Week 3: Case Studies, *R* package JM and Extensions of the Bivariate Joint Model

Laurent Briollais (laurent@lunenfeld.ca)

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital
Dalla Lana School of Public Health, UofT

May 20, 2020

Previous lecture

- Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)

- Extension of the Bivariate Joint Model for a Longitudinal Outcome and a Terminal Time-to-Event Outcome (*R* package JM)

Outline of this lecture

- Case Study = AIDS data
- Case Study = PBC data
- Extensions of the Bivariate Joint Models
- Presentation of the *R* package JM

References

- **Joint Models for Longitudinal and Time-to-Event Data With Applications in** *R* from Dimitris Rizopoulos. Chapman & Hall/CRC biostatistics series. CRC Press, Taylor & Francis Group. 2012.

  - Chapter 4 includes extensive comparisons between JM and the extended Cox model.
  - Chapter 5 presents the different parametrizations of the biomarker, interaction effects, stratified models and competing risks model.
  - Chapter 6 includes the JM related diagnostic plots.

1 Case Studies: AIDS data

1.1 Data description

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)

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

```
head(aids[c("patient", "start", "stop", "event", "CD4", "obstime", "drug")], 7)
```

```
  patient start  stop event       CD4 obstime drug
1       1     0  6.00     0 10.677078       0  ddC
2       1     6 12.00     0  8.426150       6  ddC
3       1    12 16.97     0  9.433981      12  ddC
4       2     0  6.00     0  6.324555       0  ddI
5       2     6 12.00     0  8.124038       6  ddI
6       2    12 18.00     0  4.582576      12  ddI
7       2    18 19.00     0  5.000000      18  ddI
```

1.2 Comparison extended Cox model vs. JM

- Extended Cox model

$$h_i(t) = h_0(t) R_i(t) \exp\{\gamma ddI_i + \alpha y_i(t)\},$$

```
data('aids')
td.Cox<-coxph(Surv(start, stop, event)~drug+CD4, data=aids)
summary(td.Cox)

Call:
coxph(formula = Surv(start, stop, event) ~ drug + CD4, data = aids)

  n= 1405, number of events= 188

            coef exp(coef) se(coef)      z Pr(>|z|)
drugddI  0.30948   1.36271  0.14653  2.112   0.0347 *
CD4     -0.19343   0.82412  0.02437 -7.937 2.08e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

        exp(coef) exp(-coef) lower .95 upper .95
drugddI    1.3627     0.7338    1.0225    1.8161
CD4        0.8241     1.2134    0.7857    0.8644

Concordance= 0.696  (se = 0.018 )
Likelihood ratio test= 94.62  on 2 df,   p=<2e-16
Wald test            = 65.7  on 2 df,   p=5e-15
Score (logrank) test = 73.34  on 2 df,   p=<2e-16

>
```

$\Rightarrow$ the CD4 cell count has a strong association with the risk for death, where a unit decrease in CD4 cell count corresponds to a $\exp(-(-0.19)) = 1.20$ fold increase in the risk for death (95% CI: 1.16-1.27).

$\Rightarrow$ the drug ddI has also a significant association with the risk of death, where having the drug ddI vs. ddC corresponds to a $\exp(0.31) = 1.36$ fold increase in the risk for death (95% CI: 1.02-1.82).

- JM to account for the endogeneity of CD4 cell count

```
aids.id<-aids[!duplicated(aids$patient),]

lmeFit <- lme(CD4 ~ obstime + obstime:drug,
random = ~ obstime | patient, data = aids)
coxFit <- coxph(Surv(Time, death) ~ drug, data = aids.id, x = TRUE)
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
method = "piecewise-PH-aGH")
summary(jointFit)
```

- The argument *timeVar* of *jointModel()* is used to specify the name of the time variable in the LMM, which is required in the internal computation of $m_i(t)$.

- The *method* argument specifies the type of baseline risk function.

```
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

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 of results: JM vs. Cox model

Table 1: Parameter estimates for time to death

	JM			Cox PH		
	logHR	SE	p-value	logHR	SE	p-value
Treat	0.33	0.16	0.032	0.31	0.15	0.035
CD4	-0.29	0.04	< 0.0001	-0.19	0.02	< 0.0001

- Induced causal relationships

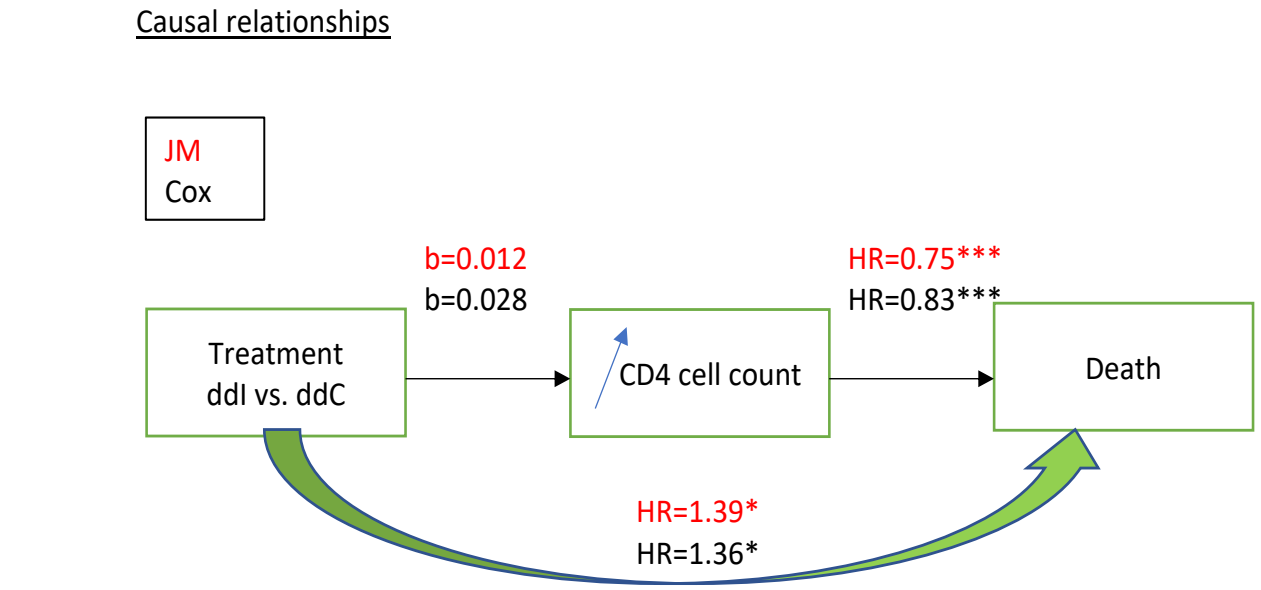
Remember:

Table 2: Parameter estimates for CD4 cell count

	JM			LMM		
	β	SE	p-value	β	SE	p-value
Time	-0.19	0.02	< 0.0001	-0.16	0.02	< 0.0001
Time $\times$ ddI	0.012	0.03	0.70	0.028	0.03	0.34

linear mixed model

Figure 1: Causal relationships in AIDS data



direct effect: residual direct effect
indirect effect?

- Conclusion: The magnitude of CD4 cell count effect on the overall survival is reduced in the Cox model compared to the JM.

  can't do causal inference: conditional on random effects
  relationship/associations are time-dependent

1.3 Effect of the baseline hazard function

Available options are:

- 'piecewise-PH-GH': PH model with piecewise-constant baseline hazard

5 knots: 7 parameters

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

The interval $[0, \tau]$ (with the last observed time among N individuals) is divided into n_{int} subintervals (6 by default). To change the default number, the control argument *lng.in.kn* can be used, whereas the argument *knots* control the position of the knots.

- 'spline-PH-GH': PH model with B-spline-approximated log baseline hazard

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

The default number of knots is 5 with equally spaced percentiles of the observed event times. The control argument *lng.in.kn* can be used to finer control.

- 'weibull-PH-GH': PH model with Weibull baseline hazard

parametric

$$h_0(t) = \sigma_t t^{\sigma_t - 1}.$$

- 'weibull-AFT-GH': AFT model with Weibull baseline hazard

$$\log T_i^* = \gamma^T w_i + \epsilon_i$$

- 'Cox-PH-GH': PH model with unspecified baseline hazard

$$h_0(t) = \begin{cases} \xi_q, & t = T_q^* \\ 0, & t \neq T_q^*, \end{cases}$$

where T_q^* denotes the qth unique true event time. It is equivalent to assuming $h_0(.)$ is discrete with point masses at the unique event times, i.e.,

different baseline hazards

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

## Baseline = spline
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "spline-PH-GH")
summary(jointFit)

## Baseline = Unspecified
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "Cox-PH-GH")
summary(jointFit)

## Baseline = Weibull-PH-GH
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "weibull-PH-GH")
summary(jointFit)

## Baseline = Weibull-AFT-GH
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "weibull-AFT-GH")
summary(jointFit)
```

Table 3: Parameter estimates for time to death

	Treatment	Assoc	
	Estimate (SE)	Estimate (SE)	AIC
Piecewise hazard	0.33 (0.16)	-0.29 (0.04)	8688
Spline	0.35 (0.16)	-0.29 (0.04)	8706
Unspecified	0.32 (0.13)	-0.27 (0.03)	9549
Weibull	0.36 (0.16)	-0.29 (0.04)	8701
Weibull AFT	0.28 (0.12)	-0.23 (0.04)	8702

not proportional hazard - AFT is a bit different

may use AIC to select which one is the best

```
## Baseline = piecewise 5 knots
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "piecewise-PH-aGH")
summary(jointFit)

## Baseline = piecewise 10 knots
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "piecewise-PH-aGH", lng.in.kn=10)
summary(jointFit)
```

