Dynamic prognostic tools using joint models for recurrent and terminal events: *Evolution after a breast cancer*

A. Mauguen [*], B. Rachet [**], S. Mathoulin-Pélissier [*], G. MacGrogan [***], A. Laurent [*], V. Rondeau[*]

[*] INSERM U897, Bordeaux
[**] London School, London
[***] Institut Bergonié, Bordeaux

Workshop
10-11 Oct. 2013, Bordeaux

Introduction

- **After a breast cancer diagnosis**
  $\rightarrow$ single or multiple events
  (recurrences, metastases, death)

Introduction

- **After a breast cancer diagnosis**
  $\rightarrow$ single or multiple events
  (recurrences, metastases, death)
- **Prediction of death**
  $\rightarrow$ clinical therapeutic decisions, and patient monitoring
  $\rightarrow$ patient information
  $\rightarrow$ trials : defining patient subpopulations

Introduction

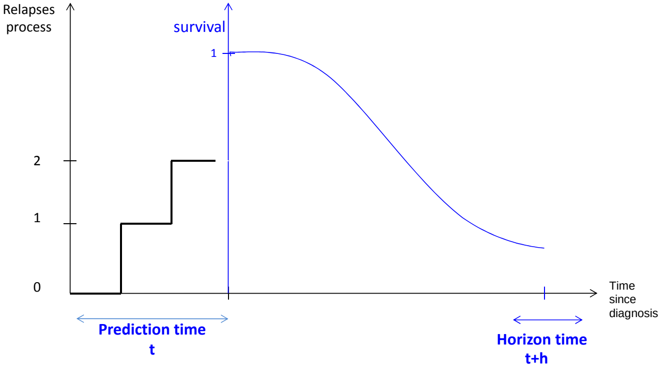
- **After a breast cancer diagnosis**
  $\rightarrow$ single or multiple events (recurrences, metastases, death)
- **Prediction of death**
  $\rightarrow$ clinical therapeutic decisions, and patient monitoring
  $\rightarrow$ patient information
  $\rightarrow$ trials : defining patient subpopulations
- **Account for**
  $\rightarrow$ individual characteristics
  $\rightarrow$ tumor characteristics
  $\rightarrow$ previous treatments
  $\rightarrow$ evolution of longitudinal markers (*Rizopoulos, 2011 ; Proust-Lima 2009*)

Introduction : Motivating example

- Cohort of patients with **operable breast cancer**
- Treated in a **comprehensive cancer center** and followed 13.9 years (median)
- **Recurrent events** observed : loco-regional relapses, distant metastases ; until 3 events per patient
- Hypothesis : individual covariates but also **recurrent event process** may improve prediction of death risk

Objective

To predict the risk of death between time t and $t + h$ given the recurrent event process before time t in the context of joint modeling



Joint Models

- Recurrent events and death **processes** are potentially **correlated**
- Standard (naive) approach of Cox with time-dependent covariate only for **external covariates** !
- Interest :
  - investigating the **strength of association** between recurrent events and death
  - allows to study impact of **covariates both** on recurrent events and death
  - treat **informative censoring** by death

Joint models : some notations

- t time of prediction and h window of prediction
- D_i time of death for subject i, $i = 1, ..., n$
- X_{ij} time of the *jth* recurrence for subject i
- Z_{ij}^R and Z_i^D covariates vectors for recurrence and death
- λ_{ij}^R and λ_i^D baseline hazards for risk of recurrence or death

Joint models

Joint modeling for the risk of recurrent event (disease relapses) and terminal event (death)

$$\begin{cases} \lambda_{ij}^{R}(t|u_i) = u_i \lambda_0^{R}(t) \exp(\beta_1' Z_{ij}^{R}) \\ \lambda_{i}^{D}(t|u_i) = u_i^{\alpha} \lambda_0^{D}(t) \exp(\beta_2' Z_{i}^{D}) \end{cases}$$

