Dynamic predictions from joint models and their evaluation

Cécile Proust-Lima

INSERM U897, Epidemiologie et Biostatistique, Bordeaux, France
Univ. Bordeaux Segalen, ISPED, Bordeaux, France

cecile.proust-lima@inserm.fr

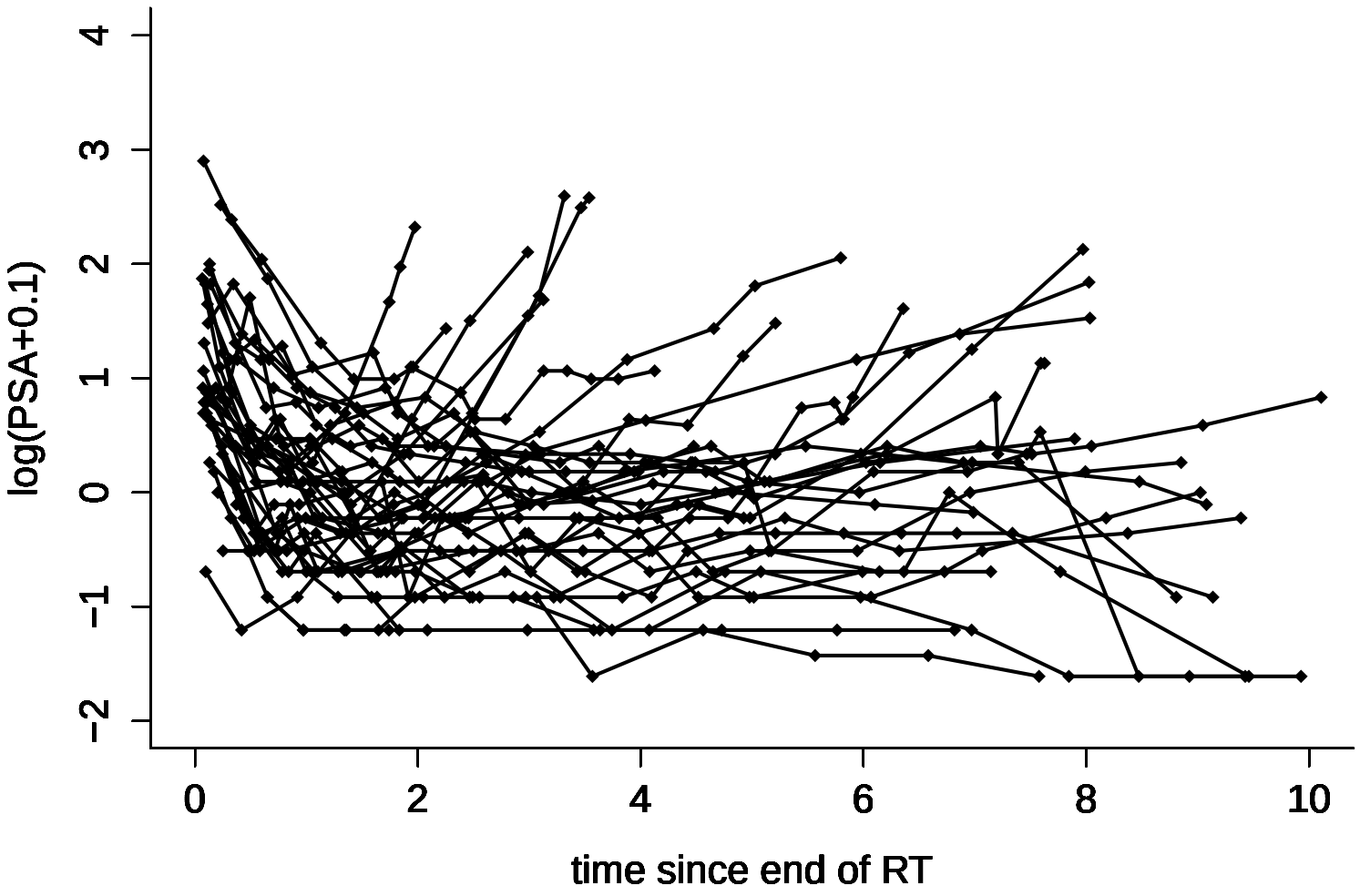
Dynamic predictions for repeated markers and repeated events
October 11th, 2013 - Bordeaux, France

Motivating example

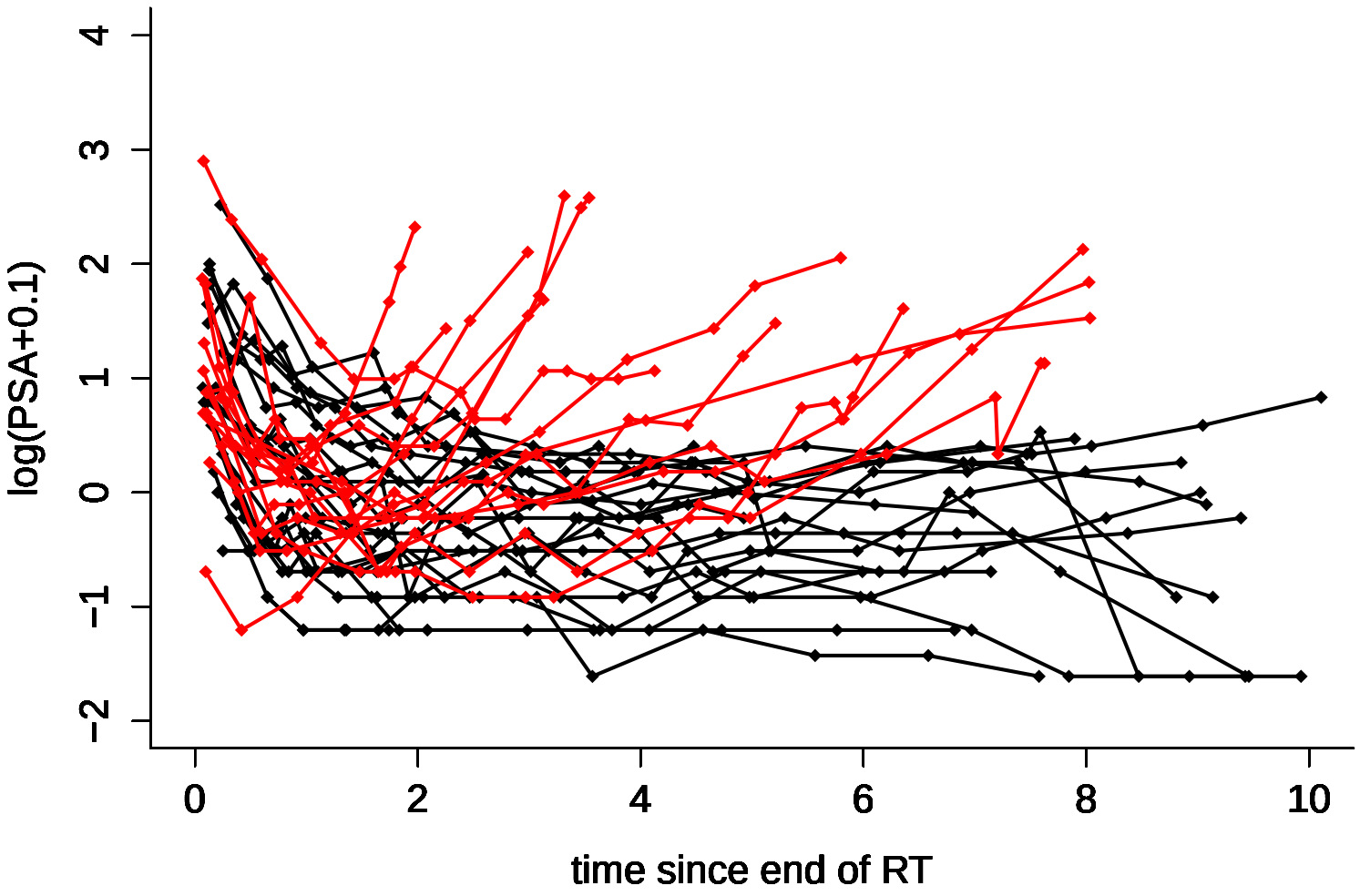
Monitoring of prostate cancer progression after radiation therapy (RT) :

- Prostate Specific Antigen (PSA)
  $\longrightarrow$ wellknown biomarker of prostate cancer progression
- Clinical recurrence of prostate cancer predicted using :
  $\longrightarrow$ diagnosis information : initial log PSA, Gleason score and T-stage
  $\longrightarrow$ last value of PSA or PSA summary (PSA doubling time)

PSA individual trajectories



PSA individual trajectories



Motivating example

Monitoring of prostate cancer progression after radiation therapy (RT) :

- Prostate Specific Antigen (PSA)
  $\longrightarrow$ wellknown biomarker of prostate cancer progression
- Clinical recurrence of prostate cancer predicted using :
  $\longrightarrow$ diagnosis information : initial log PSA, Gleason score and T-stage
  $\longrightarrow$ last value of PSA or PSA summary (PSA doubling time)

Motivating example

Monitoring of prostate cancer progression after radiation therapy (RT) :

- Prostate Specific Antigen (PSA)
  $\longrightarrow$ wellknown biomarker of prostate cancer progression
- Clinical recurrence of prostate cancer predicted using :
  $\longrightarrow$ diagnosis information : initial log PSA, Gleason score and T-stage
  $\longrightarrow$ last value of PSA or PSA summary (PSA doubling time)

Problems :

- whole PSA trajectory highly associated with the risk of recurrence
- PSA = internal noisy time-dependent covariate (see Dimitris's talk)

Solution :

