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

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

** INSERM U897, Bordeaux*
*** London School of Hygiene and Tropical Medicine, London*
**** Integraal Kankercentrum Netherlands*
***** Institut Bergonié, Bordeaux*

2 December 2013, London

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

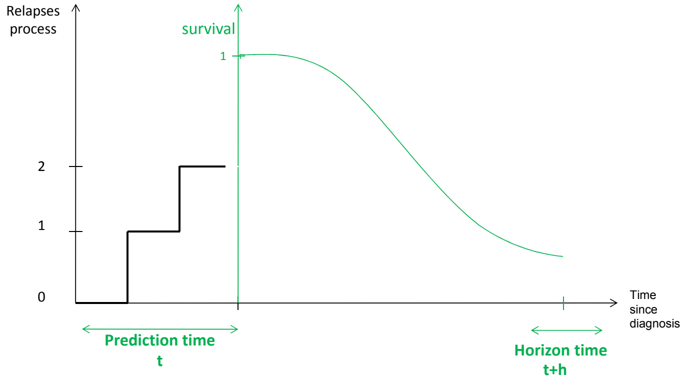
- **After a breast cancer diagnosis**
  $\rightarrow$ single or multiple events
  (recurrences, metastases, death)
- **Prediction of death**
  $\rightarrow$ clinical therapeutic decisions, and patient monitoring
  $\rightarrow$ patient information
  $\rightarrow$ trials : defining patient subpopulations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



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 modelling 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

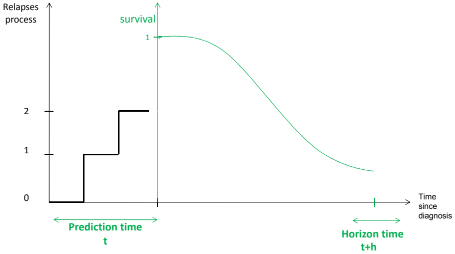
$$pl(\xi) = l(\xi) - \kappa_1 \int_0^\infty (\lambda_0^R(t))''^2 \mathrm{d}t - \kappa_2 \int_0^\infty (\lambda_0^D(t))''^2 \mathrm{d}t$$

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 :

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

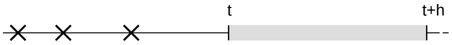


Dynamic prediction

Distinguish **two settings** for the probability of death

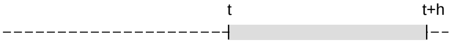
Setting 1

Exactly 3 recurrent events before t



Setting 2

Whatever the history of recurrent events before t



Recurrent event

Window of prediction of death

Period where we consider what happens

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*

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

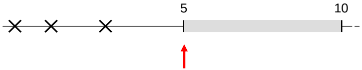
$$= \frac{\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)](u_i)^J S_{i(J+1)}^R(t|Z_{ij}^R, u_i, \xi) g(u_i) \mathrm{d}u_i}{\int_0^\infty S_i^D(t|Z_i^D, u_i, \xi)(u_i)^J S_{i(J+1)}^R(t|Z_{ij}^R, u_i, \xi) g(u_i) \mathrm{d}u_i}$$

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 Mrs Martin has developed 3 recurrences of her initial cancer, her probability of dying in the next 5 years is x%"

Exactly 3 recurrent events before t



Dynamic prediction

Setting 2 : considering the recurrence history only in the parameters estimation

$$
\begin{aligned}
& P^2(t, t+h; \xi) \\
& = P(D_i \leq t+h | D_i > t, Z_i^D, \xi) \\
& = \frac{\displaystyle\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)] g(u_i) \mathrm{d}u_i}{\displaystyle\int_0^\infty S_i^D(t|Z_i^D, \xi, u_i) g(u_i) \mathrm{d}u_i}
\end{aligned}
$$

Dynamic prediction

Setting 2 : considering the recurrence history only in the parameters estimation

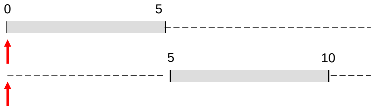
$$
\begin{aligned}
&P^2(t, t+h; \xi) \\
&= P(D_i \le t+h | D_i > t, Z_i^D, \xi) \\
&= \frac{\int_0^\infty [S_i^D(t|Z_i^D, u_i, \xi) - S_i^D(t+h|Z_i^D, u_i, \xi)] g(u_i) \mathrm{d}u_i}{\int_0^\infty S_i^D(t|Z_i^D, \xi, u_i) g(u_i) \mathrm{d}u_i}
\end{aligned}
$$