- calendar timescale (time from origin)
- $u_i \sim \Gamma(1/\theta; 1/\theta)$, i.e. $E(u_i) = 1$ and $var(u_i) = \theta$
- θ dependency between recurrent events and death
- α sense and strength of the association (more flexibility)

Liu et al. Biometrics 2004 ; Rondeau et al. Biostatistics 2007

Inference in the joint model

Penalized log-likelihood :
- smooth baseline hazard functions
- approximated by cubic M-splines

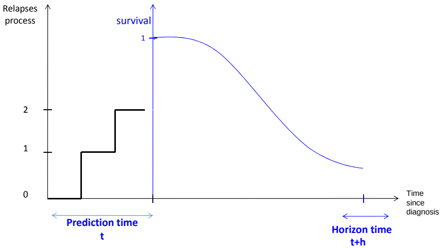
$$pl(\xi) = l(\xi) - \kappa_1 \int_0^\infty (\lambda_0^R(t))''^2 \mathrm{d}t - \kappa_2 \int_0^\infty (\lambda_0^D(t))''^2 \mathrm{d}t$$

With the vector of parameters : $\zeta = (\lambda_0^D(.), \lambda_0^R(.), \beta, \alpha, \theta)$ and κ_1 and κ_2 two smoothing parameters for the baseline hazard functions

Dynamic prediction

- Consider a new subject i **free of death at time** t (i.e. $D > t$), for whom we observe j recurrences before t and for whom the vector of covariates Z_{ij}^R and Z_{ij}^D are available at time of prediction
- The history of recurrences for patient i until time t is :

$$\mathcal{H}_i^J(t) = \{N_i^R(t) = J, X_{i1} < \ldots < X_{iJ} \leq t\}$$



Dynamic prediction

Distinguish **three setting** for the probability of death between t and $t + h$

Setting 1

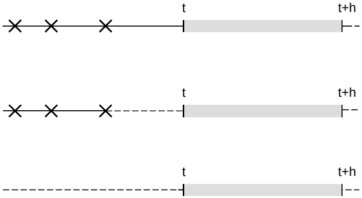
Exactly 3 recurrent events before t

Setting 2

At least 3 recurrent events before t

Setting 3

Whatever the history of recurrent events before t



× Recurrent event

— Period where we consider what happens

Window of prediction of death

- - - Period where we do not consider what happens

Dynamic prediction

Setting 1 : with exactly *j* recurrences before *t*

$$P^1(t,t+h;\xi) = P(D_i \leq t+h | D_i > t, \mathcal{H}_i^{J,1}(t), Z_{ij}^R, Z_i^D, \xi)$$
$$= \frac{\int_0^\infty [S_i^D(t|Z_i^D,u_i,\xi) - S_i^D(t+h|Z_i^D,u_i,\xi)](u_i)^J S_{i(J+1)}^R(t|Z_{ij}^R,u_i,\xi) g(u_i)\mathrm{d}u_i}{\int_0^\infty S_i^D(t|Z_i^D,u_i,\xi)(u_i)^J S_{i(J+1)}^R(t|Z_{ij}^R,u_i,\xi) g(u_i)\mathrm{d}u_i}$$

and $\mathcal{H}_i^{J,1}(t) = \{N_i^R(t) = J, X_{i1} < \ldots < X_{iJ} \leq t\}$, with $X_{i0} = 0$ and $X_{i(J+1)} > t$

Dynamic prediction

Setting 1 : with exactly j recurrences before t

$$P^1(t, t+h; \xi) = P(D_i \leq t+h | D_i > t, \mathcal{H}_i^{J,1}(t), Z_{ij}^R, Z_i^D, \xi)$$

