware (1982) to the multivariate case:

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

Joint model then captures 3 types of correlation:

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: \( W_{2i}(t) = \sum_{k=1}^{K} \gamma_{yk} W_{1i}^{(k)}(t) \)
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 \( \beta = (\beta_1^T, \ldots, \beta_K^T) \), and

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

\[
Z_i = \begin{pmatrix}
Z_{i1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & Z_{iK}
\end{pmatrix}, \quad \Sigma_i = \begin{pmatrix}
\sigma_1^2 I_{n_{i1}} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma_K^2 I_{n_{iK}}
\end{pmatrix}.
\]
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 $\theta = (\beta^T, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^T, \gamma_y^T)$
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 \( \theta = (\beta^T, \text{vech}(D), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \gamma_v^T, \gamma_y^T) \), and

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

- **EM algorithm + Monte Carlo (MC) E-step = MCEM algorithm**
- **M-step** identical except we now estimate $K$ error variances $\sigma_1^2, \ldots, \sigma_K^2$ as

\[
\hat{\sigma}_k^2 = \frac{1}{\sum_{i=1}^{n} n_{ik}} \sum_{i=1}^{n} \left\{ (y_{ik} - X_{ik}\beta_k)^\top (y_{ik} - X_{ik}\beta_k - 2Z_{ik}\mathbb{E}[b_{ik}]) \right. \\
\left. + \text{trace} \left( Z_{ik}^\top Z_{ik} \mathbb{E}[b_{ik} b_{ik}^\top] \right) \right\}
\]

\(^8\)See Wei and Tanner (1990)
E-step

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

where

\[
h(\cdot) = \text{any known function,}
\]

\[
b_i \mid y_i, \theta \sim N \left( A_i \left\{ Z_i^T \Sigma_i^{-1} (y_i - X_i \beta) \right\}, A_i \right), \text{ and}
\]

\[
A_i = \left( Z_i^T \Sigma_i^{-1} Z_i + D^{-1} \right)^{-1}
\]
Replace the Gauss-Hermite quadrature with a Monte Carlo approximation for scalability with increasing $K$ and/or $\dim(b_i)$

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

Replace the Gauss-Hermite quadrature with a Monte Carlo approximation for scalability with increasing $K$ and/or $\dim(b_i)$

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

where $b_i^{(1)}, b_i^{(2)}, \ldots, b_i^{(N)}$ are a random sample from $b_i \mid y_i, \theta$
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
In standard EM, convergence usually declared at \((m + 1)\)-th iteration if one of the following criteria satisfied

1. Relative change: \[ \Delta_{\text{rel}}^{(m+1)} = \max \left\{ \frac{|\hat{\theta}(m+1) - \hat{\theta}(m)|}{|\hat{\theta}(m)| + \epsilon_1} \right\} < \epsilon_0 \]

2. Absolute change: \[ \Delta_{\text{abs}}^{(m+1)} = \max \left\{ |\hat{\theta}(m+1) - \hat{\theta}(m)| \right\} < \epsilon_2 \]

for some choice of \(\epsilon_0, \epsilon_1,\) and \(\epsilon_2\)

**Note:** `joineR` uses criterion \#2
In MCEM framework, there are 2 complications to account for
In MCEM framework, there are 2 complications to account for:

1. Spurious convergence declared due to random chance

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

⇒ Solution: Increase Monte Carlo size as algorithm moves closer towards maximizer.
Convergence

In MCEM framework, there are 2 complications to account for:

1. spurious convergence declared due to random chance

    ⇒ **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
   ⇒ **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
   ⇒ **Solution**: require convergence for 3 iterations in succession

2. estimators swamped by Monte Carlo error, thus precluding convergence
   ⇒ **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 $N + \lceil N/\delta \rceil$ if $\text{cv}(\Delta_{\text{rel}}^{(m+1)}) > \text{cv}(\Delta_{\text{rel}}^{(m)})$ for some small positive integer $\delta$
Standard error estimation

- We could use bootstrap method again, but it is getting slow as \( \text{dim}(\theta) \) increases.
- Instead, we approximate the information matrix of \( \theta - \lambda_0(t) \) (with \( \lambda_0(t) \) profiled out of the likelihood) by the empirical information matrix (McLachlan and Krishnan 2008):
  \[
  I_e(\theta) = \frac{1}{n} \sum_{i=1}^{n} s_i(\theta) s_i^\top(\theta) - \frac{1}{n} S(\theta) S^\top(\theta),
  \]
  where
  \[
  S(\theta) = \sum_{i=1}^{n} s_i(\theta)
  \]
  is the score vector.
- Then use \( \text{SE}(\theta) \approx I_e^{-1/2}(\hat{\theta}) \)
We can implement all of this in the R package `joineRML`

- Single function required: `joineRML::mjoint()`
- Package can also be used to fit univariate joint models, but using MCEM rather than EM optimisation
- Calculates approximate SEs by default, but bootstrap SEs available via `joineRML::bootSE()`
Proposed model for PBC data