Example :

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

" if still alive in 5 years, her 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 estimator of a 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

Dynamic prediction : Error of prediction

To be able to compare different populations : residual error R^2

$$R^2 = 1 - Err_{t+h}/Err^0_{t+h}$$

with Err_{t+h} as previously defined
Err^0_{t+h} the prediction error from a Kaplan-Meier model (average survival predicted for each patient)

Graf. Stat Med 1999

Application

1. On the French cohort

Development cohort

- Model development
  - Variable selection
  - Parameters estimation
- Internal validation of the prediction
  - Apparent error
  - Cross-validated error

French cohort

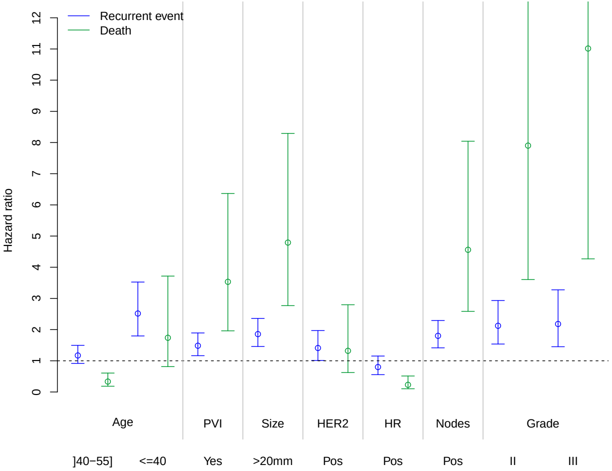
- 1067 patients
- median follow-up : 13.8 years (min=5 months)
- 427 recurrent events (locoregional relapses and distant metastases) in 362 patients (mean 0.40)

N events	0	1	2	3	All
Alive	600	114	20	3	737
Died	105	187	37	1	330
All	705	301	57	4	1067

with the R package **frailtypack**

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

Prognostic joint model



θ=1.03 (se=0.06) and α=4.66 (se=0.28)

Prediction values between 5 and 10 years

Recurrence history	$P^{Recurrence}(5, 10; \hat{\xi})$	$P^{Ignoring}(5, 10; \hat{\xi})$
No recurrence	10.8 (4.2)	12.7 (4.5)
One recurrence	30.3 (8.9)	12.7 (4.5)
Two recurrences	50.6 (11.4)	12.7 (4.5)
Three recurrences	67.4 (11.9)	12.7 (4.5)

For a given patient : age > 55y, no PVI, size ≤ 20mm, HER2 negative, HR positive, no lymph node involvement, grade II.

Prediction values between 5 and 15 years

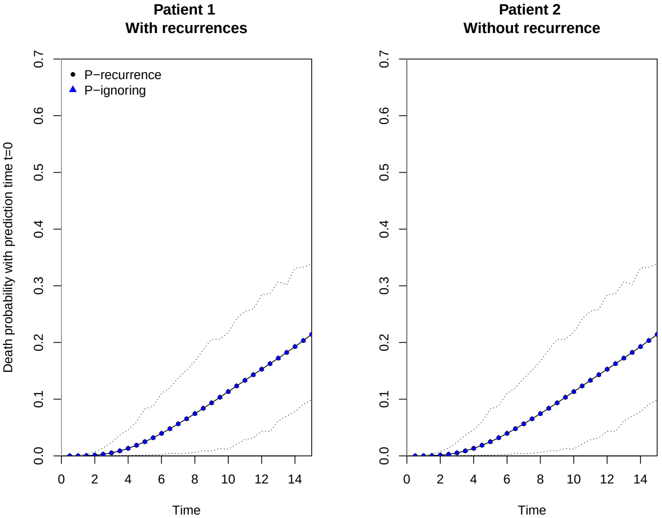
Recurrence history	$P^{Recurrence}(5, 15; \hat{\xi})$	$P^{Ignoring}(5, 15; \hat{\xi})$
No recurrence	22.7 (4.8)	25.6 (4.7)
One recurrence	53.0 (6.9)	25.6 (4.7)
Two recurrences	75.6 (6.0)	25.6 (4.7)
Three recurrences	88.4 (4.1)	25.6 (4.7)

For a given patient : age > 55y, no PVI, size $\leq$ 20mm, HER2 negative, HR positive, no lymph node involvement, grade II.