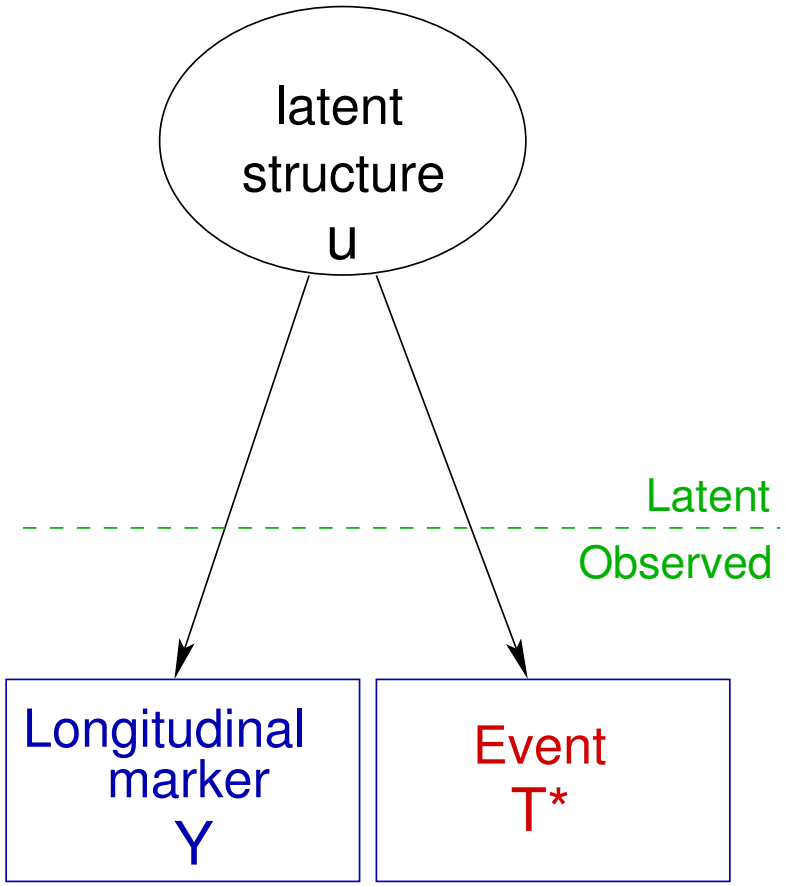
$\longrightarrow$ focus on joint models for computing & evaluating dynamic prognostic tools of prostate cancer recurrence based on the PSA trajectory

Outline of the talk

- Principle of the joint models
- Computation of individual dynamic predictions
- Evaluation of the predictive performances
- Illustration on real data
- Prediction under scenarios of treatments
- Prediction in presence of competing risks

Two joint modelling approaches

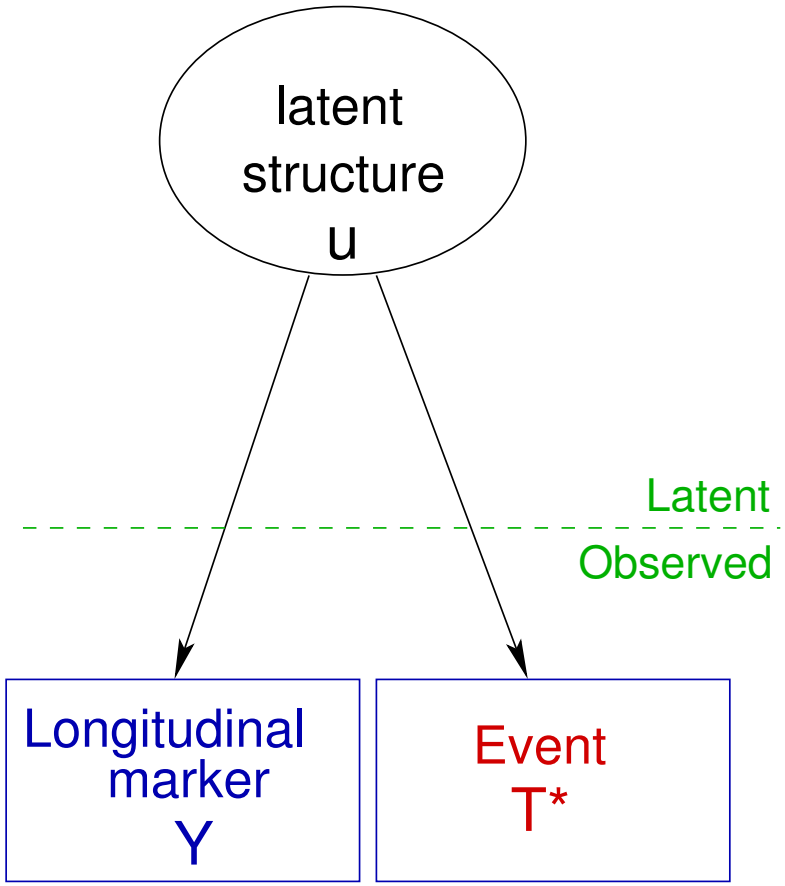
Latent structure u :



Two joint modelling approaches

Latent structure u :

1. marker characteristics
   (current level, slope,
   individual deviation,etc)
   $\longrightarrow$ *shared random-effect models*
   Wulfsohn, 1997 ; Rizopoulos, 2010,2011,...



Shared random-effect model (SREM) (Wulfsohn, Bcs 1997)

Homogeneous linear mixed model :

$$Y_i(t_{ij}) = Z_i(t_{ij})^T u_i + X_{li}(t_{ij})^T \beta + \epsilon_{ij}$$

with $u_i \sim \mathcal{N}(\mu, B)$ & $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$

Proportional hazard model including marker trajectory characteristics :

$$\lambda(t \mid u_i) = \lambda_0(t) e^{X_{ei}(t)^T \delta + f(u_i, \beta)^T \eta}$$

with $f(u_i, \beta) =$ current level, current slope, random deviation, ...

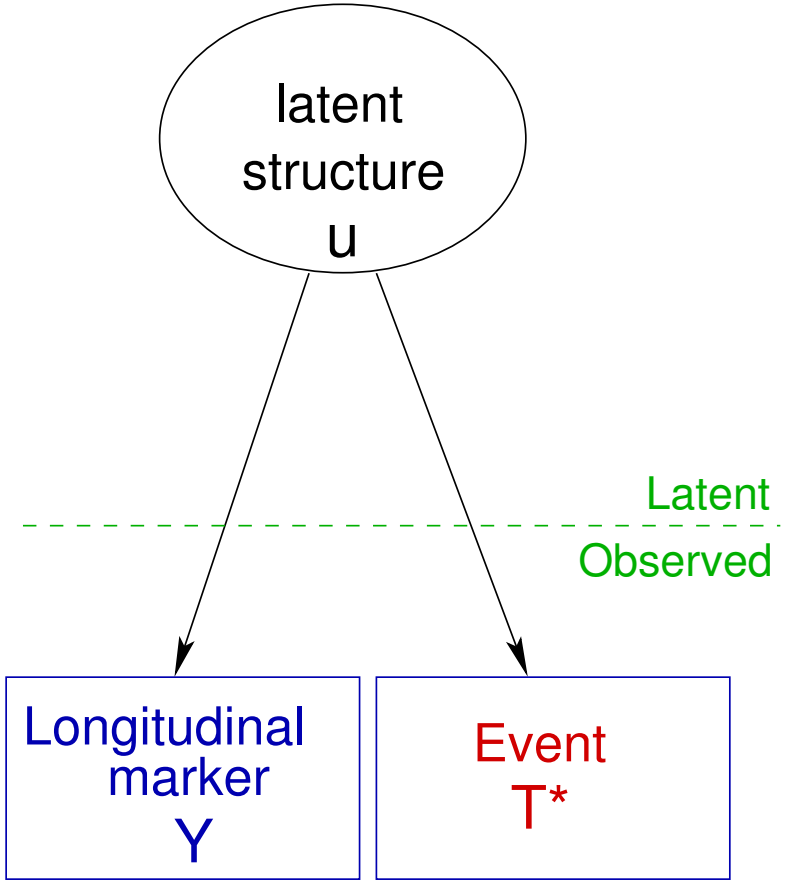
Maximum likelihood estimation
$\rightarrow$ *numerical integration over the random-effects distribution*

Implemented in **JM** R package (Rizopoulos, JSS 2010)

Two joint modelling approaches

Latent structure u :

1. marker characteristics
   (current level, slope,
   individual deviation,etc)
   $\longrightarrow$ *shared random-effect models*
   Wulfsohn, 1997 ; Rizopoulos, 2010,2011,...

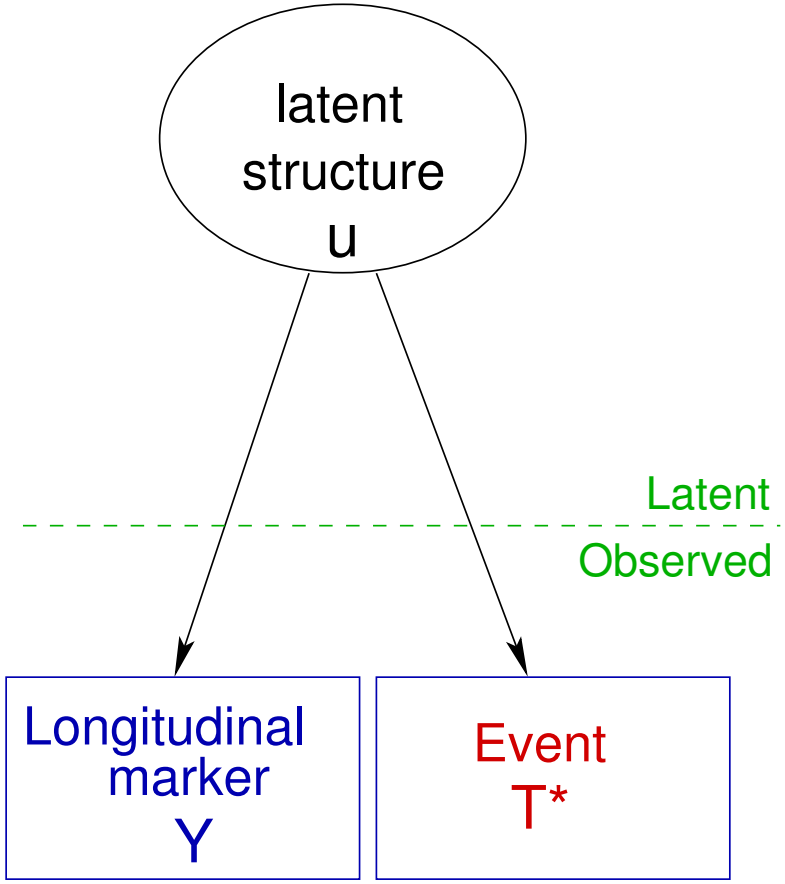


Two joint modelling approaches

Latent structure u :

1. marker characteristics
   (current level, slope,
   individual deviation,etc)
   $\longrightarrow$ *shared random-effect models*
   Wulfsohn, 1997 ; Rizopoulos, 2010,2011,...