Table 4: Parameter estimates for time to death

	Treatment	Assoc	
	Estimate (SE)	Estimate (SE)	AIC
Piecewise hazard (5 knots)	0.33 (0.16)	-0.29 (0.04)	8688
Piecewise hazard (10 knots)	0.34 (0.16)	-0.28 (0.04)	8691

1.4 Control of numerical integration for the survival functions and random effects

- The function *jointModel()* provides 3 control arguments that allow for a fine control of the numerical integration algorithms.
- The argument *GKk* controls the number of Gauss-Kronrod quadrature points of the survival function.
- The argument *GHk* controls the number of quadrature points of the Gauss-Hermite algorithm.
- The first 2 parts of the character string supplied in the *method* argument specify the baseline risk function and the type of survival model, and the last part the type of numerical integration. GH stands for the Gauss-Hermite rule and aGH for pseudo-adaptive rule.
- Examples:

if you have convergence, try these.
If not, stay with default

```
jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "piecewise-PH-aGH", GHk=15)

jointFit <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                       method = "piecewise-PH-aGH", GHk=21, GKk=15)
```

1.5 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.,

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

```
aids.id<-aids[!duplicated(aids$patient),]
lmeFit <- lme(sqrt(CD4) ~ obstime + obstime:drug,
              random = ~ obstime | patient, data = aids)
```

and the following five survival submodels

- Model I (current value - default)

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

```
coxFit <- coxph(Surv(Time, death) ~ drug, data = aids.id, x = TRUE)
jointFit1 <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                        method = "piecewise-PH-aGH")
summary(jointFit1)

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
 -2101.983 4235.965 4302.307

Variance Components:
             StdDev    Corr
(Intercept)  0.8699  (Intr)
obstime      0.0371  0.0834
Residual     0.3673

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5145  0.0425 59.1329 <0.0001
obstime         -0.0423  0.0045 -9.3082 <0.0001
obstime:drugddI  0.0052  0.0064  0.8092  0.4184

Event Process
            Value Std.Err z-value p-value
drugddI    0.3484  0.1533  2.2723  0.0231
Assoct    -1.0906  0.1163 -9.3736 <0.0001
log(xi.1) -1.6711  0.2477 -6.7461
log(xi.2) -1.3556  0.2381 -5.6942
log(xi.3) -1.0401  0.2848 -3.6513
log(xi.4) -1.5962  0.3727 -4.2829
log(xi.5) -1.4869  0.3490 -4.2597
log(xi.6) -1.4506  0.4274 -3.3937
log(xi.7) -1.4847  0.5436 -2.7311

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

of the biomarkers

- Model II (current value + current slope)

slope of the biomarkers - speed of progression of CD4 cell count

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

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

To fit this model we need to specify the 'derivForm' argument, which is a list with first component the derivative of the fixed-effects formula of 'lmeFit' with respect to 'obstime', second component the indicator of which fixed-effects coefficients correspond to the previous defined formula, third component the derivative of the random-effects formula of 'lmeFit' with respect to 'obstime', and fourth component the indicator of which random-effects correspond to the previous defined formula.

```
dForm <- list(fixed = ~ 1 + drug, indFixed = c(2, 3), random = ~ 1, indRandom = 2)
jointFit2<-jointModel(lmeFit, coxFit, timeVar = "obstime", method = "spline-PH-aGH",
           parameterization = "both", derivForm = dForm)

summary(jointFit2)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    parameterization = "both", method = "spline-PH-aGH", derivForm = dForm)

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 spline-approximated
baseline risk function
Parameterization: Time-dependent + time-dependent slope

  log.Lik      AIC     BIC
 -2095.27 4228.539 4307.32

Variance Components:
             StdDev    Corr
(Intercept)  0.8704  (Intr)
obstime      0.0369  0.0968
```

```
Residual     0.3678

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5147  0.0426 59.0379 <0.0001
obstime         -0.0429  0.0049 -8.6995 <0.0001
obstime:drugddI  0.0048  0.0064  0.7548  0.4504

Event Process
            Value Std.Err z-value p-value
drugddI    0.3687  0.1591  2.3171  0.0205
Assoct    -1.0732  0.1210 -8.8722 <0.0001
Assoct.s  -2.8447  5.7456 -0.4951  0.6205
bs1       -3.1669  0.6800 -4.6569 <0.0001
bs2       -0.4143  0.6925 -0.5983  0.5497
bs3       -3.2779  0.7468 -4.3890 <0.0001
bs4       -0.5019  0.5378 -0.9333  0.3507
bs5       -2.1079  0.5808 -3.6292  0.0003
bs6       -1.2158  0.8094 -1.5020  0.1331
bs7       -2.1258  1.4947 -1.4222  0.1550
bs8        0.0168  2.9599  0.0057  0.9955
bs9       -8.0163  9.0693 -0.8839  0.3768

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

association of current value of biomarkers: increase in CD4 cell count decrease the chance of dying.
association of current slope: not significant

model is convergence
if it's 1, it means you have a convergence problem.

- Model III (random slope)

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

```
dForm <- list(fixed = ~ 1, indFixed = c(2), random = ~ 1, indRandom = 2)
jointFit3<-jointModel(lmeFit, coxFit, timeVar = "obstime", method = "spline-PH-aGH",
                      parameterization = "slope", derivForm = dForm)

summary(jointFit3)

Call:
```

```
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    parameterization = "slope", method = "spline-PH-aGH", derivForm = dForm)

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 spline-approximated
baseline risk function
Parameterization: Time-dependent slope

   log.Lik      AIC      BIC
 -2121.845 4279.689 4354.323

Variance Components:
             StdDev    Corr
(Intercept)  0.8255  (Intr)
obstime      0.0381  0.5317
Residual     0.3837

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5227  0.0407 61.9196 <0.0001
obstime         -0.0484  0.0049 -9.8429 <0.0001
obstime:drugddI  0.0023  0.0065  0.3590  0.7196

Event Process
            Value Std.Err  z-value p-value
drugddI    0.2701  0.1755   1.5383  0.1240
Assoct.s -27.4063  4.5162  -6.0685 <0.0001
bs1       -7.1770  0.7231  -9.9248 <0.0001
bs2       -4.2147  0.6864  -6.1399 <0.0001
bs3       -6.8020  0.7007  -9.7074 <0.0001
bs4       -3.7536  0.4652  -8.0685 <0.0001
bs5       -5.2156  0.5042 -10.3439 <0.0001
bs6       -4.3429  0.7565  -5.7404 <0.0001
bs7       -4.7486  1.4110  -3.3654  0.0008
bs8       -3.3367  2.6684  -1.2504  0.2111
bs9       -7.5859  6.4539  -1.1754  0.2398

Integration:
method: (pseudo) adaptive Gauss-Hermite
```

random effects will have a normal dist with mean = 0, standard deviation of 0.04

can switch to the normal Gauss-Hermite

```
quadrature points: 5

Optimization:
Convergence: 1
```

Note: The previous model did not converge

⇒ Convergence: 1

We refit the model with more iterations iter.qN=500 and using Gauss-Hermite integration algorithm

```
jointFit3<-jointModel(lmeFit, coxFit, timeVar = "obstime", method = "spline-PH-GH",
                      parameterization = "slope", derivForm = dForm, iter.qN=500)

summary(jointFit3)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    parameterization = "slope", method = "spline-PH-GH", derivForm = dForm,
    iter.qN = 500)

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 spline-approximated
baseline risk function
Parameterization: Time-dependent slope

   log.Lik      AIC      BIC
 -2128.688 4293.375 4368.009

Variance Components:
             StdDev    Corr
(Intercept)  0.8215  (Intr)
obstime      0.0417  0.4883
Residual     0.3906
```

```
Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5535  0.0321 79.5239 <0.0001
obstime         -0.0498  0.0051 -9.8409 <0.0001
obstime:drugddI  0.0039  0.0067  0.5898  0.5554

Event Process
            Value Std.Err  z-value p-value
drugddI    0.2888  0.1819   1.5873  0.1124
Assoct.s -26.9728  4.6533  -5.7965 <0.0001
bs1       -7.2946  0.7668  -9.5129 <0.0001
bs2       -4.3234  0.7157  -6.0406 <0.0001
bs3       -6.9239  0.7252  -9.5476 <0.0001
bs4       -3.8530  0.4838  -7.9645 <0.0001
bs5       -5.2167  0.5132 -10.1645 <0.0001
bs6       -4.3433  0.7540  -5.7602 <0.0001
bs7       -4.7304  1.3909  -3.4009  0.0007
bs8       -3.3588  2.6362  -1.2741  0.2026
bs9       -7.7152  6.3591  -1.2133  0.2250

Integration:
method: Gauss-Hermite
quadrature points: 15

Optimization:
Convergence: 0
```

everything is fine

- Model IV (area)