Death prediction for 2 particular cases

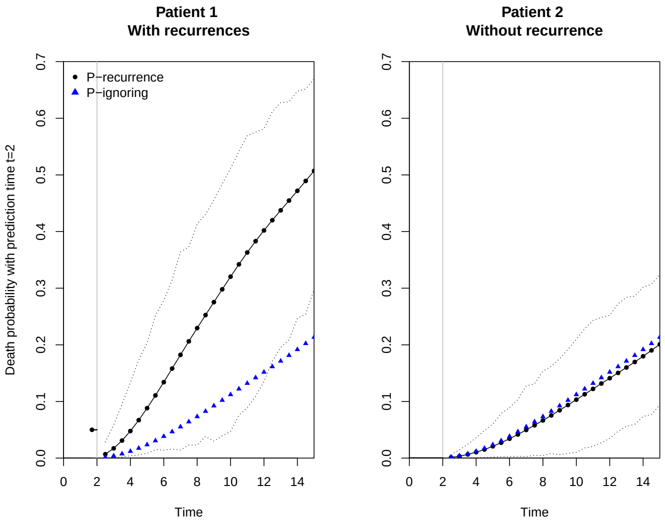
Baseline prediction

between 40 and 55 y, no peritum. vasc. invasion, tumour size $\leq$ 20 mm, HER2 -, HR +, no lymph node involv., grade II