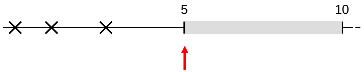
$$= \frac{\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)](u_i)^J S_{i(J+1)}^R(t|Z_{ij}^R, u_i, \xi) g(u_i) \mathrm{d}u_i}{\int_0^\infty S_i^D(t|Z_i^D, u_i, \xi)(u_i)^J S_{i(J+1)}^R(t|Z_{ij}^R, u_i, \xi) g(u_i) \mathrm{d}u_i}$$

and $\mathcal{H}_i^{J,1}(t) = \{N_i^R(t) = J, X_{i1} < \ldots < X_{iJ} \leq t\}$, with $X_{i0} = 0$ and $X_{i(J+1)} > t$

Example :

"Up to now Mr Martin has developed 3 recurrences of his initial cancer, his probability of dying in the next 5 years is x%"

Exactly 3 recurrent events before t



Dynamic prediction

Setting 2 : with at least j recurrences before t

$$P^2(t, t+h; \xi)$$
$$= \frac{\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)](u_i)^J S_{iJ}^R(X_{iJ}/Z_{ij}^R, \xi, u_i) g(u_i) \mathrm{d}u_i}{\int_0^\infty S_i^D(t|Z_i^D, \xi, u_i)(u_i)^J S_{iJ}^R(X_{iJ}/Z_{ij}^R, \xi, u_i) g(u_i) \mathrm{d}u_i}$$

and $\mathcal{H}_i^{J,2}(t) = \{N_i^R(t) \geq J, X_{i1} < \ldots < X_{iJ} \leq t\}$, with $X_{i0} = 0$

Dynamic prediction

Setting 2 : with at least j recurrences before t

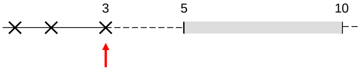
$$P^2(t, t+h; \xi) = \frac{\displaystyle\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)](u_i)^J S_{iJ}^R(X_{iJ}/Z_{ij}^R, \xi, u_i) g(u_i)\mathrm{d}u_i}{\displaystyle\int_0^\infty S_i^D(t|Z_i^D, \xi, u_i)(u_i)^J S_{iJ}^R(X_{iJ}/Z_{ij}^R, \xi, u_i) g(u_i)\mathrm{d}u_i}$$

and $\mathcal{H}_i^{J,2}(t) = \{N_i^R(t) \geq J, X_{i1} < \ldots < X_{iJ} \leq t\}$, with $X_{i0} = 0$

Example :

"Mr Martin has already developed 3 recurrences, if he is still alive in 2 years, his probability of dying between 5 and 10 years will be x% "

At least 3 recurrent events before t



Dynamic prediction

Setting 3 : considering the recurrence history only in the parameters estimation

$$
\begin{aligned}
&P^3(t, t+h; \xi) \\
&= P(D_i \leq t+h | D_i > t, Z_i^D, \xi) \\
&= \frac{\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)] g(u_i) \mathrm{d}u_i}{\int_0^\infty S_i^D(t|Z_i^D, \xi, u_i) g(u_i) \mathrm{d}u_i}
\end{aligned}
$$

Dynamic prediction

Setting 3 : considering the recurrence history only in the parameters estimation

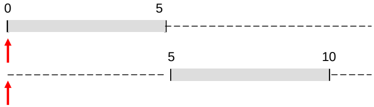
$$
\begin{aligned}
&P^3(t, t+h; \xi) \\
&= P(D_i \leq t+h | D_i > t, Z_i^D, \xi) \\
&= \frac{\displaystyle\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)] g(u_i) \mathrm{d}u_i}{\displaystyle\int_0^\infty S_i^D(t|Z_i^D, \xi, u_i) g(u_i) \mathrm{d}u_i}
\end{aligned}
$$

Example :

" his probability of dying in the next 5 years is x%"

" if his still alive in 5 years, his probability of dying over the next 5 years will be x%"

Whatever the history of recurrent events before t



Dynamic prediction : variability of the probability estimators

by **Monte Carlo** :

- at each *b* step (b=1,...,B=1000) :
  $\hat{\xi} = (\widehat{\lambda_0^R(.)}, \widehat{\lambda_0^D(.)}, \hat{\boldsymbol{\beta}}, \hat{\alpha}, \hat{\theta})$ from $\mathcal{MN}(\hat{\xi}, \hat{\Sigma}_{\xi})$.
  estimate $P^b(t, t+h; \hat{\xi})$