2. division of the population
   in homogeneous subgroups
   $\longrightarrow$ *joint latent class models*
   Lin, 2002 ; Proust-Lima, 2009-2012



Joint latent class model (JLCM) (Lin, JASA 2002)

Latent class membership c_i :

$$\pi_{ig} = P(c_i = g|X_{pi}) = \frac{e^{\xi_{0g}+{X_{pi}}^T\xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l}+{X_{pi}}^T\xi_{1l}}} \text{ with } \xi_{0G} = 0 \text{ \& } \xi_{\mathbf{1G}} = \mathbf{0}$$

Class-specific linear mixed model for $Y_i = (Y_i(t_{i1}), ..., Y_i(t_{ij}), ...Y_i(t_{in_i}))$:

$$Y_i(t_{ij}) \mid_{c_i=g} = Z_i(t_{ij})^T u_{i|_{c_i=g}} + X_{li}(t_{ij})^T \beta_g + \epsilon_{ij}$$

with $u_{i|_{c_i=g}} \sim \mathcal{N}\left(\mu_g, \omega_g^2 B\right)$ & $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$

Class-specific proportional hazard model :

$$\lambda(t \mid c_i = g) = \lambda_{0g}(t) e^{X_{ei}(t)\delta_g}$$

Maximum likelihood estimation for a fixed number of classes
→ *Optimal number of latent classes using the BIC, ICL, etc*
Implemented in **lcmm** R package (Proust-Lima, SMMR 2012)

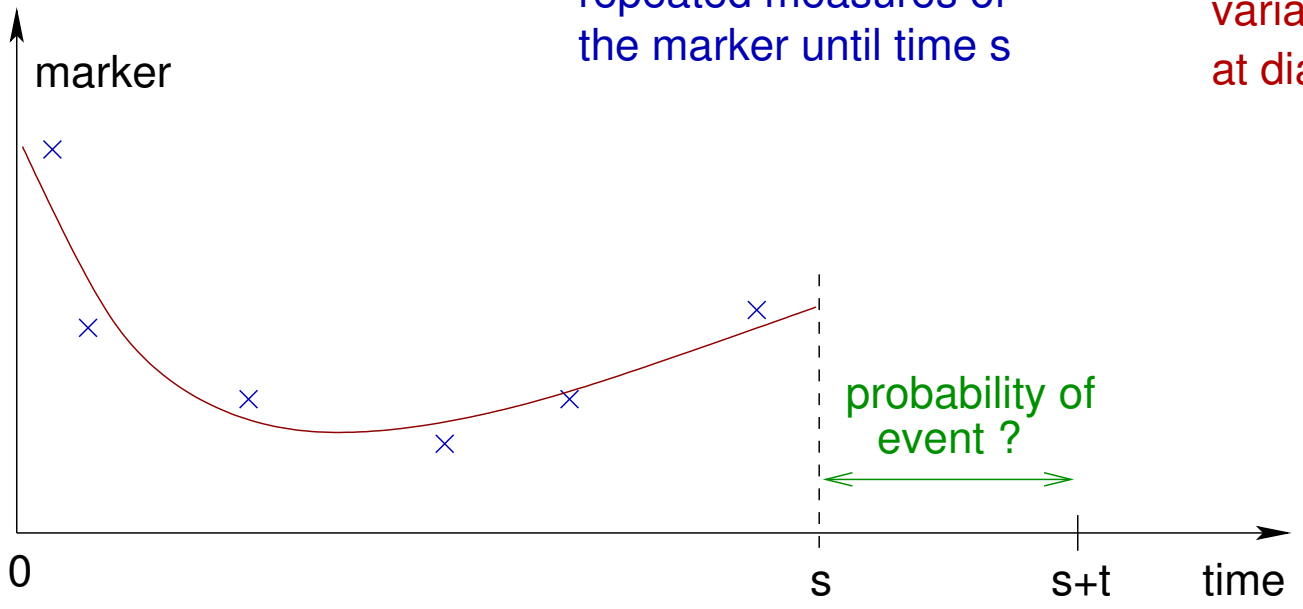
Individual dynamic predictions

Predicted probability of event given information until s, $Y_i^{(s)} = \{Y_i(t_{ij})$ such as $t_{ij} \leq s\}$:

$$P_i(s,t) = P(T_i \leq s + t \mid T_i \geq s, Y_i^{(s)}, X_i)$$

repeated measures of the marker until time s

variables known at diagnosis



Computation from joint models

→ Same formula whatever the shared latent structure :

from a JLCM :

$$P_i(s,t) = \sum_{g=1}^{G} P(T_i \leq s+t \mid T_i \geq s, c_i = g, X_i; \theta) P(c_i = g \mid T_i \geq s, Y_i^{(s)}, X_i; \theta)$$

from a SREM :

$$P_i(s,t) = \int_{u_i} P(T_i \leq s+t \mid T_i \geq s, u_i, X_i; \theta) f(u_i \mid T_i \geq s, Y_i^{(s)}, X_i; \theta) du_i$$

Derived dynamic prognostic tool

From estimated parameters $\hat{\theta}$ and variance $\hat{V(\hat{\theta})}$

For a new subject i, we know the biomarker history $Y_i^{(s)}$ and other covariates X_i

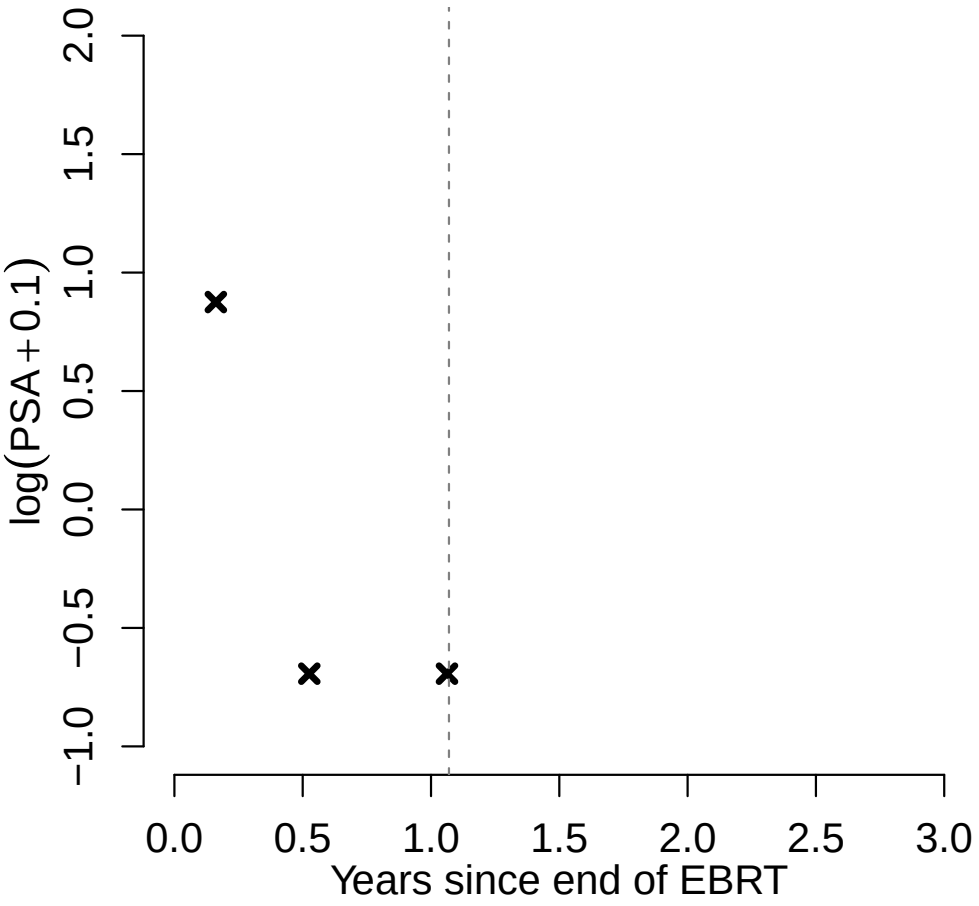
2 strategies :

1. Point estimate of the probability :
   $P_i(s,t)$ computed at $\hat{\theta}$
2. Approximation of the distribution of $P_i(s,t)$:
   D draws $\theta_d \sim \mathcal{N}\left(\hat{\theta}, \hat{V(\hat{\theta})}\right)$
   $P_i(s,t)$ computed in θ_d
   $\rightarrow$ median + 95% bands

Example (with models estimated on 459 men)

For a man with a recurrence at 3.8 years

x PSA measures



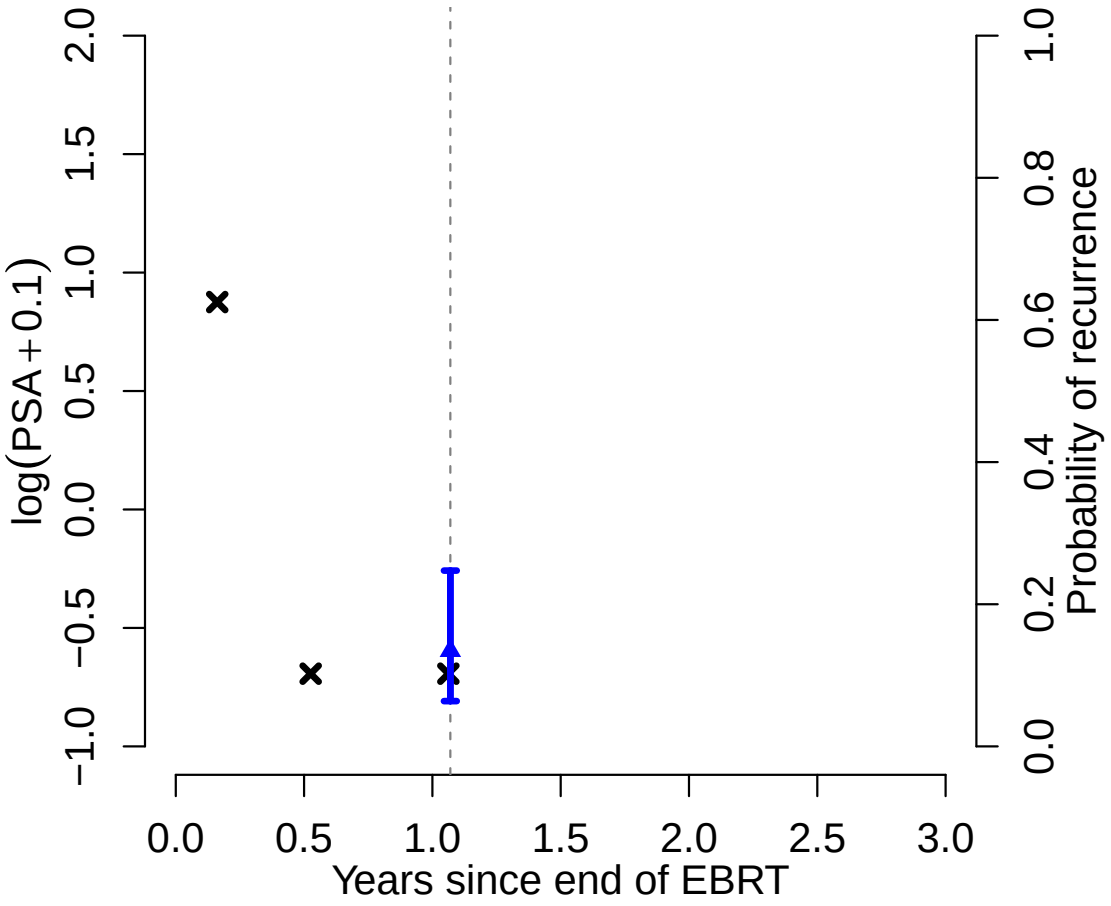
Example (with models estimated on 459 men)