Longitudinal sub-model

\[
\begin{align*}
\log(\text{serBilir}) &= (\beta_{0,1} + b_{0i,1}) + (\beta_{1,1} + b_{1i,1})\text{year} + \varepsilon_{ij1}, \\
\text{albumin} &= (\beta_{0,2} + b_{0i,2}) + (\beta_{1,2} + b_{1i,2})\text{year} + \varepsilon_{ij2}, \\
(0.1 \times \text{prothrombin})^{-4} &= (\beta_{0,3} + b_{0i,3}) + (\beta_{1,3} + b_{1i,3})\text{year} + \varepsilon_{ij3}, \\
b_i &\sim N(0, D), \text{ and } \varepsilon_{ijk} \sim N(0, \sigma_k^2) \text{ for } k = 1, 2, 3
\end{align*}
\]

Time-to-event sub-model

\[
\begin{align*}
\lambda_i(t) &= \lambda_0(t) \exp \{\gamma_{\text{age}} + W_{2i}(t)\}, \\
W_{2i}(t) &= \gamma_{\text{bil}}(b_{0i,1} + b_{1i,1}t) + \gamma_{\text{alb}}(b_{0i,2} + b_{1i,2}t) + \gamma_{\text{pro}}(b_{0i,3} + b_{1i,3}t)
\end{align*}
\]
Example code

data(pbc2)
placebo <- subset(pbc2, drug == "placebo")
fit.pbc <- mjoint(
    formLongFixed = list(
        "bil" = log(serBilir) ~ year,
        "alb" = albumin ~ year,
        "pro" = (0.1 * prothrombin)^-4 ~ year),
    formLongRandom = list(
        "bil" = ~ year | id,
        "alb" = ~ year | id,
        "pro" = ~ year | id),
    formSurv = Surv(years, status2) ~ age,
    data = placebo,
    timeVar = "year",
    control = list(tol0 = 0.001, burnin = 400))
### Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{0,1}$</td>
<td>0.5541</td>
<td>0.0858</td>
<td>(0.3859, 0.7223)</td>
</tr>
<tr>
<td>$\beta_{1,1}$</td>
<td>0.2009</td>
<td>0.0201</td>
<td>(0.1616, 0.2402)</td>
</tr>
<tr>
<td>$\beta_{0,2}$</td>
<td>3.5549</td>
<td>0.0356</td>
<td>(3.4850, 3.6248)</td>
</tr>
<tr>
<td>$\beta_{1,2}$</td>
<td>-0.1245</td>
<td>0.0101</td>
<td>(-0.1444, -0.1047)</td>
</tr>
<tr>
<td>$\beta_{0,3}$</td>
<td>0.8304</td>
<td>0.0212</td>
<td>(0.7888, 0.8719)</td>
</tr>
<tr>
<td>$\beta_{1,3}$</td>
<td>-0.0577</td>
<td>0.0062</td>
<td>(-0.0699, -0.0456)</td>
</tr>
<tr>
<td>$\gamma_v$</td>
<td>0.0462</td>
<td>0.0151</td>
<td>(0.0166, 0.0759)</td>
</tr>
<tr>
<td>$\gamma_{bil}$</td>
<td>0.8181</td>
<td>0.2046</td>
<td>(0.4171, 1.2191)</td>
</tr>
<tr>
<td>$\gamma_{alb}$</td>
<td>-1.7060</td>
<td>0.6181</td>
<td>(-2.9173, -0.4946)</td>
</tr>
<tr>
<td>$\gamma_{pro}$</td>
<td>-2.2085</td>
<td>1.6070</td>
<td>(-5.3582, 0.9412)</td>
</tr>
</tbody>
</table>
Summary
Joint models can be formulated to ‘link’ the ubiquitous Cox proportional hazards model and the linear mixed effects model. The basic model can be extended with only slight added complexity to handle competing risks or multivariate outcomes. These models can be straightforwardly fit using R.
Beyond the basics

Things not covered today:

- Alternative distributional assumptions (or lack of them)
- Alternative association structures
- How to go from fitted model to dynamic prediction
- Diagnostics
- Non-continuous longitudinal outcomes
- ...

It’s a rapidly growing field!
Where to find out more

Univariate joint model

Competing risks joint model

Multivariate longitudinal data
Where to find out more

Joint Models for Longitudinal and Time-to-Event Data

With Applications in R

Dimitris Rizopoulos

Monographs on Statistics and Applied Probability 151

Joint Modeling of Longitudinal and Time-to-Event Data

Robert M. Elashoff
Gang Li
Ning Li


Andrinopoulou, Eleni-Rosalina et al. (2015). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research* 0(0), pp. 1–18.


Tea & coffee break (outside) — 1045–1100

Catch us at the break or lunch

R practical session — 1100–1230