Week 2: Bivariate joint models

Laurent Briollais[1,2]

[1]Lunenfeld-Tanenbaum Research Institute
[2] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

May 13, 2020

Previous lecture

1. Joint Modeling (JM): General concepts
2. Original applications of JM in Health Research
3. The longitudinal component of the JM
4. The survival component of the JM

Outline of lecture #2

1. Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
2. Extension of the Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
3. Bivariate Joint Model for a Recurrent and a Terminal Time-to-Event Outcomes (*R* package Frailtypack)

Review: Extended Cox regression model

- Recall that the Cox model can be extended to handle exogenous time-dependent covariates

$$h_i(t|Y_i(t), X_i) = h_0(t)R_i(t)\exp\{X_i^T\beta + \alpha y_i(t)\}$$

where
$Y_i(t) = \{y_i(s), 0 \leq s < t\}$
$R_i(t)$ denotes the at risk process (1 if subject i still at risk at t and 0 otherwise),
$y_i(t)$ denotes the value of the time-varying covariate at t.

- $\exp(\alpha)$ denotes the relative increase in the risk for an event at time t that results from one unit increase in $y_i(t)$ at the same time point

Review: Extended Cox regression model

Recall the main assumptions:

- Assumes no measurement error
- Step-function path for $y_i(t)$
- Existence and level of the covariate (e.g., biomarker) is not related to time-to-event status

Outline of lecture #2

1. Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
2. Extension of the Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
3. Bivariate Joint Model for a Recurrent and a Terminal Time-to-Event Outcomes (*R* package Frailtypack)

The standard Joint Model

- To account for the special features of endogenous covariates a new class of models has been developed
  $\Rightarrow$ Joint models for longitudinal and time-to-event data
- Intuitive idea behind these models:
  - Use an appropriate model to describe the evolution of the time-dependent covariate (biomarker) for each patient over time
    $\Rightarrow$ Complex time functions and measurement errors are taken into account
  - The estimated evolution are then used in a Cox model
    $\Rightarrow$ The full path of the time dependent covariate is taken into account
  - The time-dependent covariate is not assumed constant between visits

The standard Joint Model

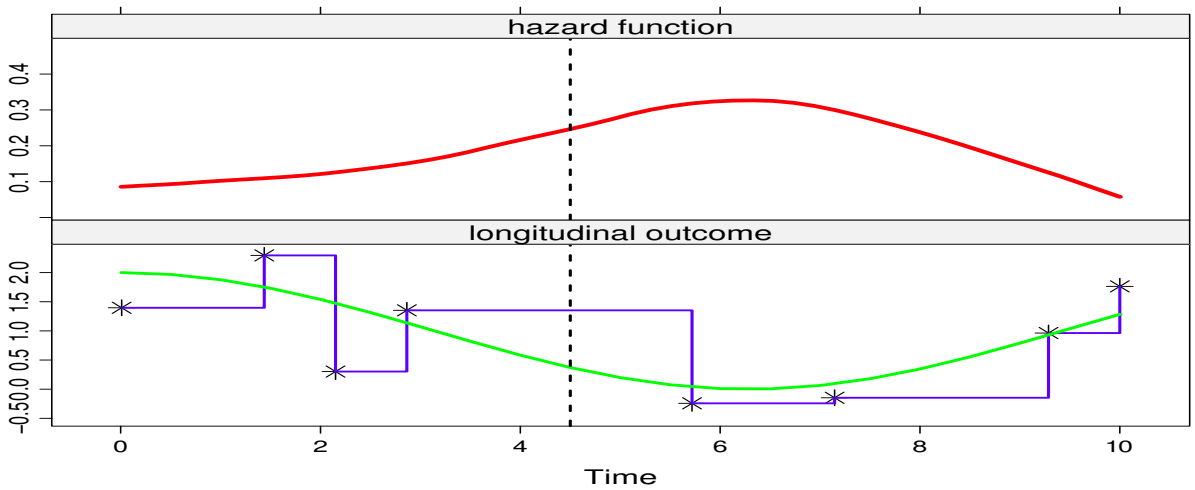


Figure: Form of the hazard function

The standard Joint Model: Notations

- T_i^* : True event time for patient i
- T_i: Observed event time for patient i
- δ_i: Event indicator, i.e., equals 1 for true events
- y_i: Longitudinal responses

Three steps in joint modeling

Step 1: the terminal event

- Let's assume that we know $m_i(t)$, i.e., the true & unobserved value of the marker at time t

- Then, we can define a standard hazard model for the terminal event

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t)\},$$

where

- $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ is the longitudinal history,
- α quantifies the strength of the association between the marker and the risk for an event,
- w_i are the baseline covariates.

Three steps in joint modeling

Step 2: the longitudinal outcome

- From the observed longitudinal response $y_i(t)$, reconstruct the covariate history for each subject

- Mixed effects model (we focus, for now, on continuous markers)

$$y_i(t) = m_i(t) + \epsilon_i(t)$$

$$= x_i^T \beta + z_i^T(t) b_i + \epsilon_i(t), \quad \epsilon_i(t) \sim \mathcal{N}(0, \sigma^2),$$

where

- $x_i(t)$ and β account for the fixed part of the model
- $z_i(t)$ and b_i account for the random part of the model, with $b_i \sim \mathcal{N}(0, D)$.

Three steps in joint modeling

Step 3: association model

- The two processes are associated $\Rightarrow$ define a model for their joint distribution

- Joint Models for such joint distributions are of the following form (Tsiatis & Davidian, Stat. Sinica, 2004)

$$p(y_i, T_i, \delta_i) = \int p(y_i|b_i)\{h(T_i|b_i)^{\delta_i} S(T_i|b_i)\} f(b_i) db_i$$

where

- b_i a vector of random effects that explains the interdependencies $\Rightarrow$ between the longitudinal and time-to-event process (e.g., terminal event) and within the longitudinal process.
- $f(.)$ density function; $S()$ survival function.

Three steps in joint modeling

Step 3

- Key assumption: Full Conditional Independence
  $\Rightarrow$ random effects explain all interdependencies
- The longitudinal outcome is independent of the time-to-event outcome given the random effects
- The repeated measurements in the longitudinal outcome are independent of each other given the random effects
  - $p(y_i, T_i, \delta_i | b_i) = p(y_i | b_i) p(T_i, \delta_i | b_i)$
  - $p(y_i | b_i) = \prod_j p(y_{ij} | b_i)$

Three steps in joint modeling

- Caveat: The conditional assumption is often difficult to test.
- The censoring and visiting processes are assumed non-informative
- The visiting process is defined as the mechanism (stochastic or deterministic) that generates the time points at which longitudinal measurements are collected.
- Decision to withdraw from the study or to appear at the next visit
  - May depend on observed past history (baseline covariates + observed longitudinal responses)
  - May depend on terminal event status
  - BUT no additional dependence on underlying, latent subject characteristics associated with prognosis

Three steps in joint modeling

- The survival function, which is a part of the likelihood of the model, depends on the whole longitudinal history (unlike the hazard function).

$$S_i(t|b_i) = \exp\left(- \int_0^t h_0(s) \exp\{\gamma^T w_i + \alpha m_i(s)\} ds \right)$$

- Therefore, care in the definition of the design matrices of the mixed model.

- When subjects have nonlinear biomarker profiles $\Rightarrow$ use splines or polynomials to allow flexibility in the modeling.