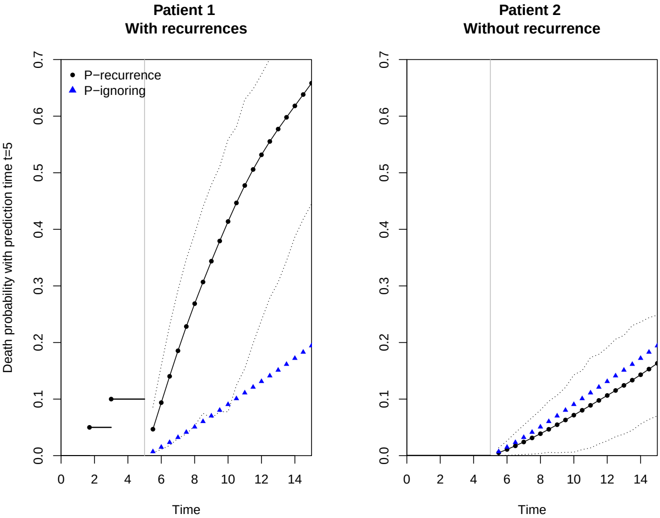
Death prediction for 2 particular cases

Prediction time t=2 years



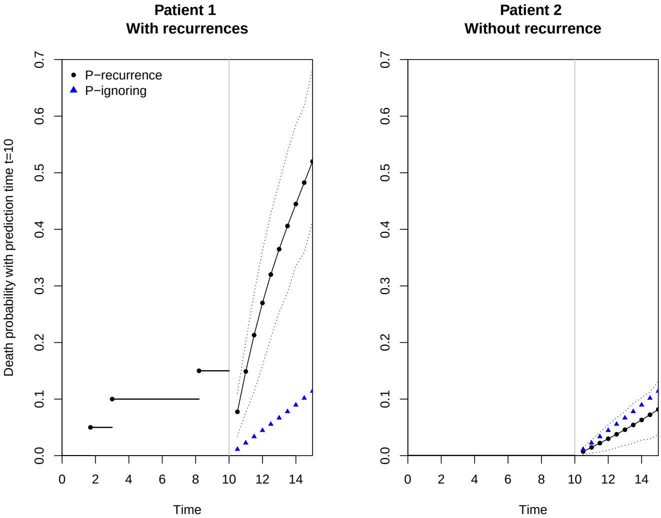
Death prediction for 2 particular cases

Prediction time t=5 years



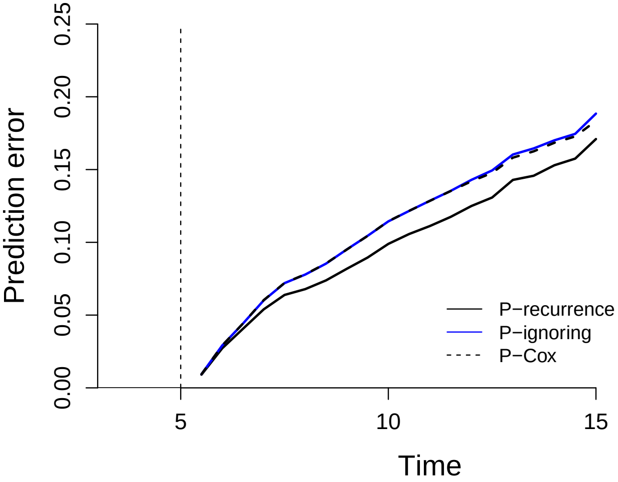
Death prediction for 2 particular cases

Prediction time t=10 years



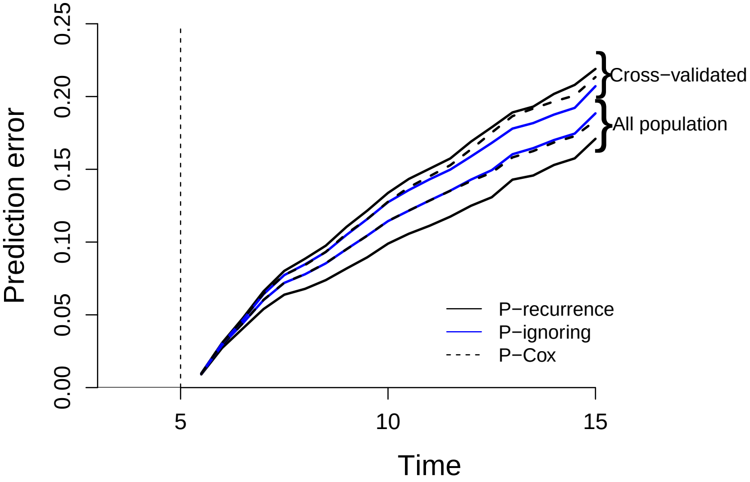
Death prediction error

Prediction at 5 years (949 patients alive)



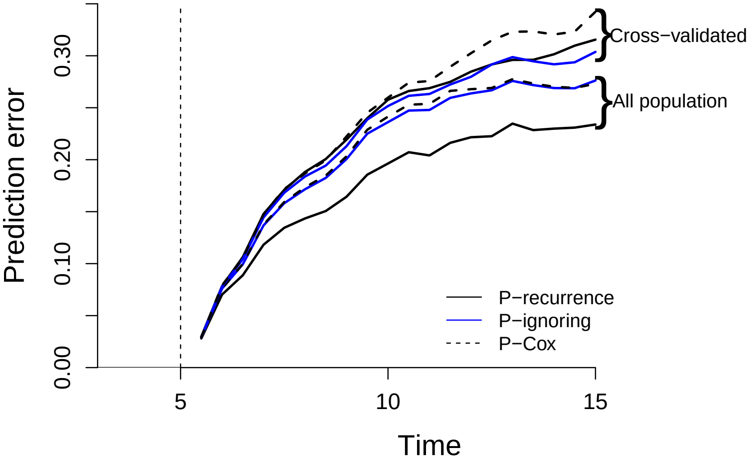
Prediction error

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



At this step

- Found the prognostic factors of interest
- Estimated parameters (factor effects, correlation between the two endpoints)
- Were able to account for relapses in the prediction of the risk of death
- Not clear whether accounting for relapses has an interest for prediction

Application

2. External validation

External validation - why ?

- Model designed to perform well on development data
  - problem with the design or methods
  - absence of an important predictor
- To check the reproducibility of the model and predictions
  - overfitting
    $\rightarrow$ correct for optimism
  - difference case-mix
- To update the proposed prognostic model

Models to be compared

- Joint frailty model
  - \+ One model $\rightarrow$ dynamic prediction
  - \+ Correlation between the two processes fully accounted for
  - \- more parameters $\rightarrow$ less stability

- Landmark Cox model
  - \+ Robust and simple model
  - \+ Time-dependent effects
  - \- One model for each prediction time t
  - \- Information about recurrent events : number of recurrent events

Populations - description

West Midlands

- 1196 subjects
- Diagnosed in 1996
- Follow-up : 16 years
- 376 relapses in 301 patients (mean=0.31)
- 613 deaths (51%)