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

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

```
cForm <- list(fixed = ~ -1 + obstime + I(obstime^2/2) + I((obstime^2 * (drug=="ddI"))/2),
              indFixed = 1:3, random = ~-1 + obstime + I(obstime^2/2), indRandom = 1:2)
jointFit4<-jointModel(lmeFit, coxFit, timeVar = "obstime", method = "spline-PH-GH",
                      parameterization = "slope", derivForm = cForm)
```

```
summary(jointFit4)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    parameterization = "slope", method = "spline-PH-GH", derivForm = cForm)

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 spline-approximated
baseline risk function
Parameterization: Time-dependent slope

   log.Lik      AIC      BIC
 -2116.463 4268.927 4343.561

Variance Components:
             StdDev    Corr
(Intercept)  0.8686  (Intr)
obstime      0.0383  0.0023
Residual     0.3752

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5587  0.0381 67.2408 <0.0001
obstime         -0.0394  0.0046 -8.6512 <0.0001
obstime:drugddI  0.0066  0.0065  1.0111  0.3120

Event Process
           Value Std.Err z-value p-value
drugddI   0.3416  0.1502  2.2741  0.0230
Assoct.s -0.1018  0.0132 -7.7221 <0.0001
bs1      -5.2835  0.5549 -9.5215 <0.0001
bs2      -1.9026  0.5706 -3.3344  0.0009
bs3      -3.6702  0.6372 -5.7595 <0.0001
bs4      -0.0388  0.4479 -0.0867  0.9309
bs5      -0.9667  0.5358 -1.8042  0.0712
bs6      -0.1028  0.8081 -0.1272  0.8988
bs7       0.4689  1.4811  0.3166  0.7516
bs8       0.6588  3.0512  0.2159  0.8291
bs9      -4.4165  9.5502 -0.4625  0.6438
```

association of biomarkers and time to death
(cumulative associated effects)
biomarkers are decreasing the chance of dying

```
Integration:
method: Gauss-Hermite
quadrature points: 15

Optimization:
Convergence: 0
```

- Model V: Weighted Cumulative Effects (convolution):

$$h_i(t) = h_0(t) \exp\{\gamma ddI_i + \alpha_5 \int_0^t w(t-s) m_i(s) ds\}$$

We will use the normal function: $w(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

In order to construct the weighted cumulative effect, we are required to evaluate integrals of the form:

$$\int_0^t w(t-s) s^j ds = \frac{1}{\sqrt{2\pi}} \int_0^t s^j \exp\{-(t-s)^2/2\} ds,$$

for $j = 0, 1$ because remember:

$m_i(t) = \beta_0 + \beta_1 t + \beta_2\{t \times ddI_i\} + b_{i0} + b_{i1} t.$

These integrals do not have a closed-form so we need to use the *integrate*() function in *R*. We define the function $g(.)$ that calculates the integrals numerically.

putting more weight at the end of the study

```
g<-function(u, pow=0){
  f<-function(t)
    integrate(function(s) s^pow*dnorm(t-s),0,t)$value
  sapply(u,f)
}
```

remove the default of intercept

```
cFormW <- list(fixed = ~ -1 + I(g(obstime)) + I(g(obstime,1))
               + I(g(obstime,1)*(drug=="ddI")), indFixed = 1:3,
               random = ~-1 + I(g(obstime)) + I(g(obstime,1)), indRandom = 1:2)
jointFit5<-jointModel(lmeFit, coxFit, timeVar = "obstime", method = "spline-PH-GH",
                      parameterization = "slope", derivForm = cFormW)

summary(jointFit5)
```

weighted cumulative effect

```
Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    parameterization = "slope", method = "spline-PH-GH", derivForm = cFormW)

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 spline-approximated
baseline risk function
Parameterization: Time-dependent slope

   log.Lik      AIC      BIC
 -2102.956 4241.913 4316.547

Variance Components:
             StdDev    Corr
(Intercept)  0.8683  (Intr)
obstime      0.0387  0.0494
Residual     0.3750

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5554  0.0374 68.3079 <0.0001
obstime         -0.0414  0.0046 -9.0433 <0.0001
obstime:drugddI  0.0054  0.0065  0.8344  0.4040

Event Process
```

```
           Value Std.Err z-value p-value
drugddI   0.3512  0.1528  2.2983  0.0215
Assoct.s -2.1819  0.2384 -9.1510 <0.0001
bs1      -4.5668  0.5111 -8.9350 <0.0001
bs2       0.5956  0.6131  0.9714  0.3313
bs3      -3.5503  0.6318 -5.6192 <0.0001
bs4      -0.1037  0.4147 -0.2501  0.8025
bs5      -1.9903  0.4762 -4.1798 <0.0001
bs6      -0.9439  0.7402 -1.2753  0.2022
bs7      -1.9685  1.4528 -1.3550  0.1754
bs8       0.0886  2.8690  0.0309  0.9754
bs9      -7.1028  8.3292 -0.8528  0.3938

Integration:
method: Gauss-Hermite
quadrature points: 15

Optimization:
Convergence: 0
```

for the weighted cumulative effect
one unit increase in CD4 will decrease the risk of dying

- Effects of parametrization on LMM and event time components

	Model I Current value	Model II Current value + Current slope	Model III Random slope	Model IV Cumulative	Model V Weighted Cum.
AIC	4236.0	4228.5	4293.4	4268.9	4241.9
LMM component					
Time	-0.04 (0.004) $p < 0.0001$	-0.04 (0.005) $p < 0.0001$	-0.05 (0.005) $p < 0.0001$	-0.04 (0.005) $p < 0.0001$	-0.04 (0.005) $p < 0.0001$
Time × drug	0.005 (0.006) $p = 0.42$	0.005 (0.006) $p = 0.45$	0.005 (0.004) $p = 0.55$	0.007 (0.007) $p = 0.31$	0.005 (0.006) $p = 0.40$
Event time component					
Drug	0.35 (0.15) $p = 0.023$	0.37 (0.16) $p = 0.020$	0.29 (0.18) $p = 0.11$	0.34 (0.15) $p = 0.023$	0.35 (0.15) $p = 0.021$
Assoc.	-1.09 (0.11) $p < 0.0001$	-1.07 (0.12) $p < 0.0001$	- -	- -	- -
Assoc. (slope)	- -	-2.85 (5.75) $p = 0.62$	-27.0 (4.65) $p < 0.0001$	-0.10 (0.01) $p < 0.0001$	-2.18 (0.24) $p < 0.0001$

Which model is the best?

According to AIC, Model II is best but the current slope parameter is not significant in this model. What's wrong?

1.6 Diagnostic plots

- To assess different formulations of the JM, residual plots can be used.

- For JMs, they are available for both the longitudinal and time-to-event components.

- The default plot() function provides four types of plots:

  - Residuals vs. fitted values from LMM
  - QQ plots of residuals from LMM
  - Marginal survival function
  - Marginal cumulative hazard function

- Example in Figure 2:

```
pdf("Model1_plot.pdf")
par(mfrow=c(2,2))
plot(jointFit1)
dev.off()
```

- For the LMM component, the standardized marginal residuals vs. marginal fitted values $X_i\hat{\beta}$ can be obtained with the commands

```
resMargY.aids<-residuals(jointFit1, process="Longitudinal", type="Marginal")
fitMargY.aids<-fitted(jointFit1, process="Longitudinal", type="Marginal")

plotResid<-function(x,y, col.loess="black",...){
  plot(x,y,...)
  lines(lowess(x,y), col=col.loess, lwd=2)
  abline(h=0, lty=3, col="grey", lwd=)
}

plotResid(fitMargY.aids, resMargY.aids, xlab="Fitted values", ylab="Marginal residuals")
```

Figure 2: Diagnostic plots for Model I: Current value of the biomarker

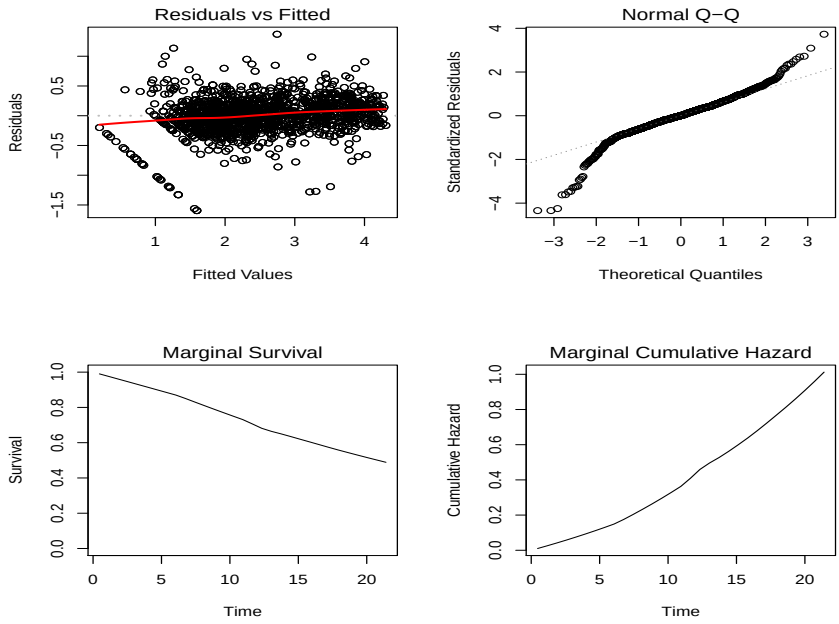
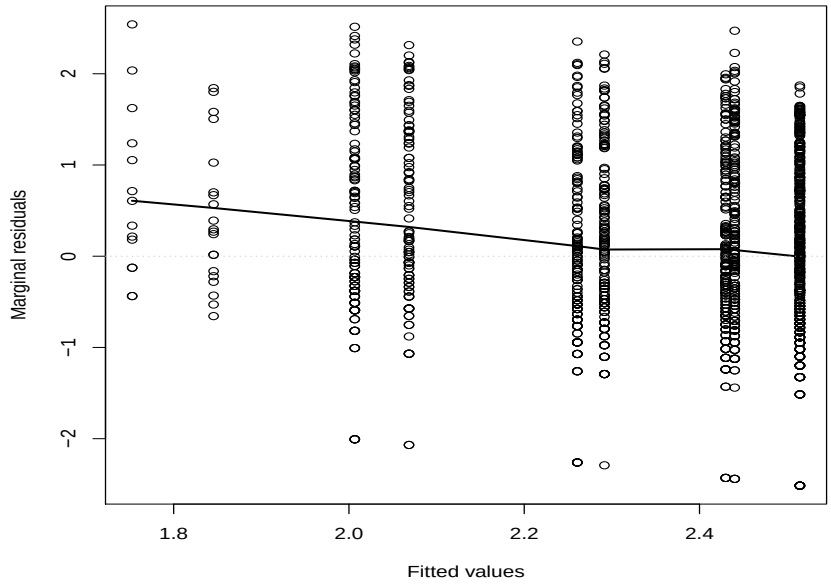