Random-effects distribution

- In mixed models, it is customary to assume normality
- However, in joint models this distribution plays a more prominent role because the random effects explain all associations
- Nevertheless, we have model robustness, especially as n_i increases (see Rizopoulos et al., 2008, Biometrika)

Assumptions for the baseline hazard function $h_0(t)$

- Parametric $\Rightarrow$ possibly restrictive, e.g. splines

$$\log h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, v)$$

where

- $B_q(t, v)$ denotes the q-th basis function of a B-spline with knots $v_1, \cdots, v_Q$
- γ_{h_0} a vector of spline coefficients

- Non-parametric

Estimation

- Mainly maximum likelihood or Bayesian approaches
- The log-likelihood contribution for subject i:

$$l(\theta_i) = \log \int \{\prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta)\}\{h_i(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta)\} f(b_i;\theta) db_i,$$

where

$$S_i(t|b_i;\theta) = \exp\Big(-\int_0^t h_0(s;\theta)\exp\{\gamma^T w_i + \alpha m_i(s)\}ds\Big)$$

- Both integrals do not have, in general, a closed-form solution $\Rightarrow$ need to be approximated numerically

Estimation

- Standard numerical integration algorithms
  - Gaussian quadrature
  - Monte Carlo

- More difficult is the integral with respect to b_i because it can be of high dimension
  - Laplace approximations
  - pseudo-adaptive Gaussian quadrature rules

Estimation

- To maximize the approximated log-likelihood

$$l(\theta) = \sum_{i=1}^{n} \log \int p(y_i|b_i;\theta)\{h_i(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta)\} f(b_i;\theta) db_i,$$

  we need to employ an optimization algorithm

- Standard choices:
  - EM (treating b_i as missing data)
  - Newton-type
  - hybrids (start with EM and continue with quasi-Newton)

Estimation

- Standard errors: Standard asymptotic MLE

$$\hat{\text{var}}(\hat{\theta}) = \left\{ -\sum_{i=1}^{n} \frac{\partial^2 \log p(y_i, T_i, \delta_i; \theta)}{\partial \theta^T \partial \theta} \Big|_{\theta=\hat{\theta}} \right\}^{-1}$$

- Standard asymptotic tests + information criteria
  - likelihood ratio test
  - score test
  - Wald test
  - AIC, BIC, . . .

Bayesian estimation

- Based on a fitted joint model, estimates for the random effects are based on the posterior distribution:

$$p(b_i|y_i, T_i, \delta_i; \theta) = \frac{p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta)f(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)}$$

$$\propto p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta)f(b_i; \theta)$$

  in which θ is replaced by its MLE $\hat{\theta}$.

- Bayesian estimation: both θ and $\{b_i, i = 1, \cdots, n\}$ are regarded as parameters
- Inference is based on the full posterior distribution $p(\theta, b|T, \delta, y)$.
- No closed-form solutions for the integrals in the normalizing constant $\Rightarrow$ MCMC

Comparison JM vs. extended Cox model

Example: To illustrate the virtues of joint modelling, we compare it with the standard time-dependent Cox model on the AIDS data

AIDS: 467 HIV infected patients who had failed or were intolerant to zidovudine therapy (AZT) (Abrams et al., NEJM, 1994)

The aim of this study was to compare the efficacy and safety of two alternative antiretroviral drugs, didanosine (ddI) and zalcitabine (ddC)

Comparison JM vs. extended Cox model

Outcomes of interest

- time to death
- randomized treatment: 230 patients ddI and 237 ddC
- CD4 cell count measurements at baseline, 2, 6, 12 and 18 months
- prevOI: previous opportunistic infections

Comparison JM vs. extended Cox model

Data

```
> aids
     patient  Time death       CD4 obstime drug gender prevOI          AZT start  stop event
1          1 16.97     0 10.677078       0  ddC   male   AIDS intolerance     0  6.00     0
2          1 16.97     0  8.426150       6  ddC   male   AIDS intolerance     6 12.00     0
3          1 16.97     0  9.433981      12  ddC   male   AIDS intolerance    12 16.97     0
4          2 19.00     0  6.324555       0  ddI   male noAIDS intolerance     0  6.00     0
5          2 19.00     0  8.124038       6  ddI   male noAIDS intolerance     6 12.00     0
6          2 19.00     0  4.582576      12  ddI   male noAIDS intolerance    12 18.00     0
7          2 19.00     0  5.000000      18  ddI   male noAIDS intolerance    18 19.00     0

> aids.id
    patient  Time death       CD4 obstime drug gender prevOI          AZT start  stop event
1          1 16.97     0 10.677078       0  ddC   male   AIDS intolerance     0  6.00     0
2          2 19.00     0  6.324555       0  ddI   male noAIDS intolerance     0  6.00     0
3          3 18.53     1  3.464102       0  ddI female   AIDS intolerance     0  2.00     0
4          4 12.70     0  3.872983       0  ddC   male   AIDS     failure     0  2.00     0
5          5 15.13     0  7.280110       0  ddI   male   AIDS     failure     0  2.00     0
```

Comparison JM vs. extended Cox model

Model

$$\begin{cases} y_i(t) = m_i(t) + \epsilon_i(t) \\ \qquad\quad = x_i^T(t)\beta + z_i^T(t)b_i + \epsilon_i(t), \quad \epsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \\ \qquad\quad = \beta_0 + \beta_1 t + \beta_2\{t \times ddl_i\} + b_{i0} + b_{i1} t + \epsilon_i(t), \\ h_i(t) = h_0(t)\exp\{\gamma ddl_i + \alpha m_i(t)\} \end{cases}$$

where $h_0(t)$ is assumed piecewise-constant.

Comparison JM vs. extended Cox model

Joint model in *R*

Joint models are fitted using function jointModel() from package JM. This function accepts as main arguments a linear mixed model and a Cox PH model based on which it fits the corresponding joint model

Comparison JM vs. extended Cox model

R Code: Mixed model

```
library(nlme)
library(splines)
library(survival)
library(JM)

data('aids')

lmeFit <- lme(CD4 ~ obstime + obstime:drug, random = ~ obstime | patient, data = aids)

> summary(lmeFit)
Linear mixed-effects model fit by REML
 Data: aids
       AIC      BIC    logLik
  7147.575 7184.295 -3566.788

Random effects:
 Formula: ~obstime | patient
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 4.5901582 (Intr)
obstime     0.1738082 -0.155
Residual    1.7497905
```

Comparison JM vs. extended Cox model

R Code: Mixed model

```
Fixed effects: CD4 ~ obstime + obstime:drug
                    Value  Std.Error  DF  t-value p-value
(Intercept)      7.188833 0.22215874 936 32.35899  0.0000
obstime         -0.163451 0.02080804 936 -7.85519  0.0000
obstime:drugddI  0.028272 0.02970929 936  0.95163  0.3415
 Correlation:
                (Intr) obstim
obstime         -0.160
obstime:drugddI  0.000 -0.682

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-4.32530054 -0.41785802 -0.04720642  0.40631129  4.32727623

Number of Observations: 1405
Number of Groups: 467
```

Comparison JM vs. extended Cox model

R Code: Cox model

```
  coxFit <- coxph(Surv(Time, death) ~ drug, data = aids.id, x = TRUE)
summary(coxFit)

Call:
coxph(formula = Surv(Time, death) ~ drug, data = aids.id, x = TRUE)

  n= 467, number of events= 188

          coef exp(coef) se(coef)     z Pr(>|z|)
drugddI 0.2102    1.2339   0.1462 1.437    0.151

        exp(coef) exp(-coef) lower .95 upper .95
drugddI     1.234     0.8104    0.9264     1.643

Concordance= 0.531  (se = 0.019 )
Rsquare= 0.004   (max possible= 0.99 )
Likelihood ratio test= 2.07  on 1 df,   p=0.15
Wald test            = 2.07  on 1 df,   p=0.1506
Score (logrank) test = 2.07  on 1 df,   p=0.1498
```

Comparison JM vs. extended Cox model

R Code: Cox model

```
  coxFit <- coxph(Surv(Time, death) ~ CD4, data = aids.id, x = TRUE)
summary(coxFit)

Call:
coxph(formula = Surv(Time, death) ~ CD4, data = aids.id, x = TRUE)

  n= 467, number of events= 188

       coef exp(coef) se(coef)      z Pr(>|z|)
CD4 -0.18162   0.83392  0.02222 -8.175 3.33e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

    exp(coef) exp(-coef) lower .95 upper .95
CD4    0.8339      1.199    0.7984     0.871

Concordance= 0.7  (se = 0.022 )
Rsquare= 0.173   (max possible= 0.99 )
Likelihood ratio test= 88.62  on 1 df,   p=0
Wald test            = 66.84  on 1 df,   p=3.331e-16
Score (logrank) test = 74.26  on 1 df,   p=0
```

Comparison JM vs. extended Cox model

R Code: JM

```
 jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
method = "piecewise-PH-aGH")
summary(jointFit)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    method = "piecewise-PH-aGH")

Data Descriptives:
Longitudinal Process Event Process
Number of Observations: 1405 Number of Events: 188 (40.3%)
Number of Groups: 467

Joint Model Summary:
Longitudinal Process: Linear mixed-effects model
Event Process: Relative risk model with piecewise-constant
baseline risk function
Parameterization: Time-dependent

   log.Lik      AIC      BIC
 -4328.261 8688.523 8754.864
```