Dutch registry

- 31,075 subjects
- Diagnosed in 2003-2006
- median follow-up : 7.7 y
- 3854 relapses in 3844 patients (mean=0.12)
- 7162 deaths (23%)

Populations - missing data

- Missing data problem not much discussed in the literature in that context
- Not an effect estimation problem
- Clinical point of view
  $\rightarrow$ complete case analysis

West Midlands

- 1196 subjects
- from 3168 cases (38%)
- HER2 and hormonal receptor unavailable

Dutch registry

- 31,075 subjects
- from 41,676 cases (75%)
- HER2 and hormonal receptor unavailable
- Perivascular invasion unavailable

Populations - Relapses definitions

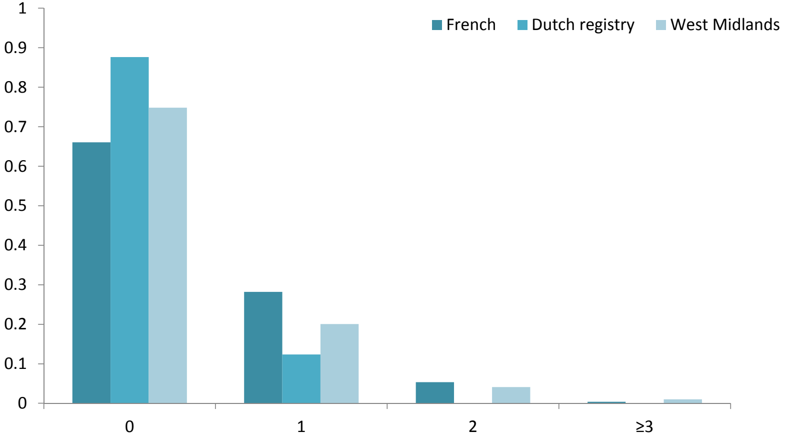
West Midlands

- Recurrence defined from treatment
- 376 relapses
  - 22% $<$2 years
  - 59% $<$5 years

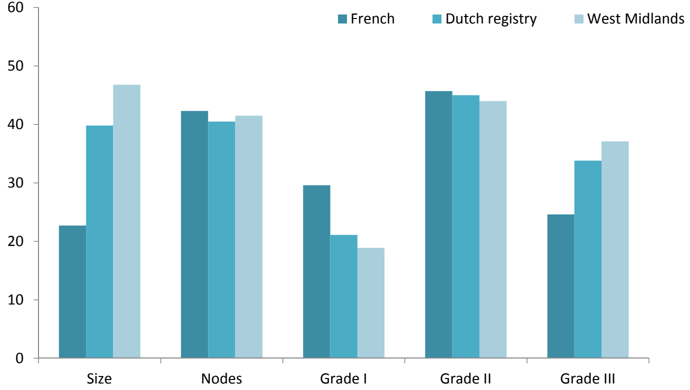
Dutch registry

- Recurrences recorded (only the 1st one of each type)
- 3854 relapses
  - 41% $<$2 years
  - 93% $<$5 years