- Percentile confidence interval : using the 2.5^{th} and the 97.5^{th} percentiles

Dynamic prediction : Error of prediction

Based on a weighted time-dependent Brier Score (IPCW error)

$$Err_{t+h} = \frac{1}{N_t}\sum_{i=1}^{N_t}[I(T_i^D > t+h) - (1 - \hat{P}(t, t+h; \hat{\xi}))]^2 \hat{w}_i(t+h, \hat{G}_N(.))$$

with

$$w_i(t+h, \hat{G}_N(.)) = \frac{I(T_i^D \leq t+h)\delta_i^D}{\hat{G}_N(T_i^D)/\hat{G}_N(t)} + \frac{I(T_i^D > t+h)}{\hat{G}_N(t+h)/\hat{G}_N(t)}$$

T_i^D = observed survival time ; δ_i = event indicator
N_t =patients alive and uncensored at t
$\hat{G}_N(t)$ = KM estimate or adjusted Cox estimate of the censoring distribution
Validated by a 10-fold cross-validation

Brier. Monthly Weather Review 1950 - Gerds et al. Biometrical J 2006

Application

- 1067 patients
- median follow-up : 13.8 years (min=5 months)
- 330 patients died
- 362 patients with recurrent events (mean 0.40), i.e. 427 obsevations (locoregional relapses and distant metastases)

N events	0	1	2	3
Alive	600	114	20	3
Died	105	187	37	1
All	705	301	57	4

with the R package **frailtypack** :
(http ://cran.r-project.org/web/packages/frailtypack/)

Prognostic joint model

Variable	% of patients	For recurrent events HR	(95%*CI*)	For death HR	(95%*CI*)
- Age					
$]40-55]$ *vs* $]55-84]$	(36.6)	1.18	(0.92-1.51)	0.36	(0.19-0.66)
$[28-40]$ *vs* $]55-84]$	(7.7)	2.54	(1.82-3.56)	1.76	(0.82-3.81)
- P. vasc. invas.	(26.7)	1.47	(1.15-1.88)	3.35	(1.80-6.25)
- Tumor size	(22.7)	1.86	(1.47-2.37)	4.68	(2.70-8.12)
> 20 *vs* ≤ 20 *mm*					
- HER2 positive	(11.2)	1.43	(1.03-1.99)	1.31	(0.62-2.77)
- HR	(83.0)	0.81	(0.57-1.16)	0.23	(0.10-0.54)
(+ vs -)					
- Nodes involv.	(42.3)	1.82	(1.42-2.32)	4.52	(2.43-8.41)
(yes vs no)					
- Grade					
II vs I	(45.7)	2.14	(1.55-2.95)	7.99	(3.39-18.85)
III vs I	(24.6)	2.21	(1.48-3.31)	10.80	(4.13-33.76)
θ			1.04 (se=0.06)		
α			4.61 (se=0.28)		
LCV			2.04		

Prediction values - between 5 and 10 years

Recurrence history	Risk of death between 5 and 10 years		
	$P^1(5, 10; \hat{\xi})$	$P^2(5, 10; \hat{\xi})$	$P^3(5, 10; \hat{\xi})$
No recurrence	10.8 (4.2)	10.8 (4.2)	12.7 (4.5)
One recurrence			
$X_{i1} = 1$	30.3 (8.9)	33.3 (8.9)	12.7 (4.5)
$X_{i1} = 2.5$	30.3 (8.9)	32.3 (8.9)	12.7 (4.5)
$X_{i1} = 4.9$	30.3 (8.9)	30.4 (8.9)	12.7 (4.5)
Two recurrences			
$X_{i1} = 1, X_{i2} = 2$	50.6 (11.4)	53.2 (11.1)	12.7 (4.5)
$X_{i1} = 2, X_{i2} = 4$	50.6 (11.4)	51.5 (11.3)	12.7 (4.5)
$X_{i1} = 4, X_{i2} = 4.9$	50.6 (11.4)	50.7 (11.4)	12.7 (4.5)
Three recurrences			
$X_{i1} = 1, X_{i2} = 2, X_{i3} = 3$	67.4 (11.9)	68.9 (11.4)	12.7 (4.5)
$X_{i1} = 1, X_{i2} = 2.5, X_{i3} = 4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)
$X_{i1} = 3, X_{i2} = 4, X_{i3} = 4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)