For a man with a recurrence at 3.8 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **with JLCM - 4 classes**



Example (with models estimated on 459 men)

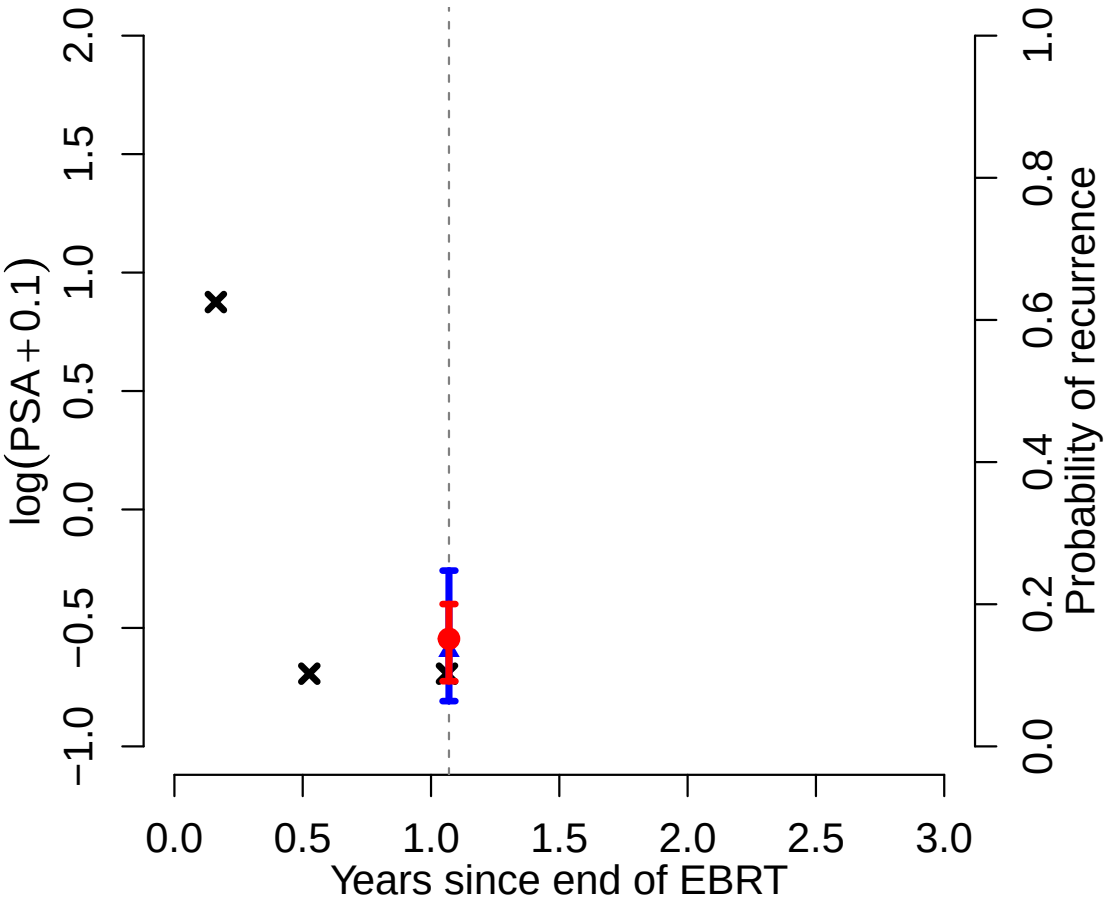
For a man with a recurrence at 3.8 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **with JLCM - 4 classes**

● **with JLCM - 1 class**
(no PSA information)



Example (with models estimated on 459 men)

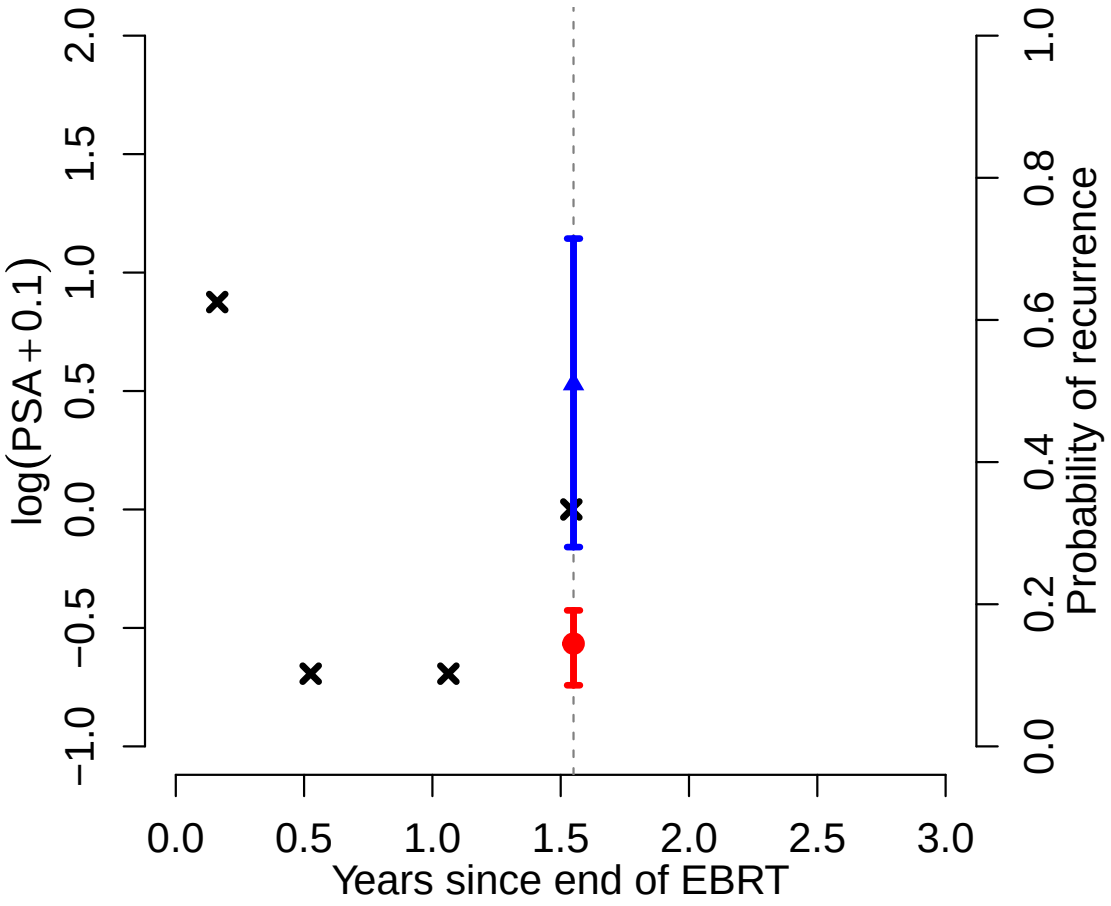
For a man with a recurrence at 3.8 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **with JLCM - 4 classes**

● **with JLCM - 1 class**
(no PSA information)



Example (with models estimated on 459 men)

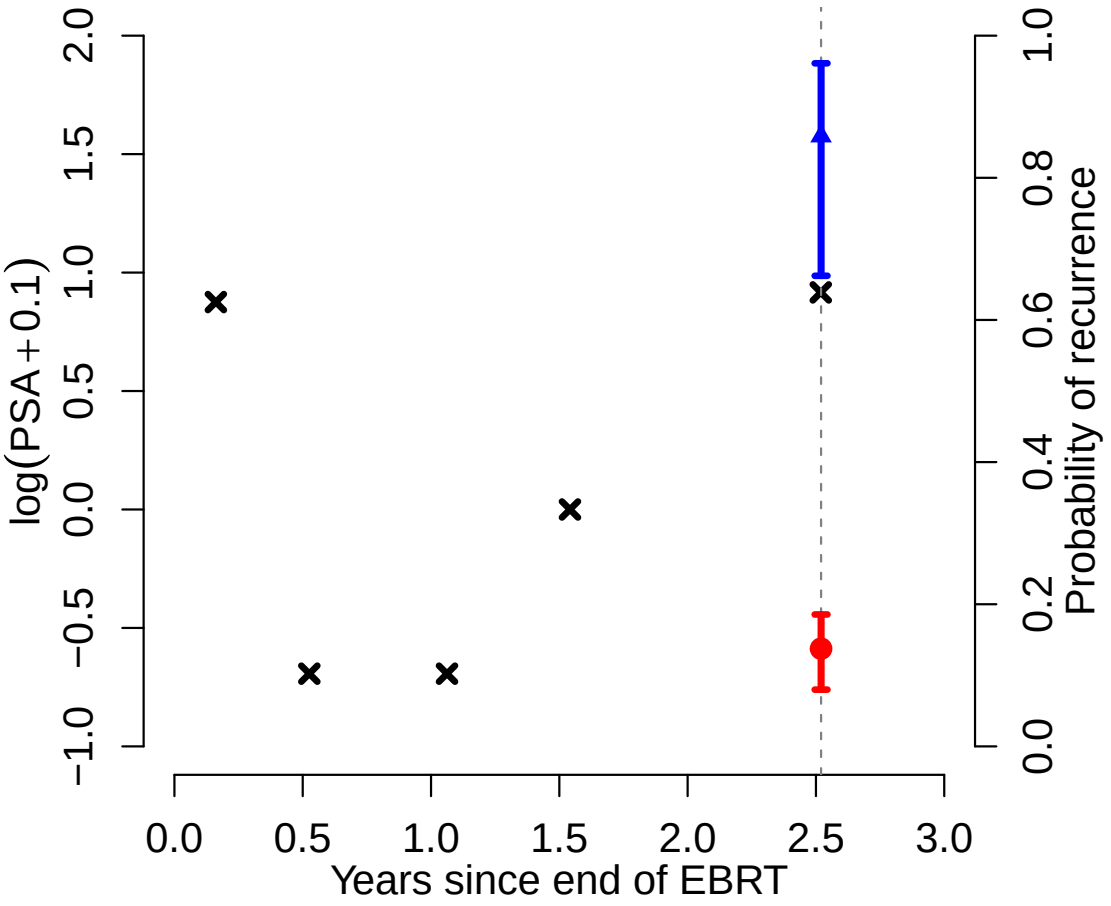
For a man with a recurrence at 3.8 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **with JLCM - 4 classes**