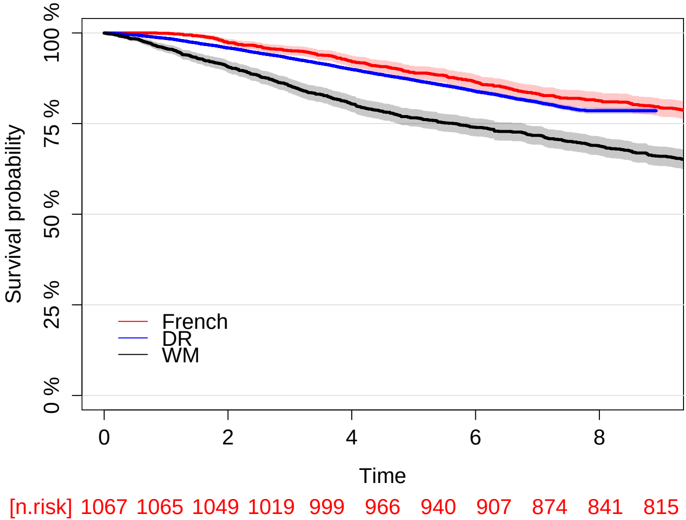
Populations - recurrent event



Populations - prognostic factors

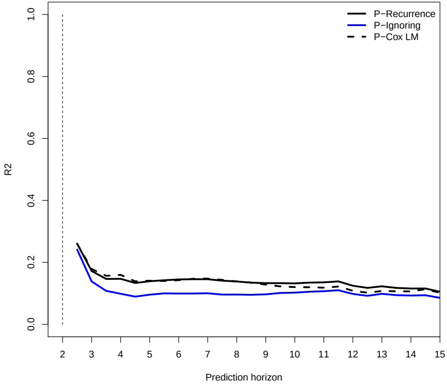


Populations - overall survival

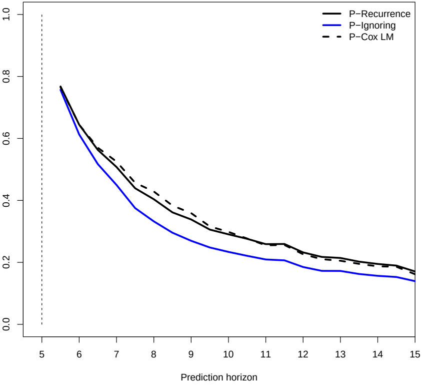


West Midlands population

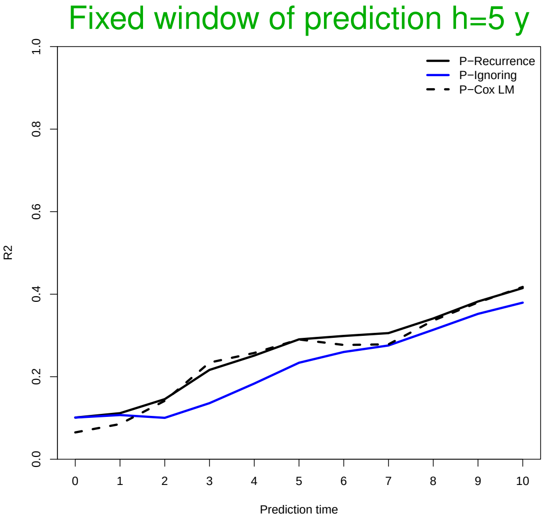
t=2 years



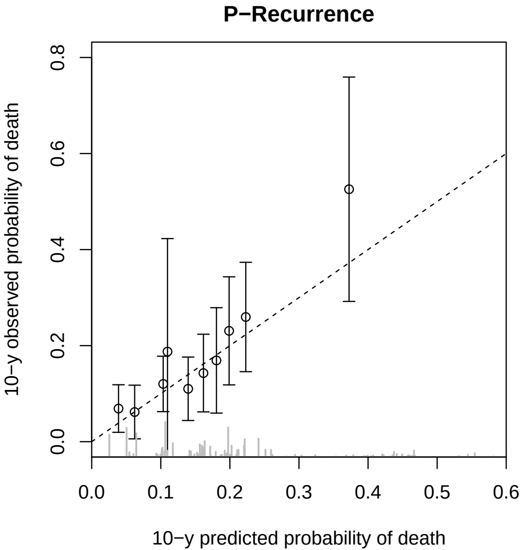
t=5 years



West Midlands population

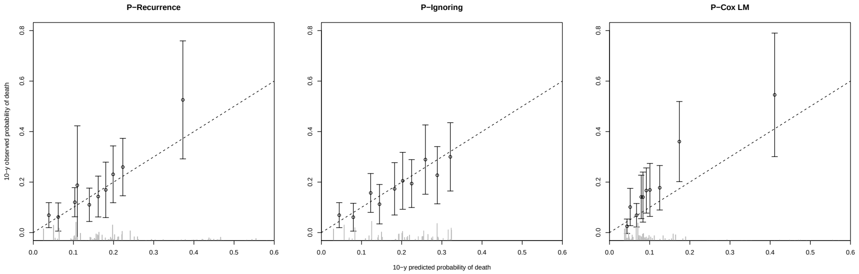


West Midlands population - Calibration at 10 years (t=5 years)