Figure 3: Diagnostic plots for Model I: Fitted values vs. marginal residuals



- For the survival component, the martingale residuals can be used. They are defined as:

$$r_i^{tm}(t) = N_i(t) - \int_0^t R_i(s) h_i(s|\hat{\mathcal{M}}_i(s); \hat{\theta}) ds$$

- They can be viewed as the difference between the observed number of events for ith subject by time t and the expected number by the same time based on the fitted model.

- They can help identify individuals who are not well fitted by the model and evaluate appropriate functional forms for the covariates

- Another type of residuals for survival models is the Cox-Snell residuals. They are calculated as the value of the estimated cumulative risk function evaluated at the observed event time T_i

difference here: hazard depends on the biomarkers

$$r_i^{tcs} = \int_0^{T_i} h_i(s|\hat{\mathcal{M}}_i(s); \hat{\theta}) ds$$

$$= \int_0^{T_i} \hat{h}_0(s) \exp\{\hat{\gamma}^T w_i + \hat{\alpha} \hat{m}_i(s)\} ds.$$

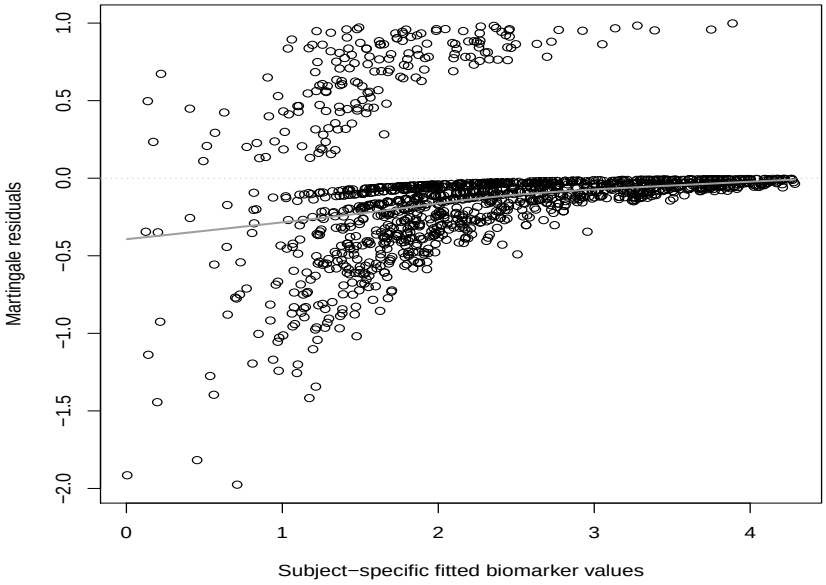
- For a model that is well fitted, the Cox-Snell residuals are unit exponentially distributed.

- Example:

```
martRes <-residuals(jointFit1, process="Event")
mi.t <-fitted(jointFit1, process="Longitudinal", type="EventTime")
plotResid(mi.t, martRes, col.loess="grey62", ylab="Martingale residuals",
xlab="Subject-specific fitted biomarker values")
```

- The option type="EventTime" indicates to calculate the fitted values of the biomarker at all time points where the event time is observed.

Figure 4: Diagnostic plots for Model I: Martingale residuals



- We can also check the martingale residuals by conditioning on baseline covariates, e.g. treatment group.

```
library(lattice)
xyplot(martRes ~ mi.t | drug, data=aids, type=c("p","smooth"), col="black",
lwd=3, ylab="Martingale residuals", xlab="Subject-specific fitted biomarker values")
```

- We can then assess the overall fit of the model by using the Cox-Snell residuals. To calculate the survival function of the Cox-Snell residuals, we use the function survfit() from the package survival.

```
resCST<-residuals(jointFit1, process="Event", type="CoxSnell")
sfit<-survfit(Surv(resCST, death)~1, data=aids.id)
plot(sfit, mark.time=F, conf.int=T, xlab="Cox-Snell residuals",
ylab="Survival probability", main="Survival fct. of Cox-Snell residuals")
curve(exp(-x), from=0, to=max(aids.id$Time), add=T, col="grey62", lwd=2)
```

Figure 5: Diagnostic plots for Model I: Martingale residuals by treatment

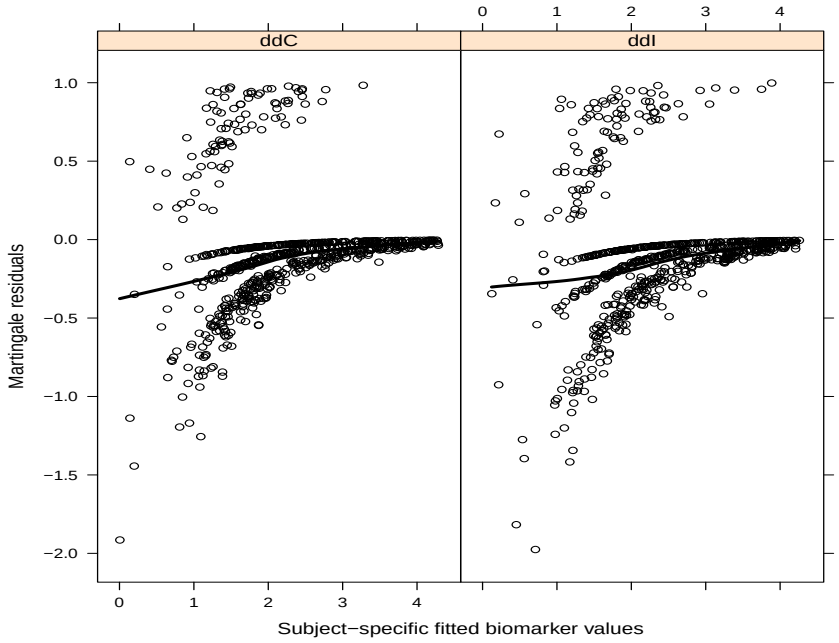
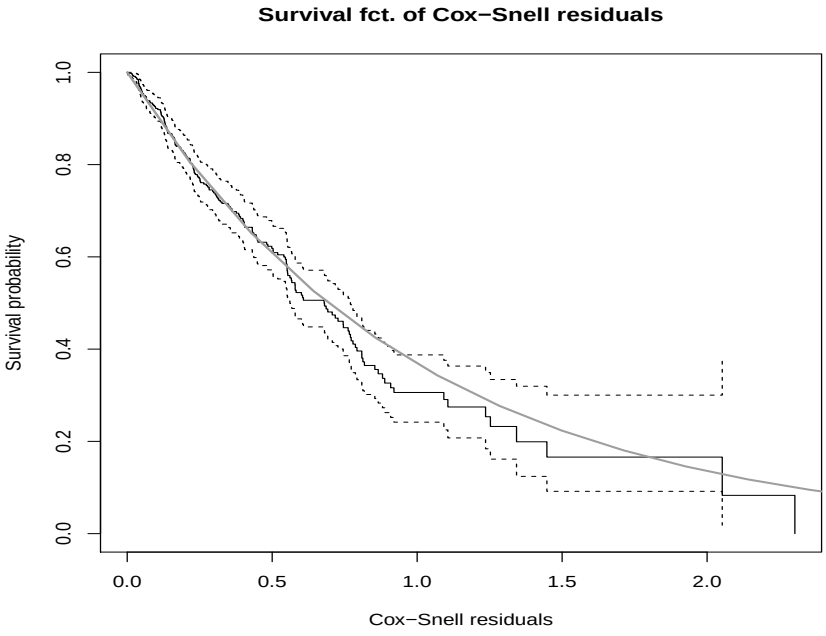