● **with JLCM - 1 class**
(no PSA information)



Example (with models estimated on 459 men)

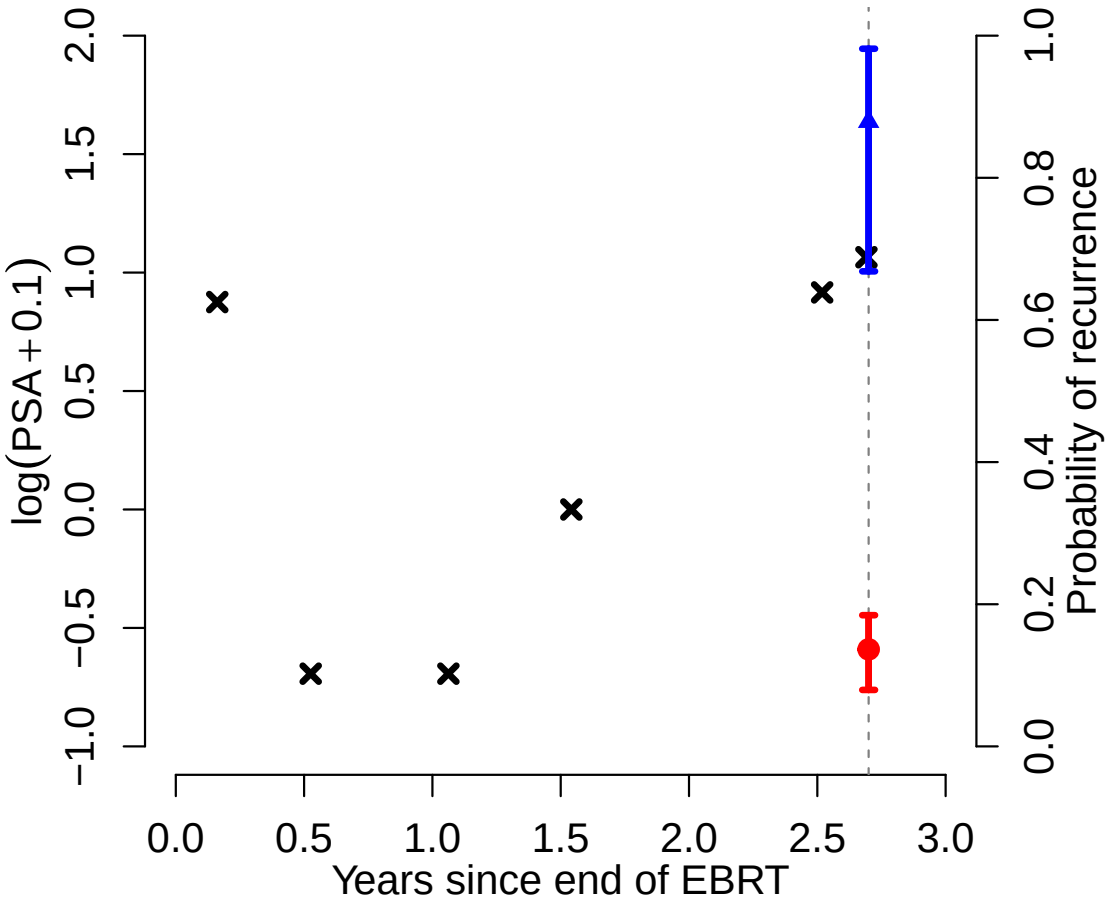
For a man with a recurrence at 3.8 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **with JLCM - 4 classes**

● **with JLCM - 1 class (no PSA information)**



Example (with models estimated on 459 men)

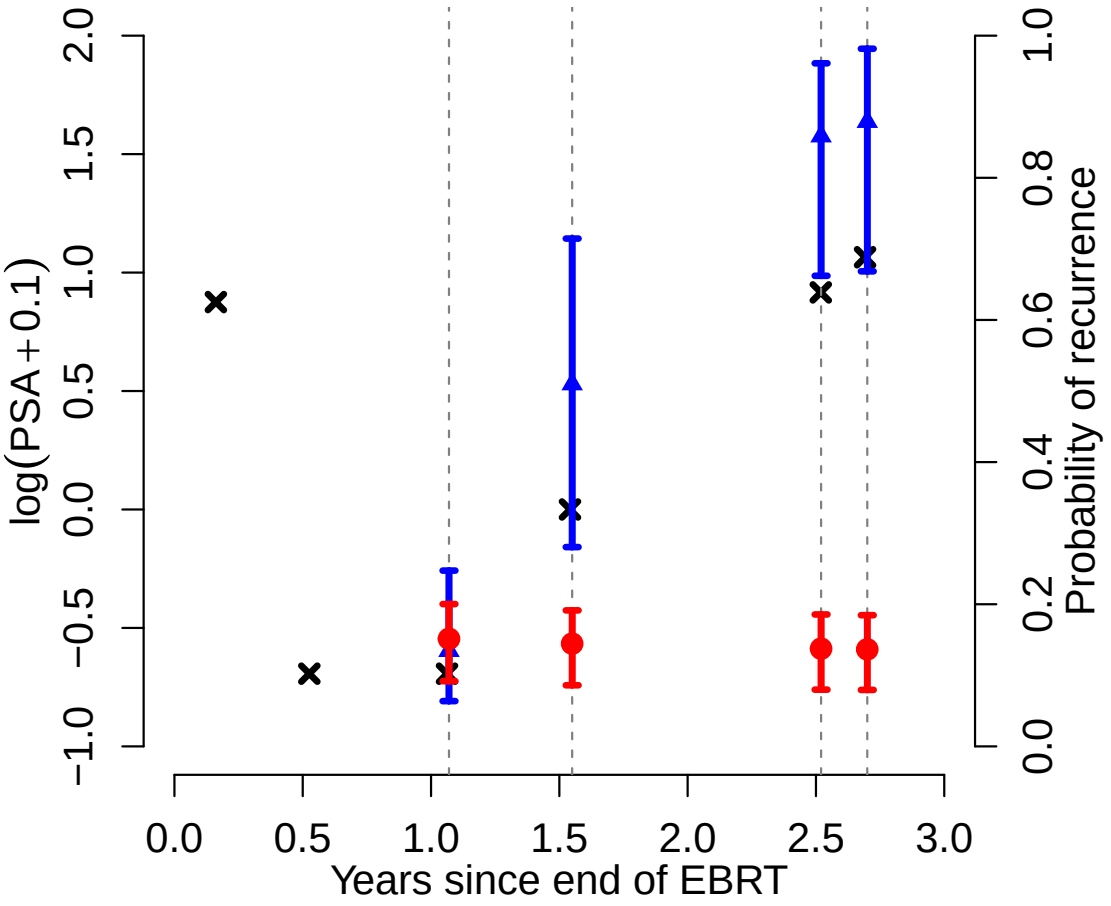
For a man with a recurrence at 3.8 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **with JLCM - 4 classes**

● **with JLCM - 1 class**
(no PSA information)



Measures of predictive accuracy

Does the model predict well the event of interest from a prediction time s ?

- Discriminative power : AUC, Se/Sp
  - $\longrightarrow$ *evaluates the concordance of $\hat{p}_i(s,t)$ with the observations*
    (Zheng, Bcs 2007 ; Rizopoulos, Bcs 2011 ; Paul's talk)
- Error of prediction : Brier score
  - $\longrightarrow$ *compares directly $\hat{p}_i(s,t)$ with the event status $\Upsilon_i(s+t)$*
    (Schoop Bcs 2008 ; Proust-Lima, SMMR 2012 ; Paul's talk)
- Prognostic information : prognostic cross-entropy (EPOCE)
  - $\longrightarrow$ *evaluates the conditional log-density of the event given the biomarker history*
    (Commenges, Bcs 2012 ; Proust-Lima, SMMR 2012 ; Sène, 2013)

Internal and External validation

On which data can we evaluate the predictiveness ?

- Training data (used for the estimation)
  - Apparent measures over evaluate the predictiveness of the model (overoptimism)
    $\rightarrow$ especially important with complex models
  - Correction by cross-validation (Gerds, Bcs 2007)
    $\rightarrow$ very long with complex models
  - Correction by approximated cross-validation
    $\rightarrow$ direct computation - available for EPOCE (Commenges, 2012) and BS (Sène, 2013)
- Validation (external) data
  - Apparent measures OK

Expected prognostic observed cross-entropy (EPOCE) (Commenges, Bcs 2012)

From time of prediction s, $EPOCE(s) = E(-\ln f_{T|Y^{(s)},T^* \geq s} \mid T^* \geq s)$

Estimator on external data :

observed conditional log-likelihood of the time-to-event data from s given the repeated measures until s : $-\mathcal{F}(\hat{\theta}; s)$

Estimator on training data :

observed conditional log-likelihood corrected by approximated cross-validation : $-\mathcal{F}(\hat{\theta}; s) +$ penalty

where penalty accounts for the model complexity (from likelihoods derivatives)

→ *the lower the better*

→ models comparison with the difference + 95% tracking interval

→ evaluation in the remaining (infinite window) or up to an horizon t

Illustration on 459 men treated by radiation therapy (Proust-Lima, 2012)

In the JLCM

- class-specific PSA trajectory in 2 phases (short term drop / long term trend)
- class-specific Weibull baseline risk functions
- covariates in the mixed model, in the survival model & in the latent class probability

→ BIC=5068.4 for 4 classes (G=4)