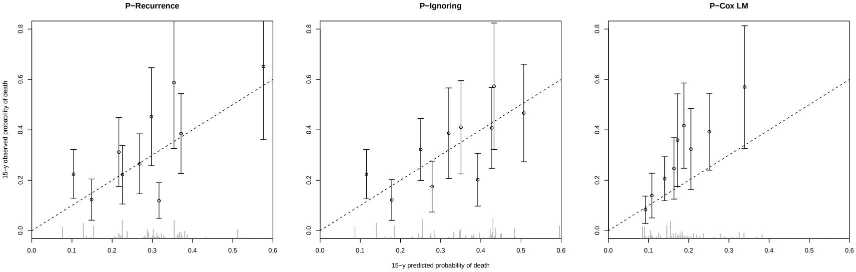
West Midlands population

Calibration at 10 years (t=5 years)



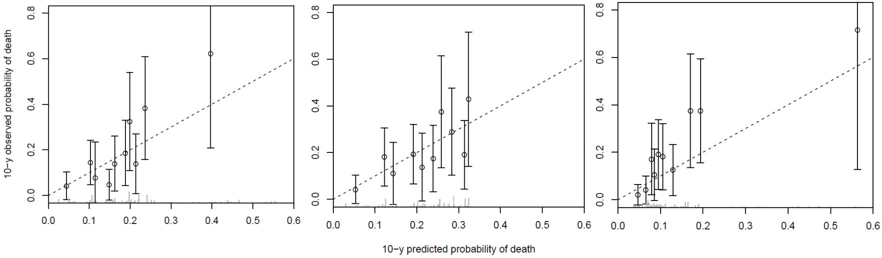
West Midlands population

Calibration at 15 years (t=5 years)



Subgroup analysis

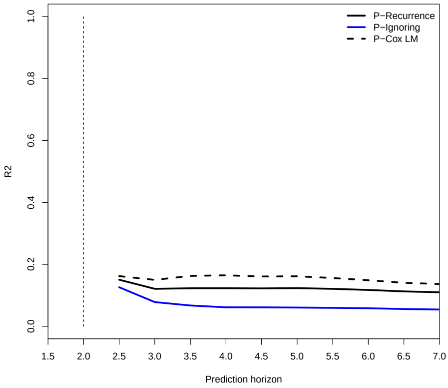
West Midlands population - operated patients

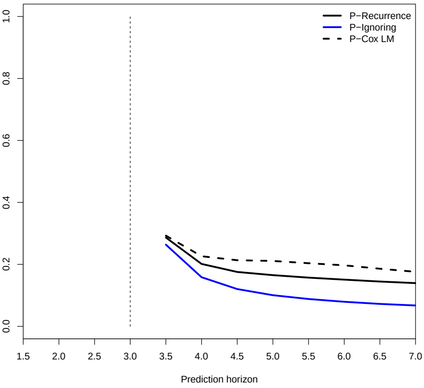


Dutch population

t=2 years

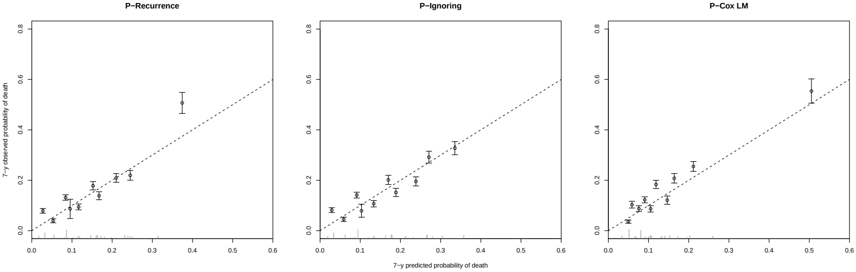
t=3 years





Dutch population

Calibration at 7 years (*t*=2 years)



At the end

- Relapses information is useful to predict the death of patients with breast cancer

- The more information, the better
  relapses information prior to 2-3 years not enough

- Two approaches (joint and landmark) give similar performance
  $\rightarrow$ Do not be afraid to use complex model (with more parameters) in prediction if needed

At the end

- The model estimated on a selected cohort of patients can be useful in more general populations
  - Good performance in West Midlands population despite
    - a different survival in the population
    - a different period of inclusion
    - a different case-mix
  - Prediction not good in Dutch registry patients
    - Short follow-up
    - Patient recently diagnosed
      impact of change in the clinical practice ?

And then ?

- Considering the type of recurrence
  Different effect of loco-regional relapse and metastasis on the risk of death

- Predict the risk of recurrence
  For example, risk of metastasis considering the previous loco-regional relapses

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***

- 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 : Institut National du Cancer and Institut Bergonié