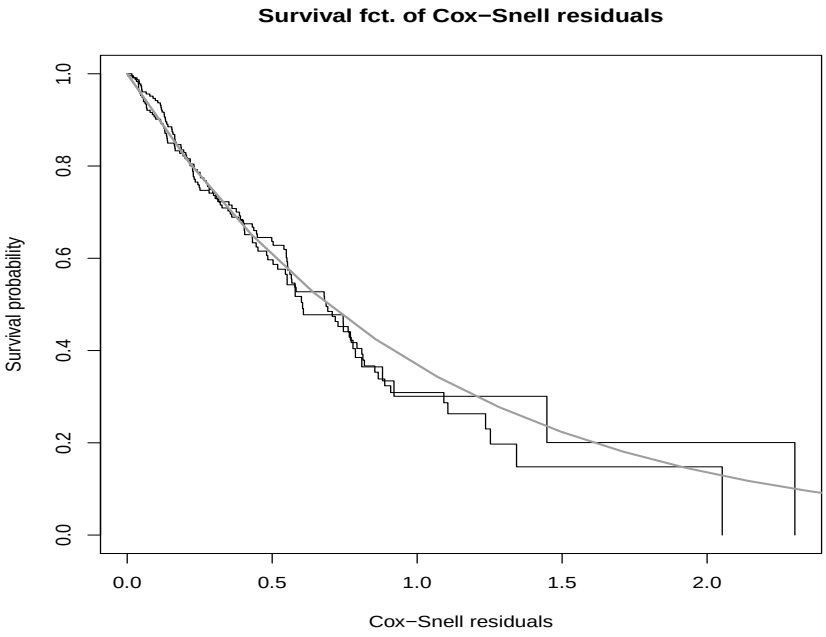
Figure 6: Diagnostic plots for Model I: Cox-Snell residuals



- We can also stratify the residuals by treatment group

```
sfit.drug<-survfit(Surv(resCST, death)~drug, data=aids.id)
pdf("Cox_snell_residuals_drug.pdf")
plot(sfit.drug, mark.time=F, conf.int=F, xlab="Cox-Snell residuals",
     ylab="Survival probability", main="Survival fct. of Cox-Snell residuals")
curve(exp(-x), from=0, to=max(aids.id$Time), add=T, col="grey62", lwd=2)
dev.off()
```

Figure 7: Diagnostic plots for Model I: Cox-Snell residuals by treatment effect



1.7 Effect of lag time

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

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

where

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

```
## Standard model
jointFit.nolag <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                        method = "piecewise-PH-aGH")

## Lag =1 for the time-dependent biomarker ##
jointFit.lag1 <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                        method = "piecewise-PH-aGH", lag=1)

## Lag =2 for the time-dependent biomarker ##
jointFit.lag2 <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                        method = "piecewise-PH-aGH", lag=2)

## Lag =3 for the time-dependent biomarker ##
jointFit.lag3 <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                        method = "piecewise-PH-aGH", lag=3)

## anova comparisons
anova(jointFit.nolag, jointFit.lag1, test=F)
anova(jointFit.nolag, jointFit.lag2, test=F)
anova(jointFit.nolag, jointFit.lag3, test=F)
```

```
> anova(jointFit.nolag, jointFit.lag1, test=F)
```

best model is nolag model

```
                   AIC     BIC  log.Lik df
jointFit.nolag 4235.97 4302.31 -2101.98
jointFit.lag1  4236.78 4303.12 -2102.39  0
```

```
> anova(jointFit.nolag, jointFit.lag2, test=F)
```

best model is nolag model

```
                   AIC     BIC  log.Lik df
jointFit.nolag 4235.97 4302.31 -2101.98
jointFit.lag2  4237.82 4304.16 -2102.91  0
```

```
> anova(jointFit.nolag, jointFit.lag3, test=F)

                   AIC     BIC  log.Lik df
jointFit.nolag 4235.97 4302.31 -2101.98
jointFit.lag3  4238.80 4305.14 -2103.40  0
```

1.8 Interaction effects between biomarkers and specific covariates

The standard parametrization assumes that the effect of the true level of the biomarker is the same in all subgroups of the target population.

This assumption might not be realistic.

A straightforward extension to handle the situation where the biomarker effect behaves differently for different subgroups of subjects.

$$h_i(t) = h_0(t) \exp[\gamma^T w_{i1} + \alpha^T \{w_{i2} \times m_i(t)\}],$$

where w_{i1} is used to accommodate the direct effects of the baseline covariate on the risk of an event, and w_{i2} contains interaction terms that expand the association of $m_i(t)$ in different subgroups of the data. The case $w_{i2} = 1$ reduces to the standard parametrization.

Interactions can be specified in the jointModel function by using the interFact statement. As an example, we use the variable 'previous opportunistic infections' prevOI from the aids data as a possible interaction effect.

```
### Model with interaction ####################
lmeFit <- lme(sqrt(CD4) ~ obstime + obstime:drug, random = ~ obstime | patient,
              data = aids)
coxFit <- coxph(Surv(Time, death) ~ drug+prevOI, data = aids.id, x = TRUE)
jointFit.interac <- jointModel(lmeFit, coxFit, timeVar = "obstime",
                               method = "piecewise-PH-aGH",
```

```
                              interFact = list(value=~ prevOI, data=aids.id))
summary(jointFit.interac)
```

I want these variabbles to be included

```
anova(jointFit.nolag, jointFit.interac)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit, timeVar = "obstime",
    method = "piecewise-PH-aGH", interFact = list(value = ~prevOI,
        data = aids.id))

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
 -2093.584 4223.167 4297.801

Variance Components:
             StdDev    Corr
(Intercept)  0.8697  (Intr)
obstime      0.0368  0.0791
Residual     0.3673

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5142  0.0425 59.1383 <0.0001
obstime         -0.0423  0.0046 -9.2892 <0.0001
obstime:drugddI  0.0062  0.0064  0.9811  0.3265

Event Process
                     Value Std.Err z-value p-value
drugddI             0.3369  0.1535  2.1948  0.0282
prevOIAIDS          1.6663  0.5832  2.8570  0.0043
Assoct             -0.6330  0.2173 -2.9135  0.0036
Assoct:prevOIAIDS -0.4579  0.2623 -1.7458  0.0808
log(xi.1)         -3.2313  0.5529 -5.8446
log(xi.2)         -2.8734  0.5444 -5.2784
log(xi.3)         -2.5280  0.5638 -4.4837
```

increase in CD4 decrease chance of dying

```
log(xi.4)          -3.0637  0.6114 -5.0109
log(xi.5)          -2.9294  0.5928 -4.9416
log(xi.6)          -2.8598  0.6373 -4.4876
log(xi.7)          -2.8376  0.7131 -3.9791

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

```
> anova(jointFit.nolag, jointFit.interac)
```

in favor of interaction model

```
                     AIC     BIC  log.Lik  LRT df p.value
jointFit.nolag   4235.97 4302.31 -2101.98
jointFit.interac 4223.17 4297.80 -2093.58 16.8  2   2e-04
```

1.9 Stratified JM

survival model is stratified

baseline hazard function - particular strata

- In many applications, it is not realistic to assume that the samples come from a homogeneous population.

- For instance, in multicentre clinical trials, the different centres are expected to have different baseline survival functions.

- A standard extension of survival models specifies multiple strata, where each stratum has its own baseline hazard function but common values for the regression coefficients γ and α.

- The hazard function is then expressed as

$$h_{ik}(t) = h_{0k}(t) \exp\{\gamma^T w_i + \alpha^T m_i(t)\},$$

where $h_{0k}(t)$ denotes the baseline hazard function for stratum k.

- Joint models with stratification are available within the JM package under the B-spline approximated baseline risk function.

- In the following, we use the variable 'gender' as a possible stratification risk factor in the aids data.

```
### Model with stratification  ####################
lmeFit <- lme(sqrt(CD4) ~ obstime + obstime:drug, random = ~ obstime | patient,
data = aids)
coxFit.strata <- coxph(Surv(Time, death) ~ drug+strata(gender), data = aids.id, x = TRUE)

jointFit.strata <- jointModel(lmeFit, coxFit.strata, timeVar = "obstime",
                              method = "spline-PH-aGH")

summary(jointFit.strata)

anova(jointFit.nolag, jointFit.strata, test=F)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit.strata, timeVar = "obstime",
    method = "spline-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: Stratified relative risk model with spline-approximated
baseline risk function
Parameterization: Time-dependent

   log.Lik      AIC      BIC
 -2090.868 4235.735 4347.686

Variance Components:
             StdDev    Corr
(Intercept)  0.8707  (Intr)
obstime      0.0370  0.0734
Residual     0.3671

Coefficients:
Longitudinal Process
```

```
                  Value Std.Err z-value p-value
(Intercept)      2.5138  0.0426 59.0668 <0.0001
obstime         -0.0418  0.0045 -9.2102 <0.0001
obstime:drugddI  0.0053  0.0064  0.8303  0.4063

Event Process
               Value Std.Err z-value p-value
drugddI       0.3541  0.1536  2.3052  0.0212
Assoct       -1.0669  0.1159 -9.2012 <0.0001
bs1(female) -1.7118  1.0718 -1.5971  0.1102
bs2(female)  0.9373  1.4934  0.6276  0.5303
bs3(female) -4.4104  2.2629 -1.9490  0.0513
bs4(female) -0.9001  1.3728 -0.6556  0.5121
bs5(female) -1.1110  1.4456 -0.7685  0.4422
bs6(female) -2.1772  2.8416 -0.7662  0.4436
bs7(female)  0.0058  5.2523  0.0011  0.9991
bs8(female)  0.4899 11.8006  0.0415  0.9669
bs9(female) -1.6987 71.6775 -0.0237  0.9811
bs1(male)   -3.4587  0.7010 -4.9338 <0.0001
bs2(male)   -0.3442  0.6587 -0.5226  0.6013
bs3(male)   -3.0389  0.6775 -4.4852 <0.0001
bs4(male)   -0.3319  0.4184 -0.7933  0.4276
bs5(male)   -2.0195  0.4915 -4.1087 <0.0001
bs6(male)   -1.0795  0.7646 -1.4118  0.1580
bs7(male)   -1.9343  1.5088 -1.2820  0.1998
bs8(male)   -0.7939  2.9840 -0.2661  0.7902
bs9(male)   -5.7604  7.7282 -0.7454  0.4560

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

spline is the baseline hazard function

```
> anova(jointFit.nolag, jointFit.strata, test=F)
                    AIC     BIC  log.Lik df
jointFit.nolag  4235.97 4302.31 -2101.98
jointFit.strata 4235.74 4347.69 -2090.87 11
```

not a big improvement when use stratification

- The function wald.strata() can be used to compare the stratified to unstratified versions of the JM.

```
wald.strata(jointFit.strata)
```

```
> wald.strata(jointFit.strata)

Wald Test for Stratification Factors

X^2 = 11.3151, df = 9, p-value = 0.2547
alternative hypothesis: spline coefficients for the baseline risk
function are not equal among strata
```

p value is not significant - not keep stratification

- We conclude that the stratification criterion is not significant.