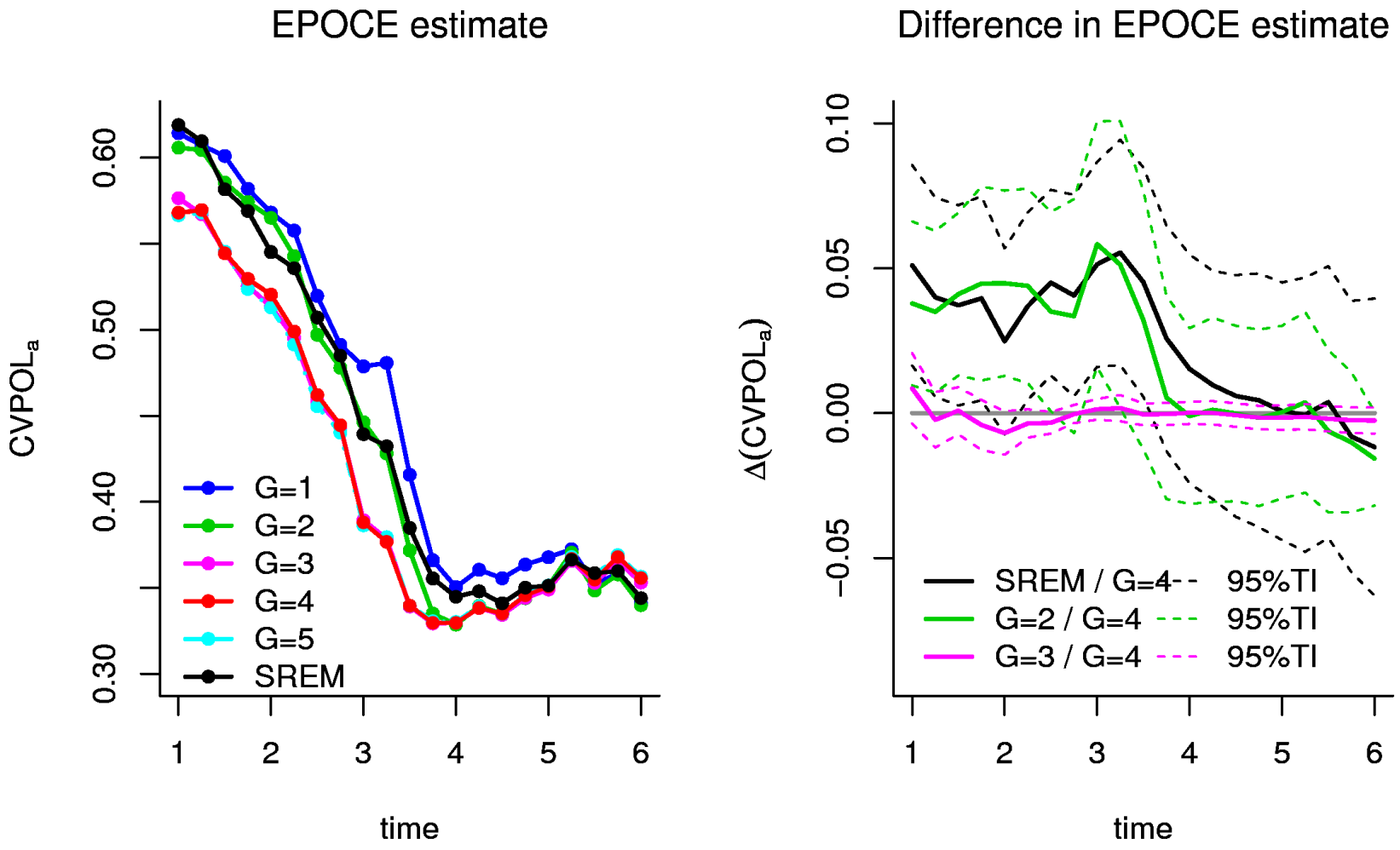
In the SREM

- PSA trajectory in 2 phases (short term drop / long term trend)
- Weibull baseline risk function
- covariates in the mixed model & in the survival model
- current PSA level and current PSA slope in the survival model

→ BIC=5445.1

Standard survival model : BIC=5598.7 (G=1)

EPOCE and difference in EPOCE on training data (N=459)



Change in treatment in the monitoring of patients after a cancer (Sène, 2013)

Dynamic predictions assume an absence of change in the follow-up

In practice, frequent initiation of second treatments :

hormonal therapy (HT) in prostate cancer

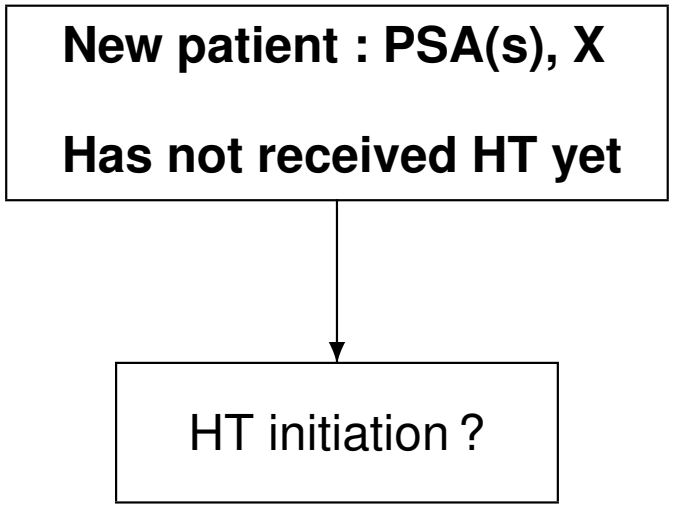
→ changes the dynamics of the biomarker

→ changes the risk of event

Solution :

- model the initiation of second treatment (ST)
- define differential dynamic predictions according to the initiation of ST