Comparison JM vs. extended Cox model

R Code: JM

```
 Variance Components:
             StdDev    Corr
(Intercept)  4.5839  (Intr)
obstime      0.1822 -0.0468
Residual     1.7377

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      7.2203  0.2218 32.5537 <0.0001
obstime         -0.1917  0.0217 -8.8374 <0.0001
obstime:drugddI  0.0116  0.0302  0.3834  0.7014

Event Process
            Value Std.Err  z-value p-value
drugddI    0.3348  0.1565   2.1397  0.0324
Assoct    -0.2875  0.0359  -8.0141 <0.0001
log(xi.1) -2.5438  0.1913 -13.2953
log(xi.2) -2.2722  0.1784 -12.7328
log(xi.3) -1.9554  0.2403  -8.1357
log(xi.4) -2.5011  0.3412  -7.3297
log(xi.5) -2.4152  0.3156  -7.6531
log(xi.6) -2.4018  0.4007  -5.9941
log(xi.7) -2.4239  0.5301  -4.5725
```

Comparison JM vs. extended Cox model

Results

Table: Parameter estimates

	JM			Cox		
	logHR	SE	*p*-value	logHR	SE	*p*-value
Treat	0.33	0.16	0.032	0.21	0.15	0.15
$CD4^{1/2}$	-0.29	0.04	< 0.0001	-0.18	0.02	< 0.0001

Comparison JM vs. extended Cox model

Results

- Clearly, there is a considerable effect of ignoring the association between the 2 processes.
- A unit decrease in $CD4^{1/2}$, results in a
  - Joint Model: 1.3-fold increase in risk (95% CI: 1.23; 1.45)
  - Time-Dependent Cox: 1.2-fold increase in risk (95% CI: 1.15; 1.25)
- The treatment effect (ddI vs. ddC) is associated with
  - Joint Model: 1.4-fold increase in risk (95% CI: 1.01; 1.92)
  - Time-Dependent Cox: 1.2-fold increase in risk (95% CI: 0.91; 1.67)

Comparison JM vs. extended Cox model

Results

- Which one to believe?

  $\Rightarrow$ Theoretical and simulation works have shown that the Cox model underestimates the true association size of markers

- Note also the very significant negative association between the 2 processes: $\alpha = -0.28$

JM *R* package options

- As before, the data frame given in lme() should be in the long format, while the data frame given to coxph() should have one line per subject.
- The ordering of the subjects needs to be the same.
- In the call to coxph() you need to set x = TRUE (or model = TRUE) such that the design matrix used in the Cox model is returned in the object fit
- Argument timeVar specifies the time variable in the linear mixed model

JM *R* package options

- Argument method specifies the type of relative risk model and the type of numerical integration algorithm, the syntax is as follows: <baseline hazard> <parameterization> <numerical integration>

- Available options are:
  - 'piecewise-PH-GH': PH model with piecewise-constant baseline hazard
  - 'spline-PH-GH': PH model with B-spline-approximated log baseline hazard
  - 'weibull-PH-GH': PH model with Weibull baseline hazard
  - 'weibull-AFT-GH': AFT model with Weibull baseline hazard
  - 'Cox-PH-GH': PH model with unspecified baseline hazard
- GH stands for standard Gauss-Hermite; using aGH invokes the pseudo-adaptive Gauss-Hermite rule

JMbayes *R* package for Bayesian estimation

```
library(JMbayes)

lmeFit <- lme(CD4 ~ obstime + obstime:drug,
random = ~ obstime | patient, data = aids)

coxFit <- coxph(Surv(Time, death) ~ drug, data = aids.id,
x = TRUE)

jointFitBayes <- jointModelBayes(lmeFit, coxFit,
timeVar = "obstime")

summary(jointFitBayes)
```

JMbayes *R* package for Bayesian estimation

- JMbayes is more flexible (in some respects):
- Directly implements the MCMC
- Allows categorical longitudinal data as well
- Allows general transformation functions
- Penalized B-splines for the baseline hazard function

R packages for JM

- In both packages methods are available for the majority of the standard generic functions + extras
  - summary(), anova(), vcov(), logLik()
  - coef(), fixef(), ranef()
  - fitted(), residuals()
  - plot()
  - xtable() (you need to load package xtable first)

Connection with missing data framework

- So far we have attacked the problem from the survival point of view
- However, often, we may be also interested in the longitudinal outcome
- Issue: When patients experience the event, they dropout from the study $\Rightarrow$ a direct connection with the missing data field
- Dropout must be taken into account when making inference on the longitudinal outcome

Connection with missing data framework

- To show this connection more clearly
  - T_i^*: true time-to-event
  - y_i^o: longitudinal measurements before T_i^*
  - y_i^m: longitudinal measurements after T_i^*
- Important to realize that the model we postulate for the longitudinal responses is for the complete vector $\{y_i^o, y_i^m\}$.

  $\Rightarrow$ Implicit assumptions about missingness

Connection with missing data framework

- Missing data mechanism:

$$p(T_i^*|y_i^o, y_i^m) = \int p(T_i^*|b_i)p(b_i|y_i^o, y_i^m)db_i$$

- Intuitive interpretation: Patients who dropout show different longitudinal evolutions than patients who do not
- Implications of nonrandom dropout $\Rightarrow$ observed data do not constitute a random sample from the target population
- This feature complicates the validation of the joint model's assumptions using standard residual plots
- What is the problem? Residual plots may show systematic behavior due to dropout and not because of model misfit

Connection with missing data framework

- What about censoring?

  $\Rightarrow$ censoring also corresponds to a discontinuation of the data collection process for the longitudinal outcome

- Likelihood-based inferences for joint models provide valid inferences when censoring is MAR

  - A patient relocates to another country (MCAR)
  - A patient is excluded from the study when his/her longitudinal response exceeds a prespecified threshold (MAR)
  - Censoring depends on random effects (MNAR)

Connection with missing data framework

- Joint models belong to the class of Shared Parameter Models

$$p(T_i^*, y_i^o, y_i^m) = \int p(y_i^o, y_i^m | b_i) p(T_i^* | b_i) f(b_i) db_i$$

$\Rightarrow$ The association between the longitudinal and missingness processes is explained by the shared random effects b_i.

- The other two well-known frameworks for MNAR data are

Connection with missing data framework

- Selection models

$$p(T_i^*, y_i^o, y_i^m) = p(y_i^o, y_i^m)p(T_i^*|y_i^o, y_i^m)$$

- Pattern mixture models:

$$p(T_i^*, y_i^o, y_i^m) = p(y_i^o, y_i^m|T_i^*)p(T_i^*)$$

$\Rightarrow$ These two model families are primarily applied to discrete dropout times and cannot be easily extended to continuous time

- A nice feature of joint models/shared parameter models is that they can automatically handle intermittent missing data

Connection with missing data framework

- Example: In the AIDS data the association parameter α was highly significant, suggesting nonrandom dropout
- A comparison between
  - linear mixed-effects model $\Rightarrow$ MAR
  - joint model $\Rightarrow$ MNAR is warranted

	LMM (MAR) value (s.e.)	JM (MNAR) value (s.e)
Inter	7.19 (0.22)	7.22 (0.22)
Time	−0.16 (0.02)	−0.19 (0.02)
Treat:Time	0.03 (0.03)	0.01 (0.03)

Connection with missing data framework

- MAR assumes that missingness depends only on the observed data

$$p(T_i^* | y_i^o, y_i^m) = p(T_i^* | y_i^o)$$

- Minimal sensitivity in parameter estimates & standard errors but this is always the case!

Outline of lecture #2

1. Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
2. Extension of the Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
3. Bivariate Joint Model for a Recurrent and a Terminal Time-to-Event Outcomes (*R* package Frailtypack)

JM: Different parametrizations

The standard joint model

$$\begin{cases} h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t)\}, \\ y_i(t) = m_i(t) + \epsilon_i(t), \\ \qquad\quad = x_i^T(t)\beta + z_i^T(t)b_i + \epsilon_i(t), \quad \epsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \end{cases}$$

where

$\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ is the longitudinal history.