- We can also extend the previous model by allowing the covariates to be specific to the strata. The leads to the following hazard model

$$h_{ik}(t) = h_{0k}(t) \exp\{\gamma_k^T w_{ik} + \alpha_k^T m_{ik}(t)\},$$

strata is the gender

```
coxFit.strata.inter <- coxph(Surv(Time, death) ~ drug*gender + strata(gender),
                             data = aids.id, x = TRUE)
jointFit.strata.inter <- update(jointFit.strata, survObject=coxFit.strata.inter,
                          interFact = list(value=~ gender, data=aids.id))
summary(jointFit.strata.inter)

anova(jointFit.strata, jointFit.strata.inter)

Call:
jointModel(lmeObject = lmeFit, survObject = coxFit.strata.inter,
    timeVar = "obstime", method = "spline-PH-aGH", interFact = list(value = ~gender,
        data = aids.id))

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: Stratified relative risk model with spline-approximated
baseline risk function
Parameterization: Time-dependent
```

```
   log.Lik      AIC      BIC
 -2090.447 4238.894 4359.138

Variance Components:
             StdDev    Corr
(Intercept)  0.8709  (Intr)
obstime      0.0369  0.0742
Residual     0.3673

Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5137  0.0426 59.0506 <0.0001
obstime         -0.0419  0.0045 -9.2267 <0.0001
obstime:drugddI  0.0053  0.0064  0.8250  0.4094

Event Process
                     Value Std.Err z-value p-value
drugddI             0.3979  0.4798  0.8293  0.4069
drugddI:gendermale -0.0414  0.5056 -0.0818  0.9348
Assoct             -0.8716  0.2884 -3.0227  0.0025
Assoct:gendermale  -0.2272  0.3143 -0.7230  0.4697
bs1(female)        -2.0735  1.2060 -1.7193  0.0856
bs2(female)         0.6174  1.5776  0.3914  0.6955
bs3(female)        -4.8466  2.3420 -2.0694  0.0385
bs4(female)        -1.2068  1.4643 -0.8242  0.4098
bs5(female)        -1.5900  1.5618 -1.0180  0.3087
bs6(female)        -2.1748  2.7649 -0.7866  0.4315
bs7(female)        -0.8457  5.2583 -0.1608  0.8722
bs8(female)         0.9138 11.6729  0.0783  0.9376
bs9(female)        -1.7227 69.3872 -0.0248  0.9802
bs1(male)          -3.4048  0.7060 -4.8230 <0.0001
bs2(male)          -0.2832  0.6637 -0.4267  0.6696
bs3(male)          -2.9914  0.6822 -4.3848 <0.0001
bs4(male)          -0.2740  0.4253 -0.6442  0.5195
bs5(male)          -1.9806  0.4982 -3.9757  0.0001
bs6(male)          -0.9827  0.7752 -1.2678  0.2049
bs7(male)          -2.0415  1.5601 -1.3086  0.1907
bs8(male)          -0.2472  3.1940 -0.0774  0.9383
bs9(male)          -7.5766  9.6129 -0.7882  0.4306

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

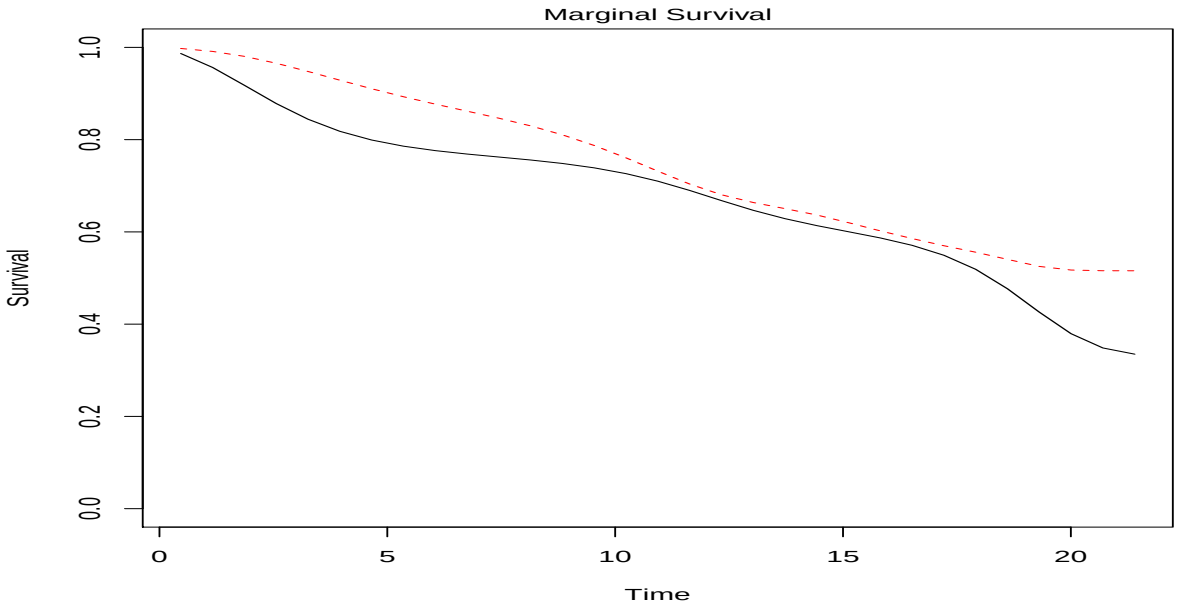
treatment effect:
baseline is female

```
Optimization:
Convergence: 0

> anova(jointFit.strata, jointFit.strata.inter)

                           AIC     BIC  log.Lik  LRT df p.value
jointFit.strata       4235.74 4347.69 -2090.87
jointFit.strata.inter 4238.89 4359.14 -2090.45 0.84  2  0.6566
```

AIC is increasing for new model and p value testing the difference between these models is not significant

- Plot of survival function

Figure 8: Marginal survival functions (dashed line is for males and solid line for females)



marginal survival for these two groups are relatively similar
indication of the stratification does not give you a big difference in marginal survival

```
pdf("Survival_plot")
plot(jointFit.strata.inter, which=3)
dev.off()
```

2 Case study: Primary Biliary Cirrhosis

2.1 Data description

- **PBC: Primary Biliary Cirrhosis**
  - A chronic, fatal but rare liver disease
  - Characterized by inflammatory destruction of the small bile ducts within the liver

- **Data collected** by Mayo Clinic from 1974 to 1984 (Murtaugh et al., Hepatology, 1994)

- **Outcomes of interest**
  - Time to death and/or time to liver transplantation
  - Randomized treatment: 158 patients received D-penicillamine and 154 placebo
  - Longitudinal serum bilirubin levels

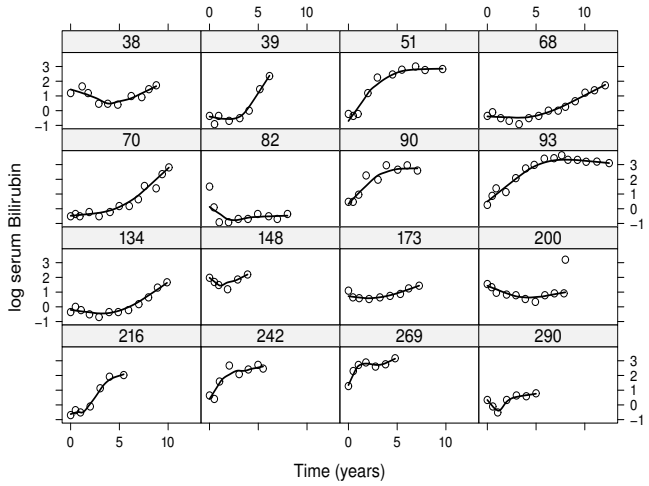


Figure 9: Longitudinal profiles of serum bilirubin levels

```
data("pbc2.id")
head(pbc2.id,5)

  id     years       status      drug      age    sex year ascites hepatomegaly spiders
1  1  1.095170         dead D-penicil 58.76684 female    0     Yes          Yes     Yes
2  2 14.152338        alive D-penicil 56.44782 female    0      No          Yes     Yes
3  3  2.770781         dead D-penicil 70.07447   male    0      No           No      No
4  4  5.270507         dead D-penicil 54.74209 female    0      No          Yes     Yes
5  5  4.120578 transplanted   placebo 38.10645 female    0      No          Yes     Yes
                    edema serBilir serChol albumin alkaline  SGOT platelets prothrombin histologic status2
1 edema despite diuretics     14.5     261    2.60     1718 138.0       190        12.2          4       1
2               No edema      1.1     302    4.14     7395 113.5       221        10.6          3       0
3      edema no diuretics      1.4     176    3.48      516  96.1       151        12.0          4       1
4      edema no diuretics      1.8     244    2.54     6122  60.6       183        10.3          4       1
5               No edema      3.4     279    3.53      671 113.2       136        10.9          3       0
```

2.2 Comparison JM vs. Cox model

- Cox model