Prediction values - between 5 and 10 years

Recurrence history	Risk of death between 5 and 10 years		
	$P^1(5,10;\hat{\xi})$	$P^2(5,10;\hat{\xi})$	$P^3(5,10;\hat{\xi})$
No recurrence	10.8 (4.2)	10.8 (4.2)	12.7 (4.5)
One recurrence			
$X_{i1}=1$	30.3 (8.9)	33.3 (8.9)	12.7 (4.5)
$X_{i1}=2.5$	30.3 (8.9)	32.3 (8.9)	12.7 (4.5)
$X_{i1}=4.9$	30.3 (8.9)	30.4 (8.9)	12.7 (4.5)
Two recurrences			
$X_{i1}=1, X_{i2}=2$	50.6 (11.4)	53.2 (11.1)	12.7 (4.5)
$X_{i1}=2, X_{i2}=4$	50.6 (11.4)	51.5 (11.3)	12.7 (4.5)
$X_{i1}=4, X_{i2}=4.9$	50.6 (11.4)	50.7 (11.4)	12.7 (4.5)
Three recurrences			
$X_{i1}=1, X_{i2}=2, X_{i3}=3$	67.4 (11.9)	68.9 (11.4)	12.7 (4.5)
$X_{i1}=1, X_{i2}=2.5, X_{i3}=4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)
$X_{i1}=3, X_{i2}=4, X_{i3}=4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)

Prediction values - between 5 and 10 years

Recurrence history	Risk of death between 5 and 10 years		
	$P^1(5, 10; \hat{\xi})$	$P^2(5, 10; \hat{\xi})$	$P^3(5, 10; \hat{\xi})$
No recurrence	10.8 (4.2)	10.8 (4.2)	12.7 (4.5)
One recurrence			
$X_{i1} = 1$	30.3 (8.9)	33.3 (8.9)	12.7 (4.5)
$X_{i1} = 2.5$	30.3 (8.9)	32.3 (8.9)	12.7 (4.5)
$X_{i1} = 4.9$	30.3 (8.9)	30.4 (8.9)	12.7 (4.5)
Two recurrences			
$X_{i1} = 1, X_{i2} = 2$	50.6 (11.4)	53.2 (11.1)	12.7 (4.5)
$X_{i1} = 2, X_{i2} = 4$	50.6 (11.4)	51.5 (11.3)	12.7 (4.5)
$X_{i1} = 4, X_{i2} = 4.9$	50.6 (11.4)	50.7 (11.4)	12.7 (4.5)
Three recurrences			
$X_{i1} = 1, X_{i2} = 2, X_{i3} = 3$	67.4 (11.9)	68.9 (11.4)	12.7 (4.5)
$X_{i1} = 1, X_{i2} = 2.5, X_{i3} = 4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)
$X_{i1} = 3, X_{i2} = 4, X_{i3} = 4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)