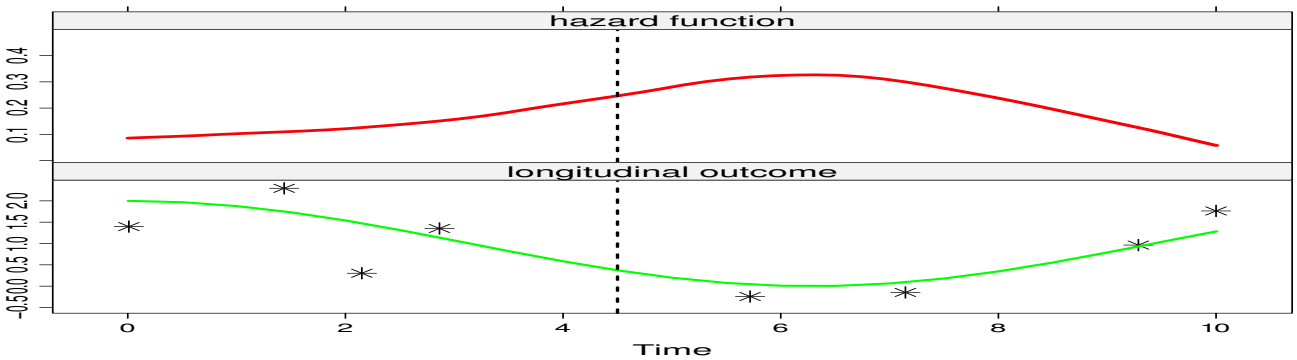


Figure: Connection between hazard and longitudinal outcome

Note: Inappropriate modelling of time-dependent covariates may result in surprising results

$\Rightarrow$ We need to carefully consider the functional form of time-dependent covariates

JM: Different parametrizations

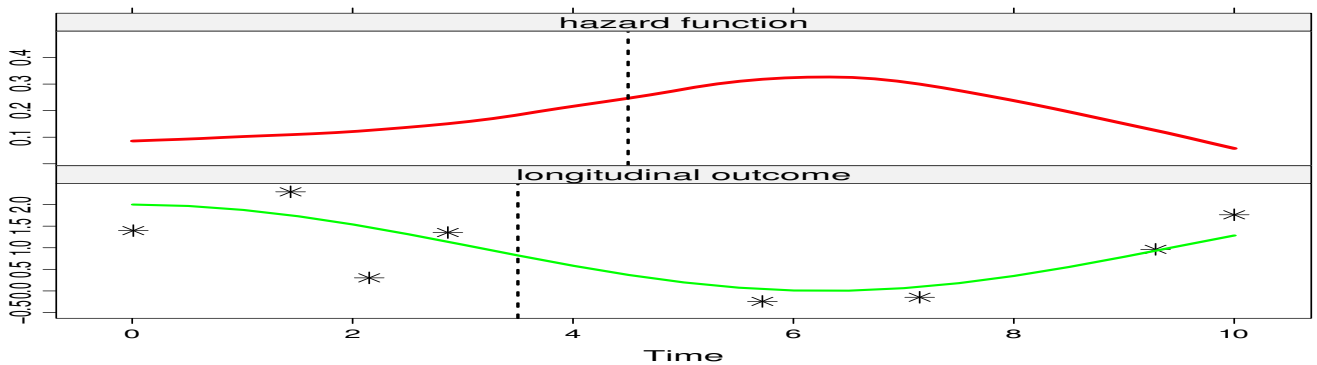
Lagged Effects

The hazard for an event at t is associated with the level of the marker at a previous time point:

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t_+^c)\}$$

where

$$t_+^c = \max(t - c, 0)$$



JM: Different parametrizations

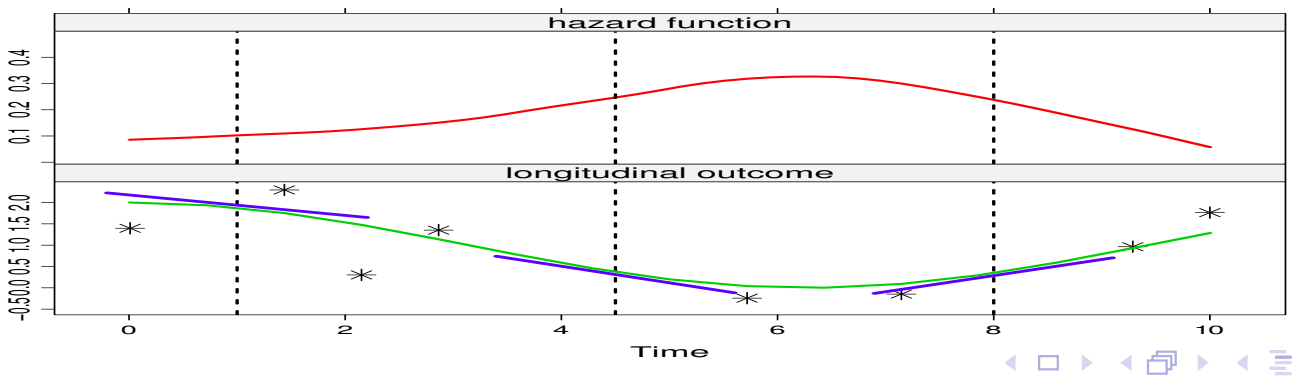
Time-dependent Slopes

The hazard for an event at t is associated with both the current value and the slope of the trajectory at t (Ye et al., 2008, Biometrics):

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\}$$

where

$$m_i'(t) = \frac{d}{dt}\{x_i^T(t)\beta + z_i^T(t)b_i\}$$



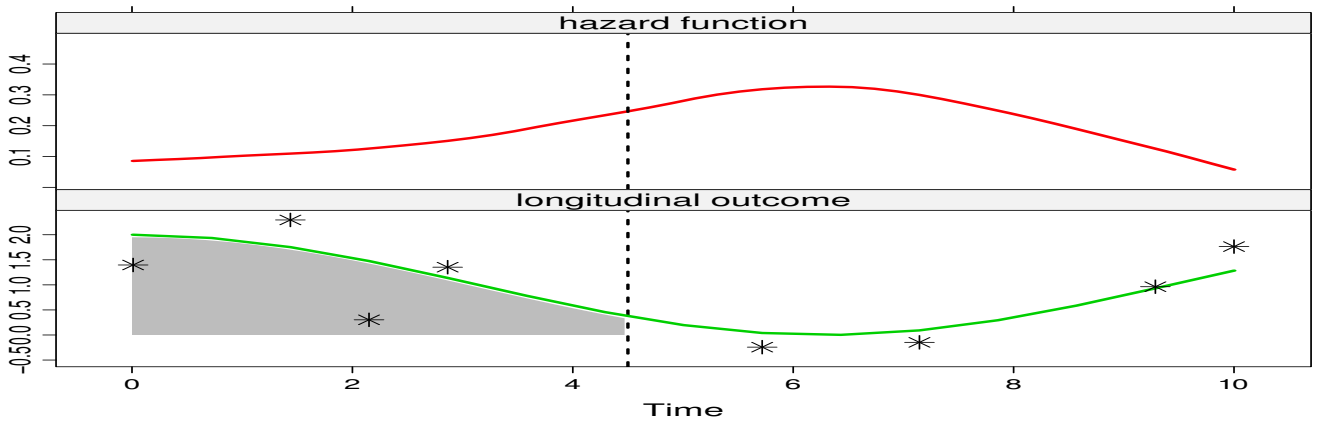
JM: Different parametrizations

Cumulative Effects

The hazard for an event at t is associated with the whole area under the trajectory up to t:

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha \int_0^t m_i(s)ds\}$$

where:
Area under the longitudinal trajectory taken as a summary of $\mathcal{M}_i(t)$



JM: Different parametrizations

Weighted Cumulative Effects (convolution)

The hazard for an event at t is associated with the area under the weighted trajectory up to t:

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha \int_0^t \omega(t-s)m_i(s)ds\}$$

where $\omega(.)$ is an appropriately chosen weight function, e.g., Gaussian, Student, etc.