```
coxFit <- coxph(Surv(years, status2) ~ log(serBilir) + drug + sex + age,
                data = pbc2.id)
summary(coxFit)

Call:
coxph(formula = Surv(years, status2) ~ log(serBilir) + drug +
    sex + age, data = pbc2.id)

  n= 312, number of events= 140

                   coef exp(coef)  se(coef)      z Pr(>|z|)
log(serBilir)  1.085856  2.961975  0.092587 11.728  < 2e-16 ***
drugD-penicil -0.127653  0.880159  0.176473 -0.723    0.469
sexfemale     -0.004180  0.995829  0.233030 -0.018    0.986
age            0.046376  1.047468  0.008222  5.640  1.7e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

              exp(coef) exp(-coef) lower .95 upper .95
log(serBilir)    2.9620     0.3376    2.4704     3.551
drugD-penicil    0.8802     1.1362    0.6228     1.244
sexfemale        0.9958     1.0042    0.6307     1.572
age              1.0475     0.9547    1.0307     1.064

Concordance= 0.811  (se = 0.018 )
Likelihood ratio test= 167.8  on 4 df,   p=<2e-16
Wald test            = 174.7  on 4 df,   p=<2e-16
Score (logrank) test = 199.9  on 4 df,   p=<2e-16
```

- JM

```
lmeFit <- lme(log(serBilir) ~ year+drug+year:drug, random = ~ year | id, data=pbc2)
coxFit <- coxph(Surv(years, status2) ~ drug + sex + age, data = pbc2.id, x = TRUE)
jointFit <- jointModel(lmeFit, coxFit, timeVar = "year", method = "piecewise-PH-aGH")
summary(jointFit)

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

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

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
 -1888.845 3815.689 3886.806

Variance Components:
             StdDev    Corr
(Intercept)  0.9986  (Intr)
year         0.1797  0.4313
Residual     0.3473

Coefficients:
Longitudinal Process
                    Value Std.Err z-value p-value
(Intercept)         0.5610  0.0826  6.7921 <0.0001
year                0.1864  0.0188  9.9328 <0.0001
drugD-penicil      -0.1353  0.1162 -1.1642  0.2443
year:drugD-penicil -0.0033  0.0255 -0.1295  0.8970

Event Process
                Value Std.Err  z-value p-value
drugD-penicil -0.0179  0.1856  -0.0966  0.9230
sexfemale      0.0992  0.2485   0.3994  0.6896
age            0.0627  0.0092   6.7812 <0.0001
```

```
Assoct          1.3438  0.1014  13.2542 <0.0001
log(xi.1)      -7.9743  0.7049 -11.3133
log(xi.2)      -7.7223  0.6963 -11.0898
log(xi.3)      -7.9591  0.7102 -11.2067
log(xi.4)      -7.9096  0.7325 -10.7981
log(xi.5)      -7.4819  0.6998 -10.6908
log(xi.6)      -7.1774  0.7231  -9.9254
log(xi.7)      -7.7526  0.7811  -9.9253

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

clinical trial may not be randomized
treatment effect at baseline

- LMM component

Table 5: Parameter estimates for log bilirubin

	JM			LMM		
	β	SE	p-value	β	SE	p-value
Drug	0.18	0.02	< 0.0001	0.19	0.02	< 0.0001
Years	-0.13	0.12	0.25	-0.14	0.12	0.24
Years $\times$ drug	-0.004	0.03	0.86	-0.003	0.03	0.90

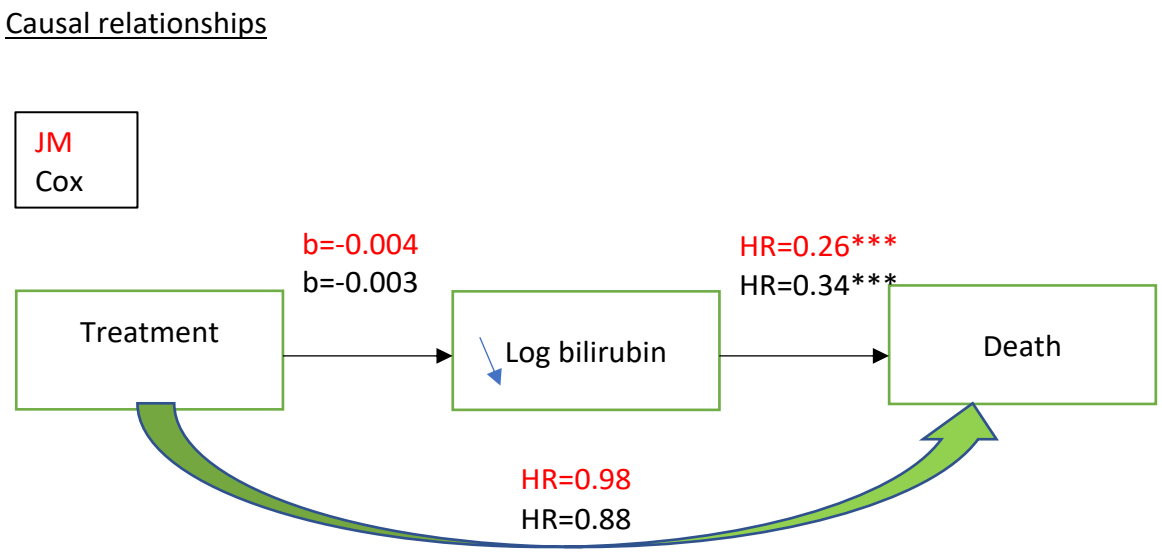
- Survival component

Table 6: Parameter estimates for time to death

	JM			Cox PH		
	logHR	SE	p-value	logHR	SE	p-value
Drug	-0.018	0.19	0.92	-0.13	0.18	0.47
Sex	0.099	0.25	0.69	-0.004	0.23	0.99
Age	0.063	0.009	< 0.0001	0.046	0.008	< 0.0001
log(bilirubin)	1.34	0.10	< 0.0001	1.09	0.09	< 0.0001

- Induced causal relationships

Figure 10: Causal relationships in PBC data



decreasing log of bilirubin, decrease risk of dying (HR is lower than 1 = 0.26) from the disease, so improving the overall survival - direct effect is very significant
hazard is close to 1 and it's not significant
indirect and treatments and death. just a direction of indirect effect here.

2.3 Competing risks model

- In longitudinal studies, we often observed more than one time-to-event outcomes.

- Competing risks concern the situation where more than one cause of failure is possible (Putter, Fiocco, and Geskus, 2007).

- A classical example relates to several causes of death (e.g. from cancer) where the occurrence of any cause of death prevents the event of interest from occurring.

- Treating the events of the competing causes as censored observations will lead to biased estimates of the survival function for the event of interest when we are in the presence of competing risks (Putter et al., 2007).

- In the PBC data, several outcomes can "compete" to be the first event observed, e.g. time to death or time to transplantation. It could be interesting to distinguish these 2 events and estimate specific effects of baseline covariates for each of them.

- If these covariates are endogenous time-dependent covariates, the JM approach should be used.

- Notations:

  - We assume K different causes of failure
  - We let $T_{i1}^*, \cdots, T_{iK}^*$ be the true failure times
  - The observed event time is $T_i = \min(T_{i1}^*, \cdots, T_{iK}^*, C_i)$
  - The event indicator is $\delta_i \in \{0, 1, \cdots, K\}$, where 0 corresponds to censoring and $1, \cdots, K$ to the competing events.

- For each of the K causes of failure, the hazard function is

$$h_{ik}(t) = h_{0k}(t) \exp\{\gamma_k^T w_i + \alpha_k m_i(t)\}.$$

- It includes baseline covariates and current values of the biomarker that affect the **cause-specific hazard** function.

- The contribution of the event time to the likelihood, conditional on the biomarker, can be written as

probability of having one event of interest given you survival all possible events before time t
if someone has no event of interest, he only has the survival component

$$p(T_i, \delta_i | b_i; \theta_t, \beta) = \prod_{k=1}^{K} \Big[h_{0k}(T_i) \exp\{\gamma_k^T w_i + \alpha_k m_i(T_i)\} \Big]^{I(\delta_i = k)}$$

$$\times \exp\Big(- \sum_{k=1}^{K} \int_0^{T_i} h_{0k}(s) \exp\{\gamma_k^T w_i + \alpha_k m_i(s)\} ds \Big).$$

two events of interest - so need to create the data framwork

- For the data format, each subject has K rows, one for each possible cause of failure. The function crLong() can be used to create the 'competing risks long format' data set.