Prediction values - between 5 and 10 years

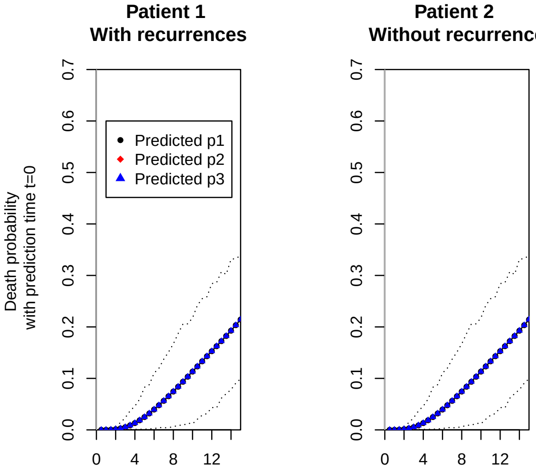
Recurrence history	Risk of death between 5 and 10 years		
	$P^1(5, 10; \hat{\xi})$	$P^2(5, 10; \hat{\xi})$	$P^3(5, 10; \hat{\xi})$
No recurrence	10.8 (4.2)	10.8 (4.2)	12.7 (4.5)
One recurrence			
$X_{i1} = 1$	30.3 (8.9)	33.3 (8.9)	12.7 (4.5)
$X_{i1} = 2.5$	30.3 (8.9)	32.3 (8.9)	12.7 (4.5)
$X_{i1} = 4.9$	30.3 (8.9)	30.4 (8.9)	12.7 (4.5)
Two recurrences			
$X_{i1} = 1, X_{i2} = 2$	50.6 (11.4)	53.2 (11.1)	12.7 (4.5)
$X_{i1} = 2, X_{i2} = 4$	50.6 (11.4)	51.5 (11.3)	12.7 (4.5)
$X_{i1} = 4, X_{i2} = 4.9$	50.6 (11.4)	50.7 (11.4)	12.7 (4.5)
Three recurrences			
$X_{i1} = 1, X_{i2} = 2, X_{i3} = 3$	67.4 (11.9)	68.9 (11.4)	12.7 (4.5)
$X_{i1} = 1, X_{i2} = 2.5, X_{i3} = 4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)
$X_{i1} = 3, X_{i2} = 4, X_{i3} = 4.9$	67.4 (11.9)	67.5 (11.9)	12.7 (4.5)

Prediction values - **between 5 and 15 years**

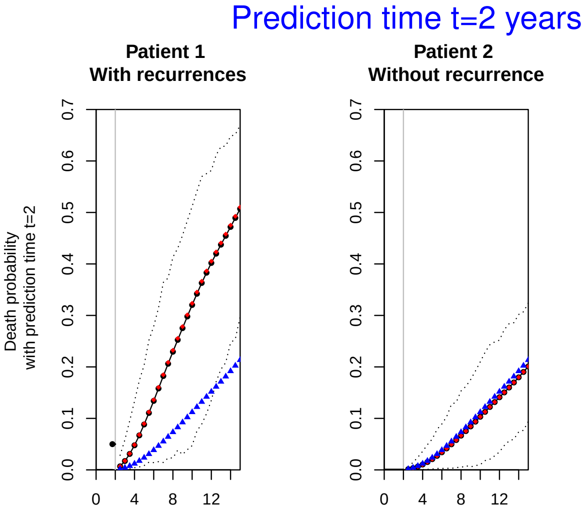
Recurrence history	Risk of death between 5 and 15 years		
	$P^1(5, 15; \hat{\xi})$	$P^2(5, 15; \hat{\xi})$	$P^3(5, 15; \hat{\xi})$
No recurrence	22.7 (4.8)	22.7 (4.8)	25.6 (4.7)
One recurrence			
$X_{i1} = 1$	53.0 (6.9)	56.2 (6.3)	25.6 (4.7)
$X_{i1} = 2.5$	53.0 (6.9)	55.2 (6.5)	25.6 (4.7)
$X_{i1} = 4.9$	53.0 (6.9)	53.1 (6.9)	25.6 (4.7)
Two recurrences			
$X_{i1} = 1, X_{i2} = 2$	75.6 (6.0)	77.5 (5.4)	25.6 (4.7)
$X_{i1} = 2, X_{i2} = 4$	75.6 (6.0)	76.3 (5.8)	25.6 (4.7)
$X_{i1} = 4, X_{i2} = 4.9$	75.6 (6.0)	75.6 (6.0)	25.6 (4.7)
Three recurrences			
$X_{i1} = 1, X_{i2} = 2, X_{i3} = 3$	88.4 (4.1)	89.2 (3.7)	25.6 (4.7)
$X_{i1} = 1, X_{i2} = 2.5, X_{i3} = 4.9$	88.4 (4.1)	88.4 (4.1)	25.6 (4.7)
$X_{i1} = 3, X_{i2} = 4, X_{i3} = 4.9$	88.4 (4.1)	88.4 (4.1)	25.6 (4.7)