JM: Different parametrizations

Random Effects

The hazard for an event at t is associated only with the random effects of the longitudinal model:

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha^T b_i\}$$

Features:

- Avoids numerical integration for the survival function
- Interpretation of α more difficult, especially in high-dimensional random-effects settings

JM: Sensitivity to the different parametrizations

- Sensitivity of inferences for the longitudinal process to the choice of the parameterization for the AIDS data

- We use the same mixed model as before, i.e.,

$$\begin{aligned} y_i(t) &= m_i(t) + \epsilon_i(t) \\ &= \beta_0 + \beta_1 t + \beta_2 \{t \times ddl_i\} + b_{i0} + b_{i1} t + \epsilon_i(t) \end{aligned}$$

JM: Sensitivity to the different parametrizations

Assume the following four survival submodels

- Model I (current value)

$$h_i(t) = h_0(t) \exp\{\gamma ddl_i + \alpha_1 m_i(t)\}$$

- Model II (current value + current slope)

$$h_i(t) = h_0(t) \exp\{\gamma ddl_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\}$$

where $m_i'(t) = \beta_1 + \beta_2 ddl_i + b_{i1}$

JM: Sensitivity to the different parametrizations

- Model III (random slope)

$$h_i(t) = h_0(t) \exp\{\gamma ddl_i + \alpha_3 b_{i1})\}$$

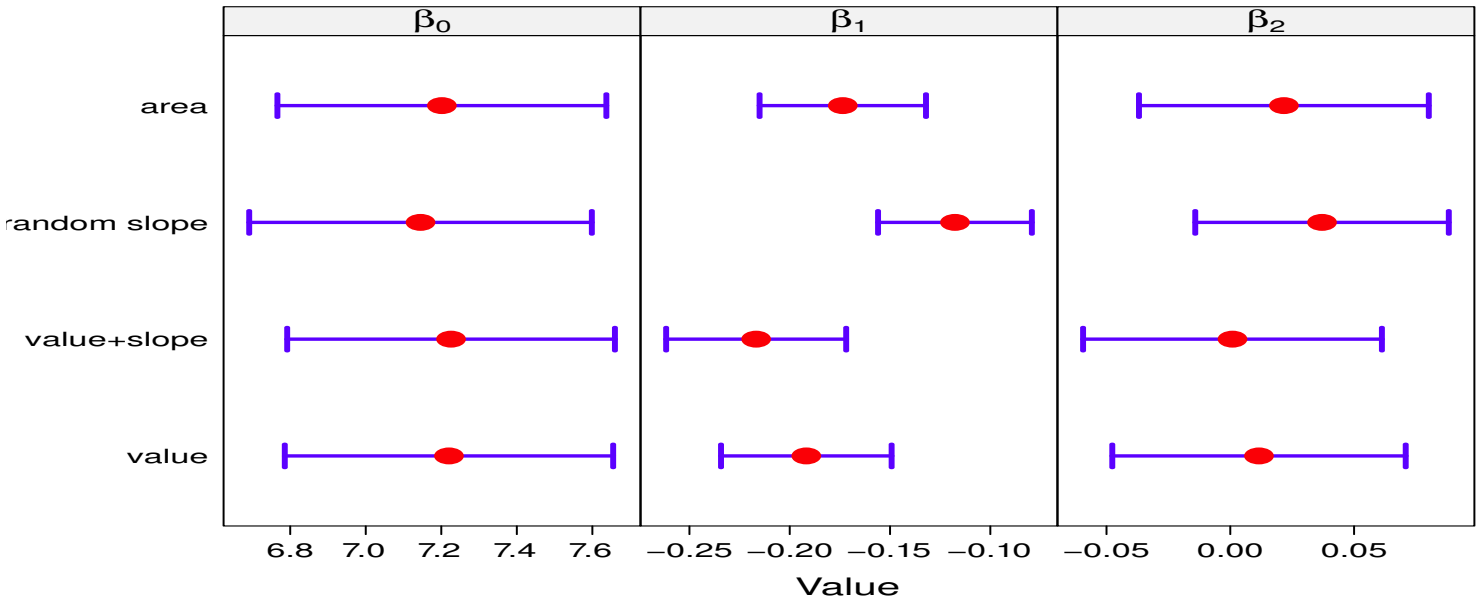
- Model IV (area)

$$h_i(t) = h_0(t) \exp\{\gamma ddl_i + \alpha_4 \int_0^t m_i(s)ds\}$$

where $\int_0^t m_i(s)ds = \beta_0(t) + \frac{\beta_1}{2} t^2 + \frac{\beta_2}{2}\{t^2 \times ddl_i\} + b_{i0} t + \frac{b_{i1}}{2} t^2$

JM: Sensitivity to the different parametrizations

Effect on the LMM component:



Parametrizations: *R* codes

Lagged effects can be fitted using the lag argument of jointModel().

For example, the following code fits a joint model for the PBC dataset with

- Random intercepts and random slopes for log serum bilirubin, and
- A relative risk model with piecewise-constant baseline hazard and the true effect at the previous year

Parametrizations: *R* codes

```
lmeFit <- lme(log(serBilir) ~ year, random = ~ year | id, data=pbc2)

coxFit <- coxph(Surv(years, status2) ~ 1, data = pbc2.id, x = TRUE)

jointFit <- jointModel(lmeFit, coxFit, timeVar = "year",
method = "piecewise-PH-aGH", lag = 1)

summary(jointFit)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar="year",
    method = "piecewise-PH-aGH", lag = 1)

Data Descriptives:
Longitudinal Process Event Process
Number of Observations: 1945 Number of Events: 140 (44.9%)
Number of Groups: 312
```

Parametrizations: *R* codes

```
Joint Model Summary:
Longitudinal Process: Linear mixed-effects model
Event Process: Relative risk model with piecewise-constant
baseline risk function
Parameterization: Time-dependent

   log.Lik       AIC       BIC
 -1920.731 3869.461 3921.863

Variance Components:
             StdDev    Corr
(Intercept)  1.0029  (Intr)
year         0.1760  0.4170
Residual     0.3479
```

Parametrizations: *R* codes

```
Coefficients:
Longitudinal Process
             Value Std.Err z-value p-value
(Intercept) 0.4897  0.0583  8.3966 <0.0001
year        0.1796  0.0131 13.7484 <0.0001

Event Process
                Value Std.Err  z-value p-value
Assoct(lag=1)  1.2938  0.0975  13.2628 <0.0001
log(xi.1)     -4.2262  0.2317 -18.2437
log(xi.2)     -4.0188  0.2417 -16.6302
log(xi.3)     -4.3434  0.2930 -14.8225
log(xi.4)     -4.3256  0.3486 -12.4099
log(xi.5)     -4.0445  0.3201 -12.6339
log(xi.6)     -3.6795  0.3360 -10.9520
log(xi.7)     -4.5785  0.4773  -9.5924

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5
Optimization:
Convergence: 0
```

Parametrizations: *R* codes

- For the time-dependent slopes and cumulative effects parameterizations, arguments parameterization and derivForm of jointModel() should be used

- The first one just specifies whether we want to include a single or two terms involving $m_i(t)$ in the linear predictor of the survival submodel, options are

  - parameterization = ”value”
  - parameterization = ”slope”
  - parameterization = ”both”

- The second one requires a few extra steps to specify

Outline of lecture #2

1. Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
2. Extension of the Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)
3. Bivariate Joint Model for a Recurrent and a Terminal Time-to-Event Outcomes (*R* package Frailtypack)

Bivariate JM for recurrent and terminal events

Goal

- Model dependence between two time-dependent outcomes, e.g. a terminal event and a recurrent event.
- Study the impact of covariates both on recurrent events and death
- To treat informative censoring by terminal event (e.g. death)

Bivariate JM for recurrent and terminal events

Notations

T_i^* : true time of terminal event for subject i, $\quad i = 1, \cdots, N$

T_i : observed terminal event time for patient i, i.e., $T_i = \min(C_i; T_i^*)$

δ_i : Event indicator for the terminal event, i.e., $\delta_i = I_{\{T_i = T_i^*\}}$

T_{ij}^* : true time of the jth recurrent event for subject i

T_{ij} : observed time of the jth recurrent event for subject i
$\Rightarrow T_{ij} = \min(T_{ij}^*, C_i, T_i^*)$ and $\delta_{ij} = I_{\{T_{ij} = T_{ij}^*\}}$

$X_{R_{ij}}$ and X_{T_i} : covariate vectors for recurrent and terminal events
r_0 and λ_0 : baseline hazards for risk of recurrent and terminal events

Bivariate JM for recurrent and terminal events

Formulation

$$\begin{cases} r_{ij}(t|u_i) = u_i r_0(t) \exp(X_{R_{ij}}^T \beta_R), \\ \\ \lambda_i(t|u_i) = u_i^\alpha \lambda_0(t) \exp(X_{T_i}^T \beta_T), \end{cases}$$

Frailty $u_i \sim \Gamma(1/\theta, 1/\theta)$, i.e. $E(u_i) = 1$ and $\text{var}(u_i) = \theta$.