- Usual data format (for survival analysis)

```
head(pbc2.id[c("id","years","status")],5)
```

```
> head(pbc2.id[c("id","years","status")],5)
  id     years       status
1  1  1.095170         dead
2  2 14.152338        alive
3  3  2.770781         dead
4  4  5.270507         dead
5  5  4.120578 transplanted
```

an event for both - transplantation status and death status
I need to create a long format dataset - use crLONG - two rows for one individual - one for death status and one for transplantations status

- New format

```
pbc2.idCR<-crLong(pbc2.id, statusVar="status", censLevel="alive", nameStrata="CR")
head(pbc2.idCR[c("id","years","status","CR","status2")],10)
```

```
> head(pbc2.idCR[c("id","years","status","CR","status2")],10)
    id     years       status           CR status2
1    1  1.095170         dead         dead       1
1.1  1  1.095170         dead transplanted       0
2    2 14.152338        alive         dead       0
2.1  2 14.152338        alive transplanted       0
3    3  2.770781         dead         dead       1
3.1  3  2.770781         dead transplanted       0
4    4  5.270507         dead         dead       1
4.1  4  5.270507         dead transplanted       0
5    5  4.120578 transplanted         dead       0
5.1  5  4.120578 transplanted transplanted       1
```

- Note that each patient has 2 rows (i.e. 2 causes of failure, death or transplantation) but the time variable 'years' is similar.

- The variable 'CR' denotes the cause for the specific line of the dataset.

- LMM for the PBC data. We assume now a more complicated model

polynomial function with quadratic terms

$$y_i(t) = \beta_0+\beta_1\text{drug}+\beta_2 t+\beta_3 t^2+\beta_4\{\text{drug}\times t\}+\beta_5\{\text{drug}\times t^2\}+b_{i0}+b_{i1}t+b_{i2}t^2+\epsilon_i(t).$$

```
lmeFit.pbc <- lme(log(serBilir) ~ drug*(year + I(year^2)),
                  random = ~ year + I(year^2)| id, data=pbc2)
summary(lmeFit.pbc)
```

```
> summary(lmeFit.pbc)
Linear mixed-effects model fit by REML
 Data: pbc2
       AIC      BIC    logLik
  2929.358 3001.767 -1451.679
```

```
Random effects:
 Formula: ~year + I(year^2) | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.99941237 (Intr) year
year        0.30424697  0.171
I(year^2)   0.02525548  0.011 -0.882
Residual    0.30486888

Fixed effects: log(serBilir) ~ drug * (year + I(year^2))
                             Value  Std.Error   DF   t-value p-value
(Intercept)              0.5855771 0.08268683 1629  7.081867  0.0000
drugD-penicil           -0.1435083 0.11623757  310 -1.234612  0.2179
year                     0.1624491 0.03112533 1629  5.219194  0.0000
I(year^2)                0.0015047 0.00314793 1629  0.477995  0.6327
drugD-penicil:year       0.0072035 0.04387470 1629  0.164182  0.8696
drugD-penicil:I(year^2) -0.0023505 0.00448354 1629 -0.524245  0.6002
 Correlation:
                        (Intr) drgD-p year   I(y^2) drgD-:
drugD-penicil           -0.711
year                     0.047 -0.034
I(year^2)                0.088 -0.062 -0.842
drugD-penicil:year      -0.033  0.050 -0.709  0.597
drugD-penicil:I(year^2) -0.061  0.085  0.591 -0.702 -0.842

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-4.847446422 -0.453308510  0.000722912  0.446415136  4.557128170

Number of Observations: 1945
Number of Groups: 312
```

- For the cause-specific hazard models, we assume

$$h_{i1}(t) = h_{01}(t)\exp\{\gamma_{11}\text{drug}_i + \gamma_{12}\text{age}_i + \alpha_1 m_i(t)\},$$

baseline hazard function to the first event

$$h_{i2}(t) = h_{02}(t)\exp\{(\gamma_{11}+\gamma_{21})\text{drug}_i + (\gamma_{12}+\gamma_{22})\text{age}_i + (\alpha_1+\alpha_2) m_i(t)\},$$

baseline hazard function and parameters for the first event
biomarkers will be specific to the second event

where event 1 denotes the risk of transplantation and event 2 the risk of death.

all the variables will be dependent on the event indicator

hazard baseline function will also dependent on the event indicator

- Model fitting is obtained by

```
coxFit.pbc <- coxph(Surv(years, status2)~(drug + age) * CR + strata(CR),
                    data=pbc2.idCR, x=T)
```

at the end, we are getting two baseline hazard functions and two survival functions

main effect variables in the model - that's not estimated

```
Call
coxph(formula = Surv(years, status2) ~ (drug + age) * CR + strata(CR),
    data = pbc2.idCR, x = T)

  n= 624, number of events= 169

                          coef exp(coef) se(coef)      z Pr(>|z|)
drugD-penicil        -0.23680   0.78915  0.37723 -0.628    0.530
age                  -0.09649   0.90802  0.02265 -4.259 2.05e-05 ***
CRdead                     NA        NA  0.00000     NA       NA
drugD-penicil:CRdead  0.07473   1.07759  0.41480  0.180    0.857
age:CRdead            0.14221   1.15281  0.02419  5.878 4.15e-09 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

                     exp(coef) exp(-coef) lower .95 upper .95
drugD-penicil           0.7891     1.2672    0.3768    1.6530
age                     0.9080     1.1013    0.8686    0.9492
CRdead                      NA         NA        NA        NA
drugD-penicil:CRdead    1.0776     0.9280    0.4779    2.4296
age:CRdead              1.1528     0.8674    1.0994    1.2088

Concordance= 0.646  (se = 0.021 )
Likelihood ratio test= 51.47  on 4 df,   p=2e-10
Wald test            = 47.58  on 4 df,   p=1e-09
Score (logrank) test = 49.58  on 4 df,   p=4e-10
```

- For fitting the JM, only the option 'spline-PH-aGH' or 'spline-PH-GH' are available.

indicate that we have compete risk

```
jointFit.pbc <- jointModel(lmeFit.pbc, coxFit.pbc, timeVar="year",
                           method="spline-PH-aGH", CompRisk=T,
                           interFact=list(value=~CR, data=pbc2.idCR))
```

```
summary(jointFit.pbc)

Call:
jointModel(lmeObject = lmeFit.pbc, survObject = coxFit.pbc, timeVar = "year",
    method = "spline-PH-aGH", interFact = list(value = ~CR, data = pbc2.idCR),
    CompRisk = T)

Data Descriptives:
Longitudinal Process Event Process
Number of Observations: 1945 Number of Events: 169 (54.2%)
Number of Groups: 312

Joint Model Summary:
Longitudinal Process: Linear mixed-effects model
Event Process: Competing risks relative risk model with spline-approximated
baseline risk function
Parameterization: Time-dependent

   log.Lik      AIC      BIC
 -1896.393 3866.785 4005.276
```

treatment for the first event - which is transplantation

```
Variance Components:
             StdDev    Corr
(Intercept)  0.9966  (Intr)    year
year         0.3090  0.1981
I(year^2)    0.0251  0.0233 -0.8590
Residual     0.3027

Coefficients:
Longitudinal Process
                          Value Std.Err z-value p-value
(Intercept)              0.5841  0.0753  7.7540 <0.0001
drugD-penicil           -0.1448  0.1069 -1.3545  0.1756
year                     0.1650  0.0313  5.2646 <0.0001
I(year^2)                0.0031  0.0033  0.9383  0.3481
drugD-penicil:year       0.0064  0.0430  0.1494  0.8813
drugD-penicil:I(year^2) -0.0025  0.0043 -0.5736  0.5662

Event Process
                          Value Std.Err  z-value p-value
drugD-penicil           -0.2986  0.3844  -0.7766  0.4374
age                     -0.0855  0.0245  -3.4952  0.0005
drugD-penicil:CRdead     0.2721  0.4213   0.6461  0.5182
age:CRdead               0.1516  0.0259   5.8603 <0.0001
Assoct                   1.0189  0.1917   5.3146 <0.0001
Assoct:CRdead            0.4505  0.2182   2.0647  0.0389
```

transplantation

transplantation

people have a unit increaes in bilirubin have

i

```
bs1(transplanted)     -4.8115  4.2422  -1.1342  0.2567
bs2(transplanted)     -3.9341  2.4143  -1.6295  0.1032
bs3(transplanted)      0.0880  1.7452   0.0504  0.9598
bs4(transplanted)     -2.3802  1.4399  -1.6530  0.0983
bs5(transplanted)      0.0968  1.4255   0.0679  0.9458
bs6(transplanted)     -2.2655  1.8350  -1.2346  0.2170
bs7(transplanted)     -4.2260  5.4507  -0.7753  0.4381
bs8(transplanted)     -4.2056 10.5547  -0.3985  0.6903
bs9(transplanted)     -3.7498 13.5878  -0.2760  0.7826
bs1(dead)             -8.0430  0.7625 -10.5483 <0.0001
bs2(dead)             -8.7315  0.8341 -10.4687 <0.0001
bs3(dead)             -7.9882  0.8535  -9.3599 <0.0001
bs4(dead)             -8.3107  0.7464 -11.1341 <0.0001
bs5(dead)             -8.2169  0.7947 -10.3393 <0.0001
bs6(dead)             -8.1018  0.7966 -10.1710 <0.0001
bs7(dead)             -6.2104  1.2790  -4.8557 <0.0001
bs8(dead)            -10.0470  2.1937  -4.5800 <0.0001
bs9(dead)             -6.9176  2.3545  -2.9380  0.0033

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

- The interFact option is used the include in the survival model the interaction between the biomarker $m_i(t)$ and the variable 'CR'.

- The results indicate that a one unit increase of the current value of the log serum bilirubin is associated with a $\exp(1.02) \approx 2.8$ fold increase (95% CI: 1.9-4.1) in a patient's risk of transplantation and a $\exp(1.02+0.45) \approx 4.3$ fold increase (95% CI: 1.9-9.9) in the patient's risk of death.

- Other parametrizations for the effect of the biomarker can be introduced as well, e.g. effect of current value and current slope.