Death prediction for 2 particular cases

Baseline prediction :
between 40 and 55 years, no peritumoral vasc. invasion, tumor size $\leq$ 20 mm, HER2 -, RH +, no lymph node involv., grade II

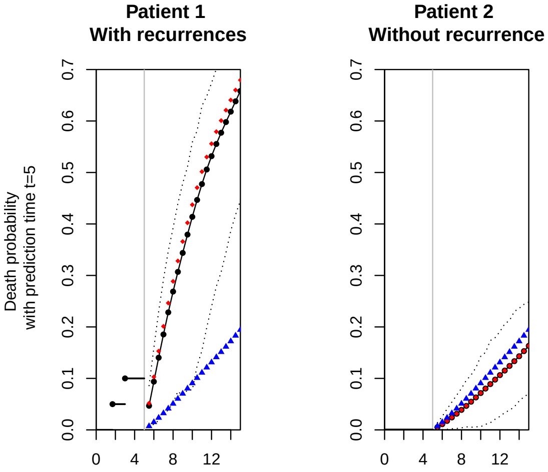


Death prediction for 2 particular cases

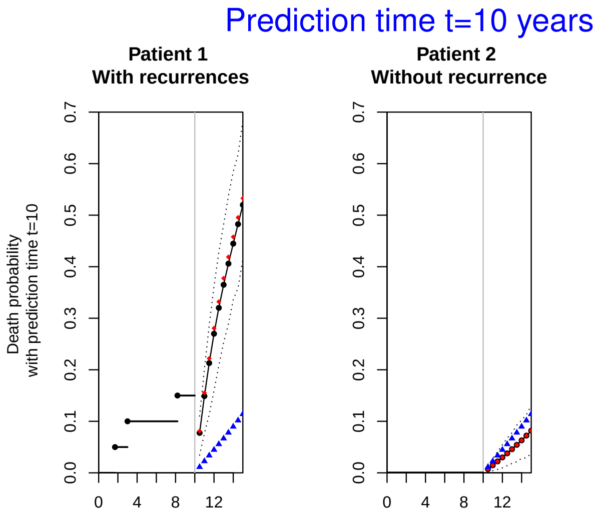


Death prediction for 2 particular cases

Prediction time t=5 years

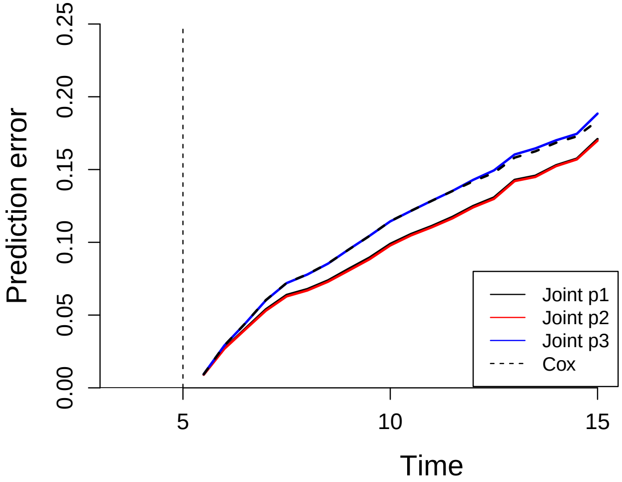


Death prediction for 2 particular cases



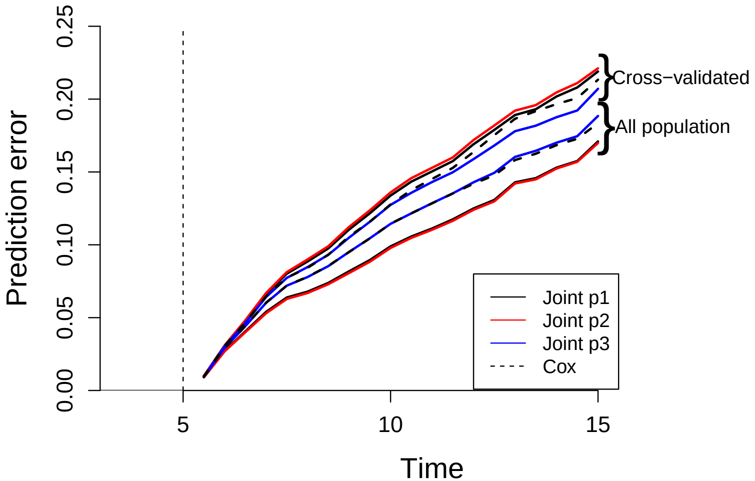
Death prediction error

Prediction at 5 years (949 patients alive)



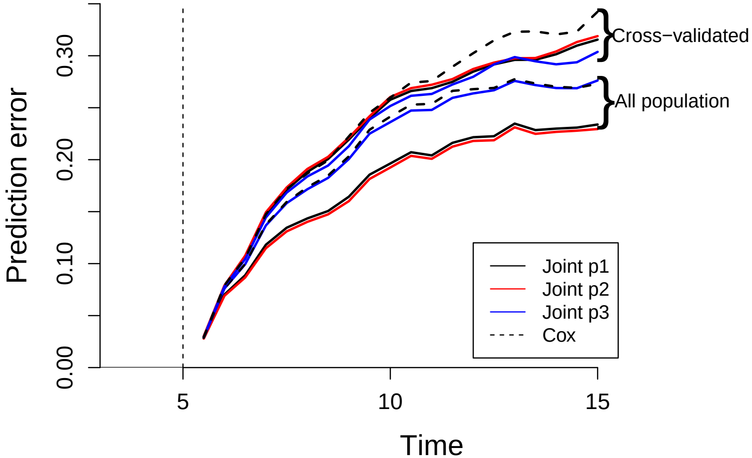
Prediction error

Prediction at 5 years (949 patients alive), with 10-fold cross-validation



Prediction error

Prediction at 5 years (267 patients alive with recurrence), with 10-fold cross-validation



Conclusion

- Recurrent event process seems interesting to predict the risk of death, in framework of joint models
- Dynamic prediction : updated with new events
- Joint modeling gives better results than Cox model with lower prediction error
- However, the 10-fold cross-validation suggests a higher risk of over-fitting
- Conditional prediction possible, but interest is limited
- Perspective :
    - Independent external validation (See A. Mauguen poster)
    - To study the prediction of the risk of events along with the risk of death
    - Prediction using alternative models ?

References

- Liu et al. (2004). Shared frailty models for recurrent events and a terminal event. ***Biometrics***

- Rondeau et al. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation : application on cancer events. ***Biostatistics***

- Mauguen et al. (2013). Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models ***Stat Med***

- Gerds et al. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. ***Biometrical J***

- Proust-Lima et al. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA : a joint modeling approach. ***Biostatistics***

- Rizopoulos et al. (2011) Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. ***Biometrics***

- Brier (1950). ***Monthly weather review***

- Rondeau V, and Gonzalez, JR. (2012) FRAILTYPACK : An R package for the analysis of correlated data with frailty models using the penalized likelihood estimation. ***JSS.***

http ://cran.r-project.org/web/packages/**frailtypack**/
http ://cran.r-project.org/web/packages/pec/

Supported by : INCA and Institut Bergonié