- Heterogeneity between subjects associated with unobserved prognostic factors
- Within-subject correlation for the dependence between the recurrent events
- Association between recurrent events and terminal event $\Rightarrow$ power term α

Bivariate JM for recurrent and terminal events

Likelihood

Parameters to estimate: $\xi = (r_0(.), \lambda_0(.), \beta_R^T, \beta_T^T, \theta, \alpha)$

$$L_i(\xi) = \int_{u_i} \prod_{j=1}^{n_i} f(T_{ij}, \delta_{ij}|u_i;\xi) f(T_i, \delta_i|u_i;\xi) f(u_i;\xi) du_i$$

$$= \int_0^{+\infty} \prod_{j=1}^{n_i} \left[r_{ij}(T_{ij}, \delta_{ij}|u_i;\xi)^{\delta_{ij}} S_{ij}(T_{ij}|u_i;\xi) \right] \lambda_i(T_i, \delta_i|u_i;\xi)^{\delta_i} S_i(T_i|u_i;\xi) f(u_i;\xi) du_i$$

$S_i(j)$: conditional survival function for terminal event
$S_{ij}(j)$: conditional survival function for jth recurrent event
n_i : number of recurrent events for individual i

Bivariate JM for recurrent and terminal events

Marginal log-likelihood (gamma distribution of the frailty)

$$l(\xi) = \sum_{i=1}^{N} \Big(\sum_{j=1}^{n_i} \delta_{ij} \log r_{ij}(T_{ij}|u_i) + \delta_i \log \lambda_i(T_i|u_i) - \log \Gamma(1/\theta) - \frac{1}{\theta} \log \theta$$

$$+ \log \int_0^{+\infty} u_i^{(\sum_{j=1}^{n_i} \delta_{ij} + \alpha\delta_i + 1/\theta - 1)} \exp^{-u_i} \int_0^{T_i} r_{ij}(t|u_i)dt - u_i^{\alpha} \int_0^{T_i} \lambda_i(t)dt - \frac{u_i}{\theta} du_i \Big)$$

Bivariate JM for recurrent and terminal events

Estimation

- Using penalized likelihood (Rondeau, Biostat 2007)
- Using the EM algorithm (Liu, Biometrics 2004)

Bivariate JM for recurrent and terminal events

Baseline risk function

Let $h_0(t)$ be a baseline hazard function.

Weibull baseline hazards (parametric)

- $h_0(t) = (at^{a-1})/b^a$
- with a $>$ 0 the shape parameter and b $>$ 0 the scale parameter

Piecewise constant baseline hazards (parametric)

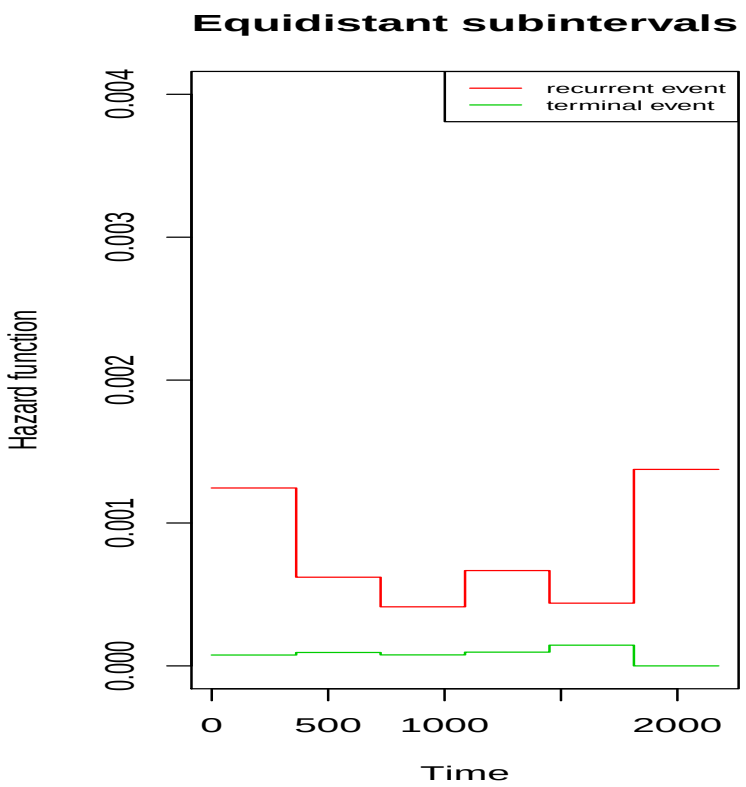
$$h_0(t) = \sum_{i=1}^{n_{int}} I_{\{t \in (t_{i-1}, t_i)\}} c_i$$

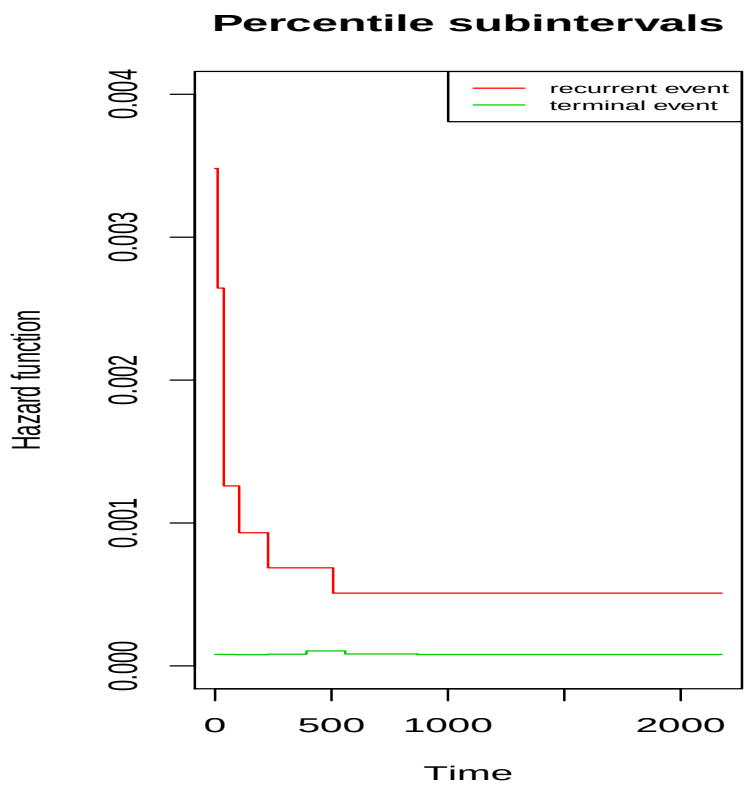
- The interval $[0, \tau]$ (with the last observed time among N individuals) is divided into n_{int} subintervals:
- Equidistant intervals between two knots or percentiles

Bivariate JM for recurrent and terminal events

Baseline risk function

Piecewise constant baseline hazards - Example ($n_{int} = 6$)





Bivariate JM for recurrent and terminal events

Baseline risk function

Baseline hazards approximated with splines (semi-parametric)

- cubic M-splines : adapted for the approximation of hazard functions (Ramsay, 1988)

$$h_0(t) = \sum_{i=1}^{m} \zeta_i M_i(.).$$

with $M_i(.)$ cubic M-splines of order 4,
$\zeta = (\zeta_1, \cdots, \zeta_m)$ is the vector of splines coefficients,
$m = T + 2$ with T the number of knots.

- Antiderivatives, called I-splines : useful for approximating the cumulative hazard function using the same spline parameters ζ
- We can use equidistant or percentiles knots

Bivariate JM for recurrent and terminal events

Estimation

Maximization of log-likelihood: Marquardt algorithm (Marquardt, 1963)

- Combines the Newton-Raphson and the steepest descent algorithms
- More stable behavior than the Newton-Raphson algorithm in complex problems while preserving fast convergence
- Three conditions for convergence
  - whether the coefficients are stable (difference between two consecutive iterations $< 10^{-4}$)
  - a condition on the difference between values of the penalized log-likelihood between two iterations ($< 10^{-4}$)
  - a condition on the gradient of the penalized log-likelihood, whether it is small enough ($< 10^{-4}$)

Bivariate JM for recurrent and terminal events

Maximum penalized likelihood baseline hazards with splines

- In medical studies, it is natural to assume a smooth hazard function with low local variations $\rightarrow$ $l(\xi)$ is penalized by a term that has large values for rough functions

- The penalized log-likelihood is defined as:

$$pl(\xi) = l(\xi) - \kappa_1 \int_0^{+\infty} r_0''^2(t)dt - \kappa_2 \int_0^{+\infty} \lambda_0''^2(t)dt$$

κ_1 and κ_2 control the trade-off between the to the data and the smoothness of the hazard functions.