Differential individual dynamic predictions in patients free of HT

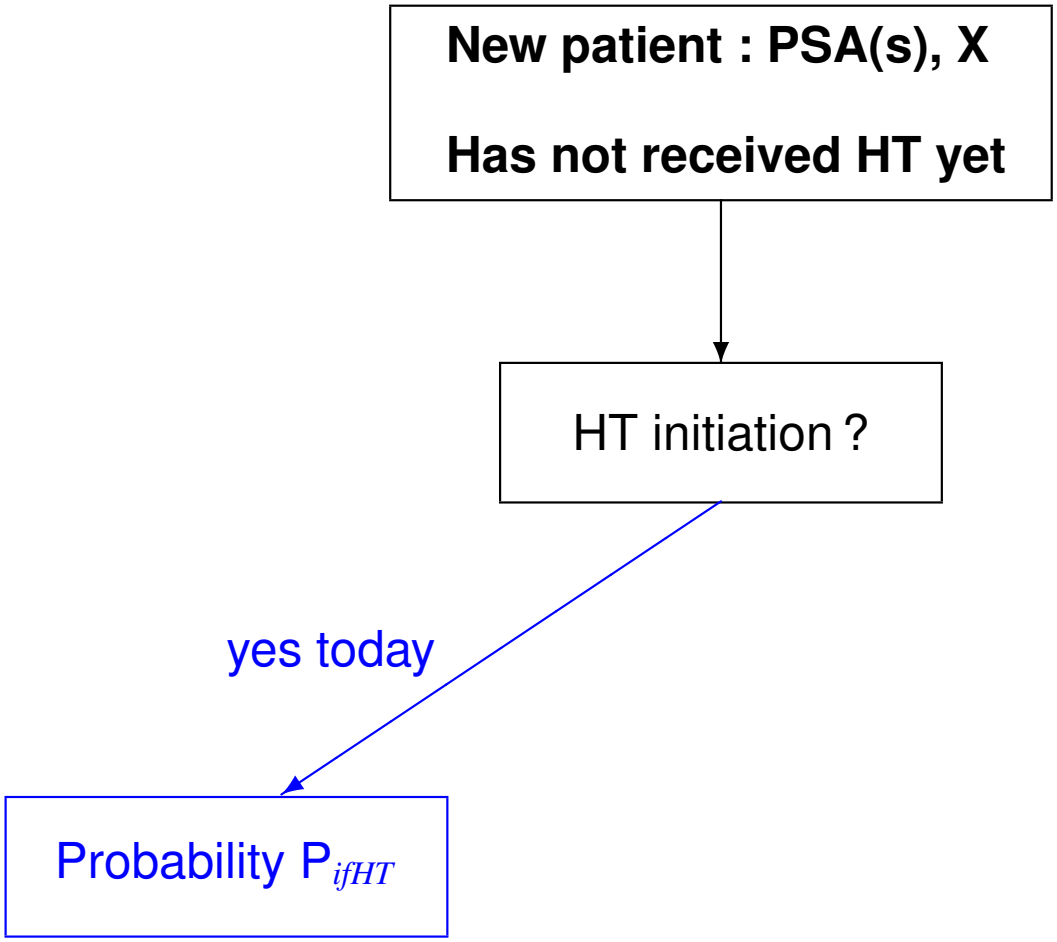


PSA(s)= collected PSA until today

X= available covariates at diagnosis

τ = time of initiation of hormonal therapy

Differential individual dynamic predictions in patients free of HT



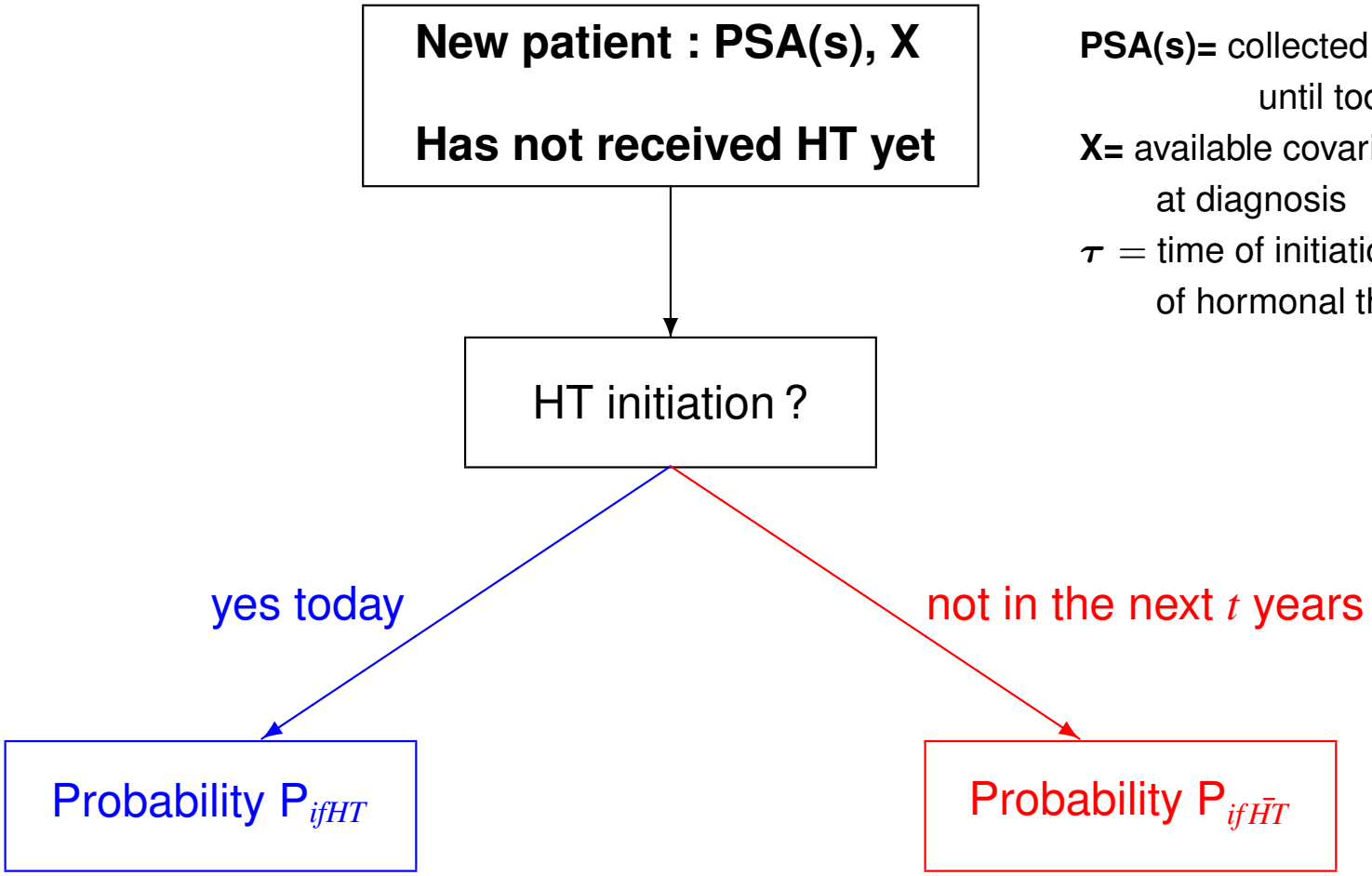
PSA(s)= collected PSA until today

X= available covariates at diagnosis

τ = time of initiation of hormonal therapy

$$P(T_i \leq s + t \mid T_i \geq s, \tau_i = s, Y_i^{(s)}, X_i)$$

Differential individual dynamic predictions in patients free of HT



PSA(s)= collected PSA until today

X= available covariates at diagnosis

τ = time of initiation of hormonal therapy

$$P(T_i \leq s+t \mid T_i \geq s, \tau_i = s, Y_i^{(s)}, X_i) \qquad P(T_i \leq s+t \mid T_i \geq s, \tau_i > \min(T_i, s+t), Y_i^{(s)}, X_i)$$

Illustration among 2386 men treated by radiation therapy (Sène, 2013)

SREM with different natures of dependency in the survival model :

$$\lambda(t \mid u_i) = \lambda_0(t) e^{X_{ei}(t)^T \delta + f(u_i, \beta, \tau_i)^T \eta}$$

- $f(u_i, \beta, \tau_i) =$ current level and slope before and after HT with change of risk in τ_i
- $f(u_i, \beta, \tau_i) =$ current level and slope before and after HT stratified before and after τ_i
- $f(u_i, \beta, \tau_i) =$ random-effects with change of risk in τ_i
- ....

Illustration among 2386 men treated by radiation therapy (Sène, 2013)

SREM with different natures of dependency in the survival model :

$$\lambda(t \mid u_i) = \lambda_0(t) e^{X_{ei}(t)^T \delta + f(u_i, \beta, \tau_i)^T \eta}$$

- $f(u_i, \beta, \tau_i) =$ current level and slope before and after HT with change of risk in τ_i
- $f(u_i, \beta, \tau_i) =$ current level and slope before and after HT stratified before and after τ_i
- $f(u_i, \beta, \tau_i) =$ random-effects with change of risk in τ_i
- ....

Differential evaluation of predictive performances :

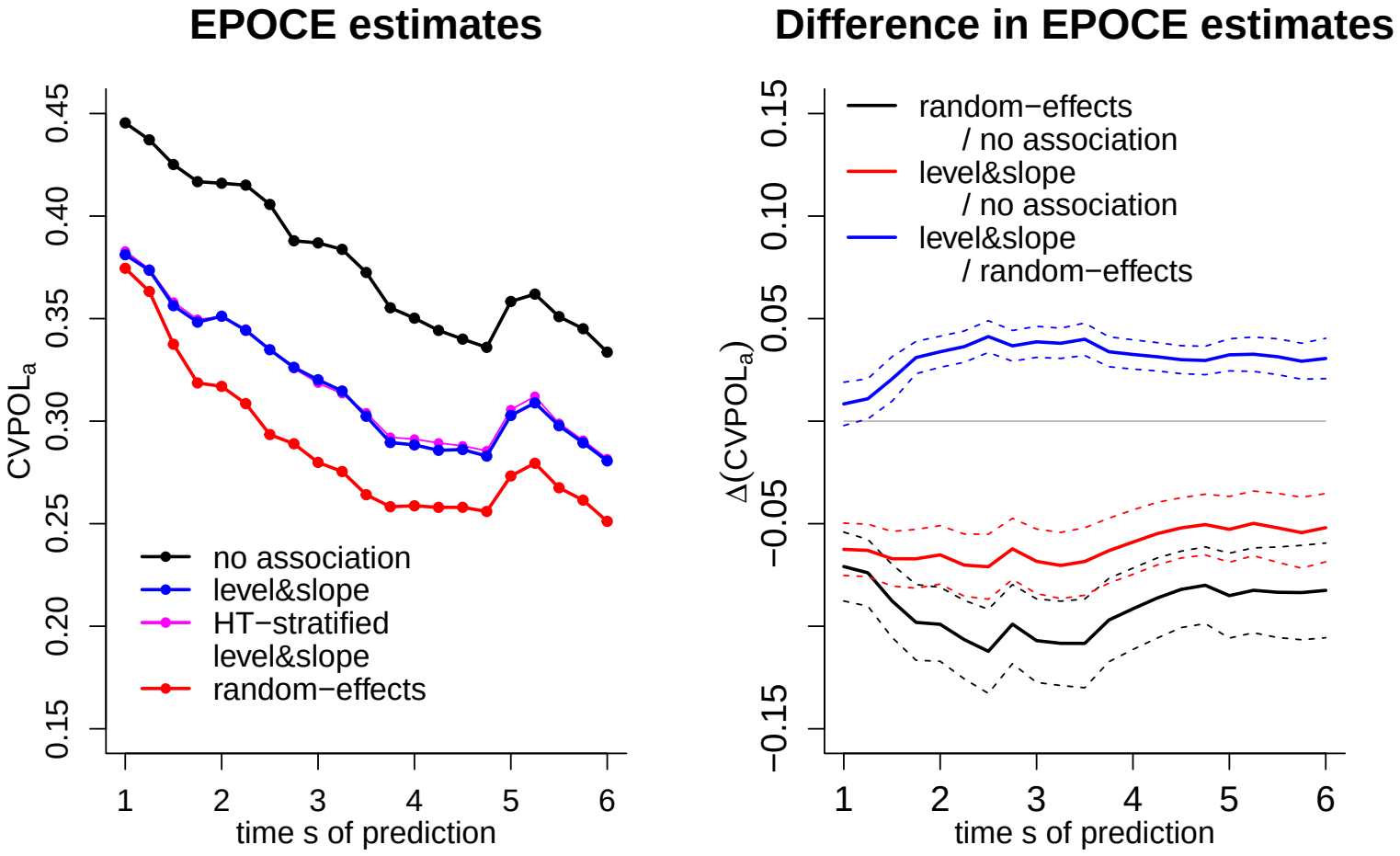
$\longrightarrow$ In absence of HT initiation in the window of prediction :

EPOCE curve computed at different times of prediction s with an horizon of 3 years

$\longrightarrow$ After immediate initiation of HT :

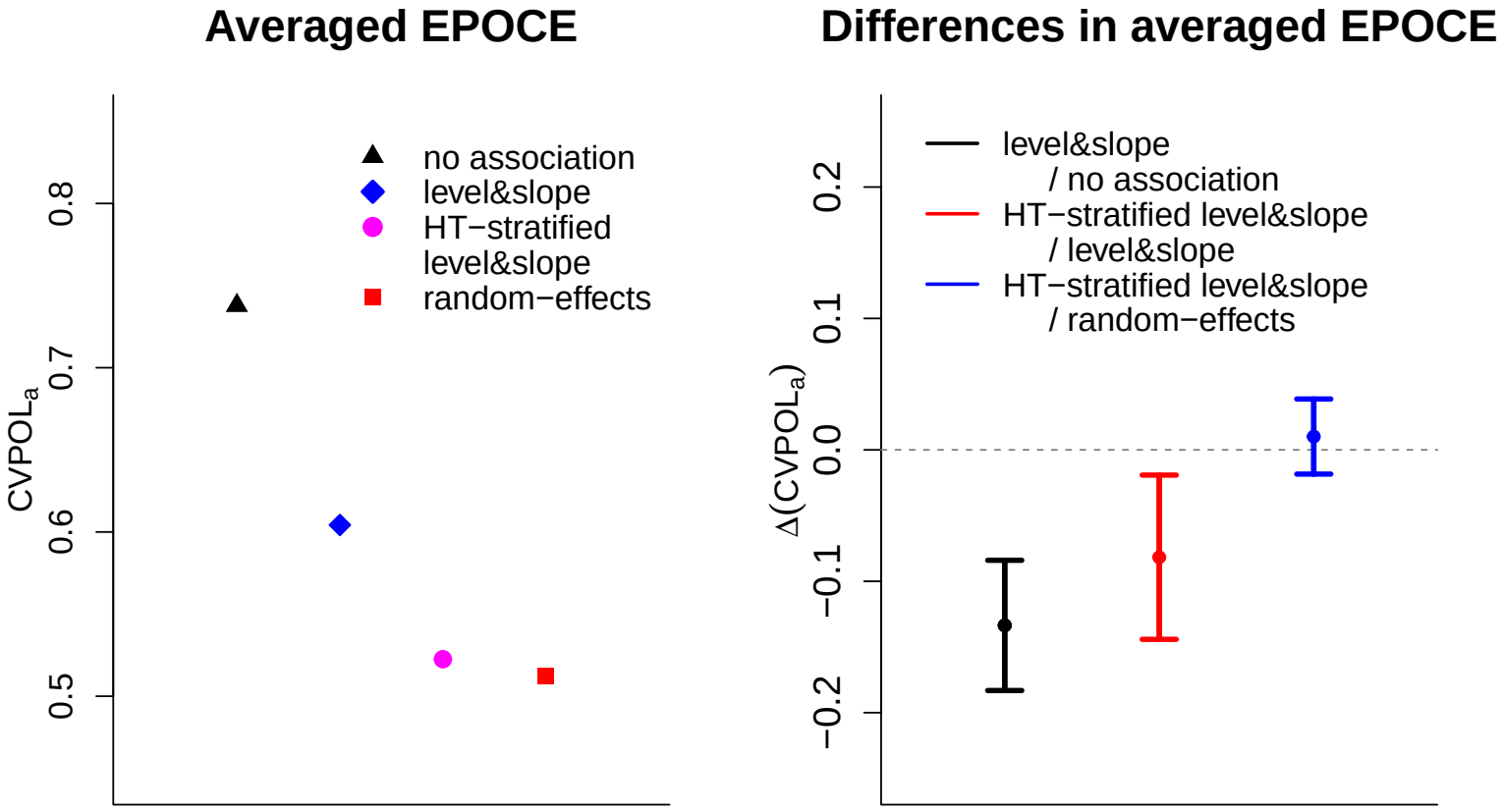
EPOCE computed in each τ_i and averaged over τ_i distribution