Bivariate JM for recurrent and terminal events

Maximum penalized likelihood baseline hazards with splines

No automatic algorithm for deciding smoothing parameters, but one can choose

- Graphically by assessing the smoothness of several estimated baseline hazard functions and selecting the one which seems the most realistic
- Using an approximated cross-validation criterion applied to corresponding reduced models (shared frailty model for the recurrent events and the Cox model for the survival) (O'Sullivan, 1988 ; Rondeau et al., 2003)

Bivariate JM for recurrent and terminal events

Available softwares

Limited choice of softwares, to my knowledge :

- SAS: proc nlimixed
- *R* packages:
  - joint.Cox : MPL estimation under the copula-based joint Cox models for time to clustered events (progression) and a terminal event
  - frailtypack

Bivariate JM for recurrent and terminal events

Example: readmission data set

- Rehospitalization of patients after a surgery and diagnosed with colorectal cancer (Gonzalez et al. 2005 ; Rondeau et al. 2012)
- Information on times (in days) of successive hospitalizations and death (or last registered time of follow-up for right-censored patients) counting from date of surgery
- Patients characteristics : type of treatment, sex, Dukes' tumoral stage, comorbidity Charlson's index and survival status
- Includes 403 patients with 861 rehospitalization events in total
- 112 (28%) patients died during the study

Bivariate JM for recurrent and terminal events

Example: readmission data set

- Justification for a joint frailty model for this data ?

`readmission` data set

id	t.start	t.stop	time	event	chemo	sex	dukes	charlson	death
1	0	24	24	1	1	F	D	3	0
1	24	457	433	1	1	F	D	0	0
1	457	1037	580	0	1	F	D	0	0
2	0	489	489	1	0	M	C	0	0
2	489	1182	693	0	0	M	C	0	0
3	0	15	15	1	0	M	C	3	0
3	15	783	768	0	0	M	C	3	1
4	0	163	163	1	1	F	A-B	0	0
4	163	288	125	1	1	F	A-B	0	0
4	288	638	350	1	1	F	A-B	0	0
4	638	686	48	1	1	F	A-B	0	0
4	686	2048	1362	0	1	F	A-B	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Bivariate JM for recurrent and terminal events

Example: readmission data set

- Implementation of joint frailty model

```
# load the package
library(frailtypack)
# load the data
data(readmission)

# fit a joint frailty model using gap time and
    splines for baseline hazards
modJoint.gap <- frailtyPenal(Surv(time, event) ~
  cluster(id) + dukes + charlson + sex + chemo +
  terminal(death), formula.terminalEvent = ~ dukes
  charlson + sex + chemo, data = readmission,
  n.knots = 8, kappa = c(2.11e+08, 9.53e+11))
```

Bivariate JM for recurrent and terminal events

Example: readmission data set

- Output: Recurrent event

```
> modJoint.gap
Call:
frailtyPenal(formula = Surv(time, event) ~ cluster(id) + dukes +
    charlson + sex + chemo + terminal(death), formula.terminalEvent = ~dukes +
    charlson + sex + chemo, data = readmission, n.knots = 8,
    kappa = c(2.11e+08, 9.53e+11))


  Joint gamma frailty model for recurrent and a terminal event processes
  using a Penalized Likelihood on the hazard function

Recurrences:
-------------
                  coef exp(coef) SE coef (H) SE coef (HIH)        z          p
dukesC        0.344028  1.410617    0.165718      0.165718  2.07598 3.7896e-02
dukesD        1.260629  3.527639    0.206285      0.206285  6.11111 9.8938e-10
charlson1-2   0.410971  1.508282    0.256589      0.256589  1.60167 1.0923e-01
charlson3     0.413175  1.511609    0.137136      0.137136  3.01289 2.5878e-03
sexFemale    -0.537464  0.584228    0.141488      0.141488 -3.79865 1.4549e-04
chemoTreated -0.154436  0.856899    0.147137      0.147137 -1.04961 2.9390e-01

           chisq df global p
dukes    41.6554  2 9.01e-10
charlson 11.6428  2 2.96e-03
```

Bivariate JM for recurrent and terminal events

Example: readmission data set

- Output: Terminal event

```
Terminal event:
-----------------
                  coef exp(coef) SE coef (H) SE coef (HIH)         z          p
dukesC        1.318393  3.737410    0.367781      0.367781  3.584724 3.3743e-04
dukesD        3.185326 24.175173    0.432573      0.432573  7.363668 1.7897e-13
charlson1-2   0.507129  1.660516    0.634718      0.634718  0.798983 4.2430e-01
charlson3     1.274563  3.577139    0.257553      0.257553  4.948735 7.4698e-07
sexFemale    -0.234725  0.790789    0.230204      0.230204 -1.019638 3.0790e-01
chemoTreated  1.096094  2.992456    0.258081      0.258081  4.247090 2.1657e-05

           chisq df global p
dukes    61.1967  2 5.14e-14
charlson 24.4911  2 4.81e-06

 Frailty parameters:
   theta (variance of Frailties, w): 0.738406 (SE (H): 0.105178 ) p = 1.1049e-12
   alpha (w^alpha for terminal event): 0.861979 (SE (H): 0.250164 ) p = 0.00056968

   penalized marginal log-likelihood = -4133.31
   Convergence criteria:
   parameters = 1.18e-05 likelihood = 4.88e-05 gradient = 1.21e-07

   LCV = the approximate likelihood cross-validation criterion
         in the semi parametric case      = 4.84003

   n observations= 861  n subjects= 403
   n recurrent events= 458
   n terminal events= 109
   n censored events= 403
   number of iterations:  9

   Exact number of knots used:  8
   Value of the smoothing parameters:  2.11e+08 9.53e+11
```

Bivariate JM for recurrent and terminal events

Model fit evaluation

- Approximate likelihood cross-validation criterion (LCV)
- Measures the relative goodness of fit among a collection of models
- Lower values indicate a better fitting

$$LCV_a = \frac{1}{\sum_{i=1}^{N} r_i}(\text{trace}(H_{pl}^{-1} H_l) - l(.))$$

  - H_{pl} is minus times the converged hessian of the penalized log-likelihood,
  - H_l minus times the converged hessian of the log-likelihood
  - $l(.)$ is the full log-likelihood

Bivariate JM for recurrent and terminal events

Example: readmission data set

- Example readmission data set : gamma vs. log-normal joint frailty model

```
# gamma joint frailty model  (default option)
m_gamma <- frailtyPenal(Surv(time, event) ~ cluster(id)
 + dukes + sex + chemo + terminal(death),
 formula.terminalEvent =~ dukes + sex + chemo,n.knots=8,
 data = readmission, kappa = c(2.1e+08, 9.5e+11))
m_gamma$LCV    # LCV = 4.852029

# log-normal joint frailty model
m_logN <- frailtyPenal(Surv(time, event) ~ cluster(id)
 + dukes + sex + chemo + terminal(death),
 formula.terminalEvent =~ dukes +sex + chemo, n.knots=8,
 data = readmission, kappa = c(2.1e+08, 9.5e+11),
 RandDist = 'LogN')
m_LogN$LCV    # LCV = 4.848132
```

Bivariate JM for recurrent and terminal events

Martingale residuals

- Models whether the number of observed events is correctly predicted by a model
- Based on the counting processes theory
- The martingale residuals can be expressed by (recurrences)

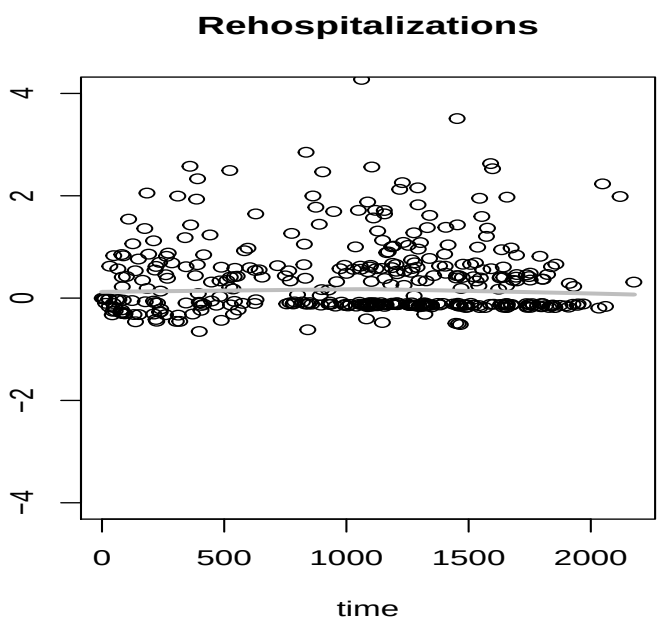
$$M_i(t) = N_i(t) - \hat{u}_i \int_0^t W_i(s)\hat{r}_i(s)ds$$

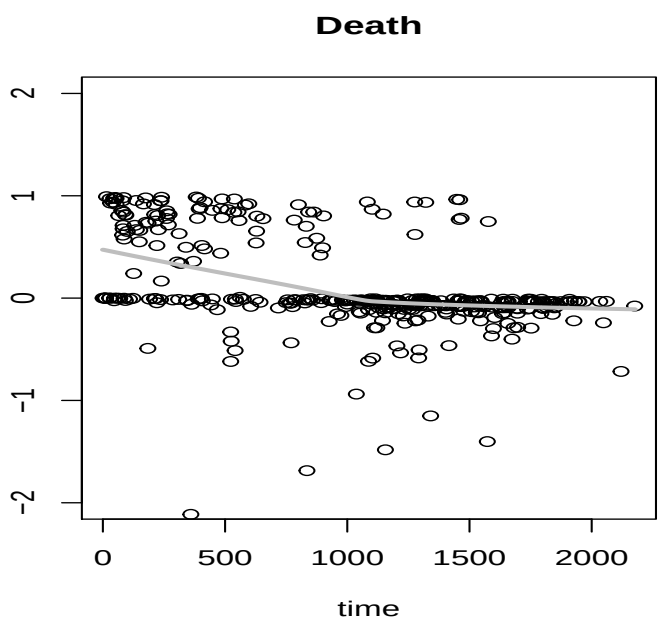
where $W_i(t)$ is equal to 1 if the individual is at risk of the event at time t and 0 otherwise, $N_i(t)$ is number of events until t, and $\hat{r}_i(t) = u_i r_i(t|u_i)$.

Bivariate JM for recurrent and terminal events

Martingale residuals

Calculated at T_i, the mean of the martingale residuals at a given time point should be equal to 0





Bivariate JM for recurrent and terminal events

Time-dependent variables

- B-splines

$$\hat{\beta}(t) = \sum_{j=-q+1}^{m} \hat{\zeta}_j B_{j,q}(t)$$

where $B_{j,q}(t)$ is the basis of B-splines calculated using a recurring expression

- Quadratic B-splines, i.e., $q = 3$, with a small number of interior knots ($m \leq 5$) ensure stable estimation

Bivariate JM for recurrent and terminal events

Time-dependent variables

- Helpful to verify the proportional hazard (PH) assumption using a likelihood ratio (LR) test
- Time-dependency tested for both events :

$$H_0 : \beta_R(t) = \beta_R, \quad \beta_T(t) = \beta_T \text{ vs. } H_1 : \beta_R(t) \neq \beta_R, \quad \beta_T(t) \neq \beta_T$$

- The LR statistic has a χ^2 distribution of degree $2 \times (m + q - 1)$

Bivariate JM for recurrent and terminal events

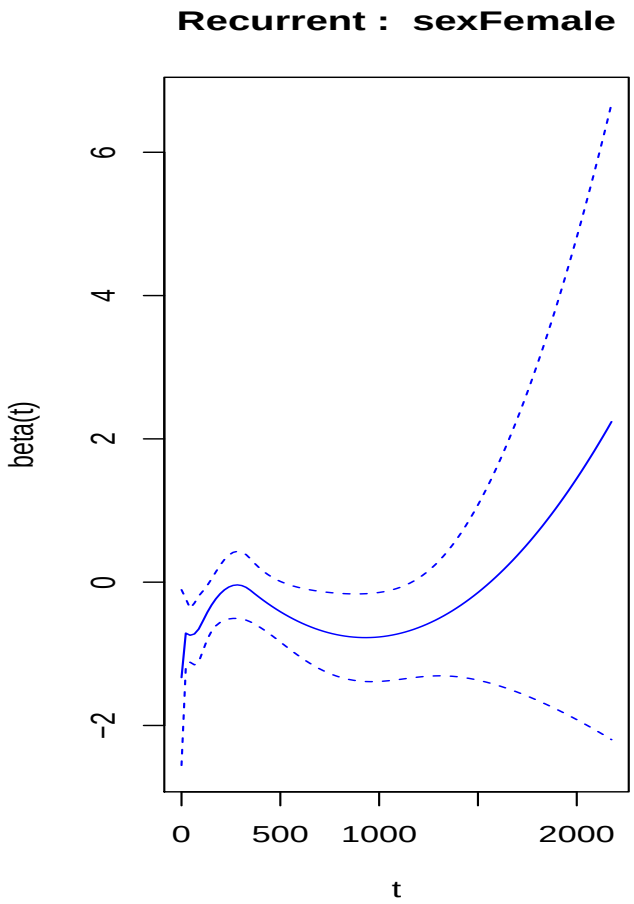
Readmission data

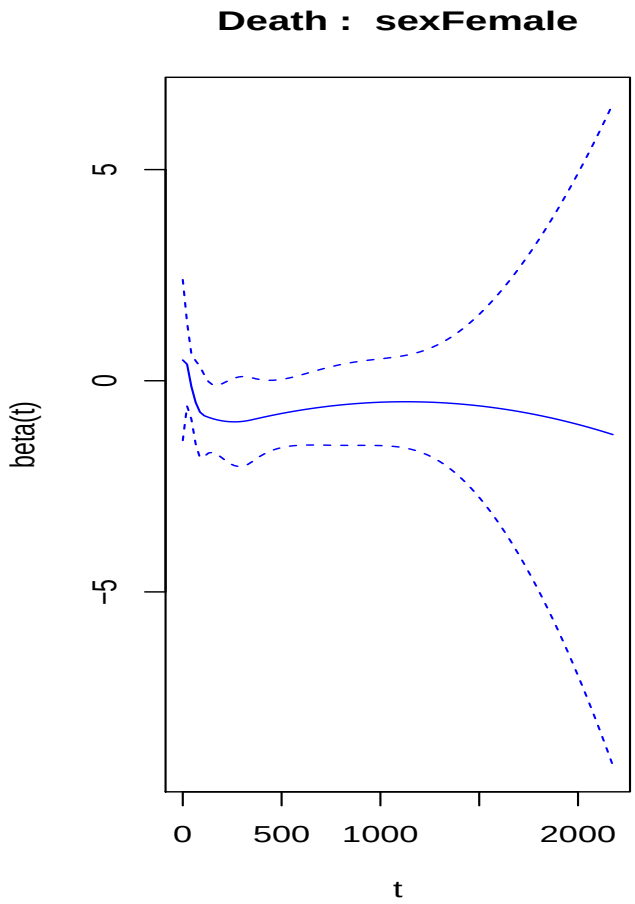
```
# Model with time-dependent effect of gender
modJoint.gap.timedep <- frailtyPenal(Surv(time, event) ~
  cluster(id) + dukes + charlson + timedep(sex) + chemo
  + terminal(death), formula.terminalEvent = ~ dukes +
  timedep(sex) + chemo, data = readmission, n.knots = 8,
  kappa = c(2.11e+08, 9.53e+11), betaorder = 3,
  betaknots = 3)

# LR test
LR.statistic <- -2 * modJoint.gap$logLik +
 2 * modJoint.gap.timedep$logLik
p.value <- signif(1 - pchisq(LR.statistic, df = 10), 5)
# p.value = 0.049
```

Bivariate JM for recurrent and terminal events

Readmission data





Conclusion

Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome are useful for

- Account for endogenous time-dependent variables
- Model joint association between a longitudinal and time-to-event processes
- Account for informative dropout from longitudinal process
- Get better estimation

Conclusion

Bivariate Joint Model for a Recurrent and a Terminal Time-to-Event Outcomes are useful for:

- Account for endogenous time-dependent variables
- When the interest is to analyze strength of the association between the survival and recurrent events processes
- When the process of recurrent events is informatively censored by a terminal event
- Get better estimation
- Specific applications to meta-analyses and multi-centre studies