Predictive accuracy assessment : in absence of HT



Among men who did not initiate HT during the window of prediction (3y from time s)

Predictive accuracy assessment : after immediate HT initiation

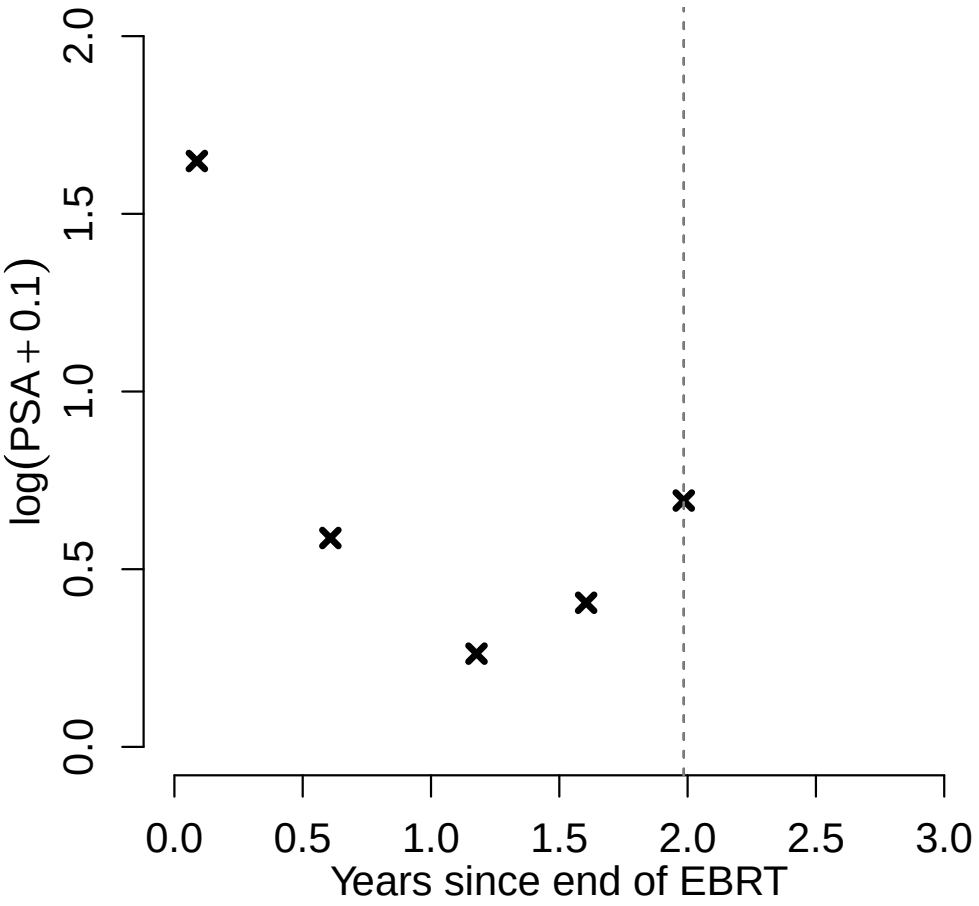


Cross-entropy **averaged** over the distribution of the times of HT

Example of differential dynamic predictions

For a man with a recurrence at 2.7 years

x PSA measures



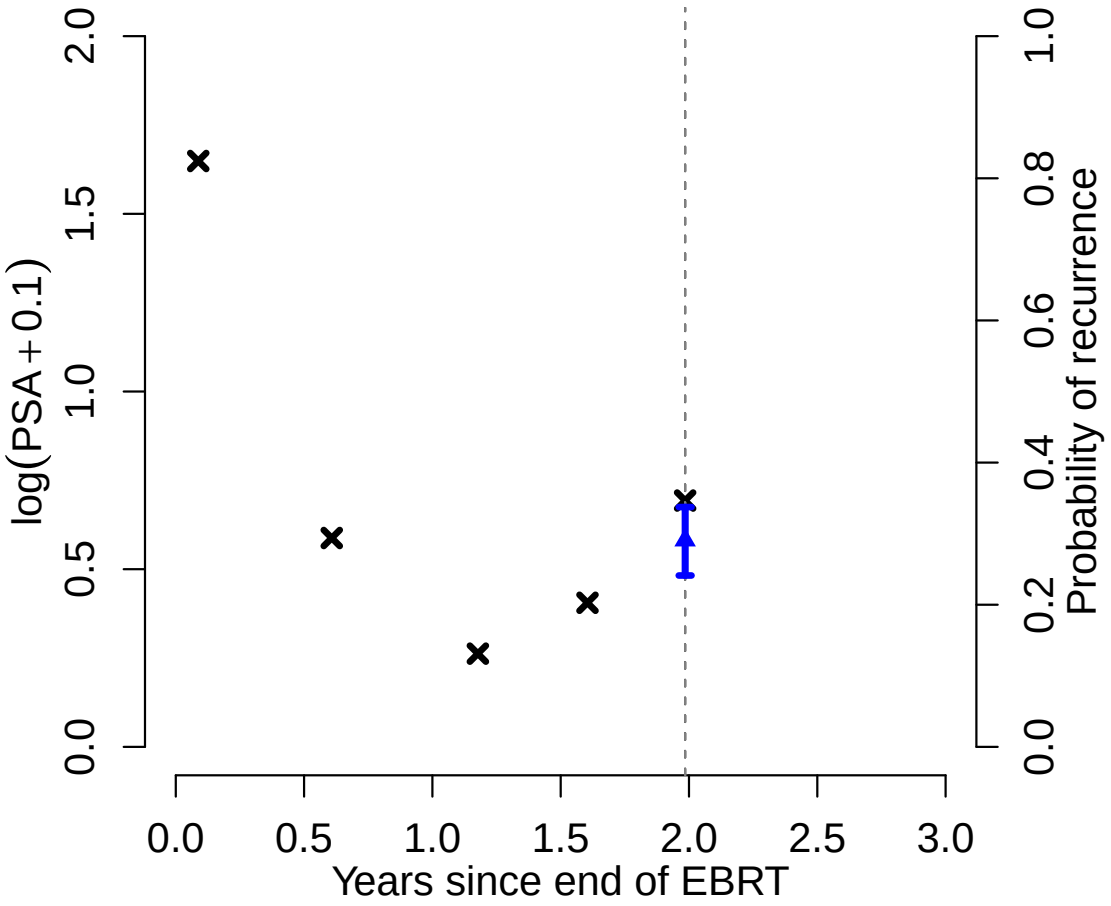
Example of differential dynamic predictions

For a man with a recurrence at 2.7 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **in absence of HT**



Example of differential dynamic predictions

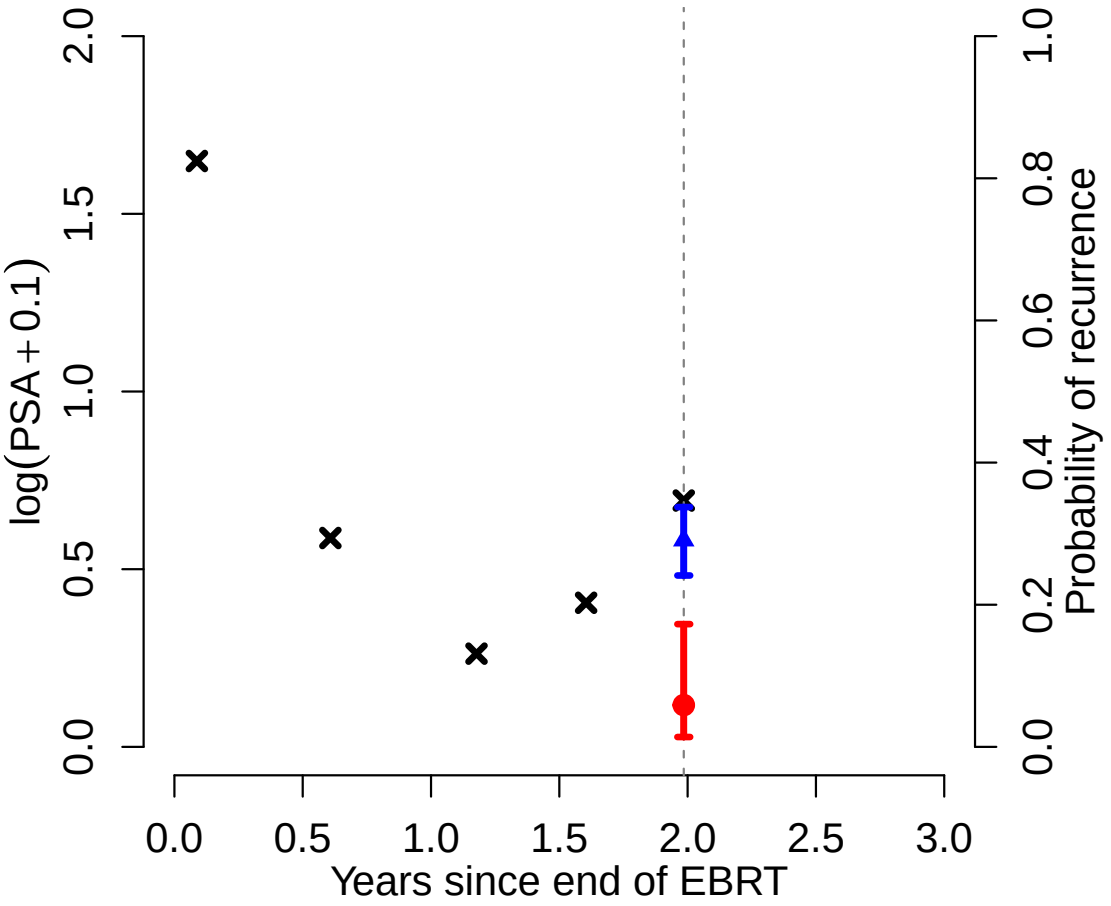
For a man with a recurrence at 2.7 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **in absence of HT**

● **if immediate initiation of HT**



Example of differential dynamic predictions

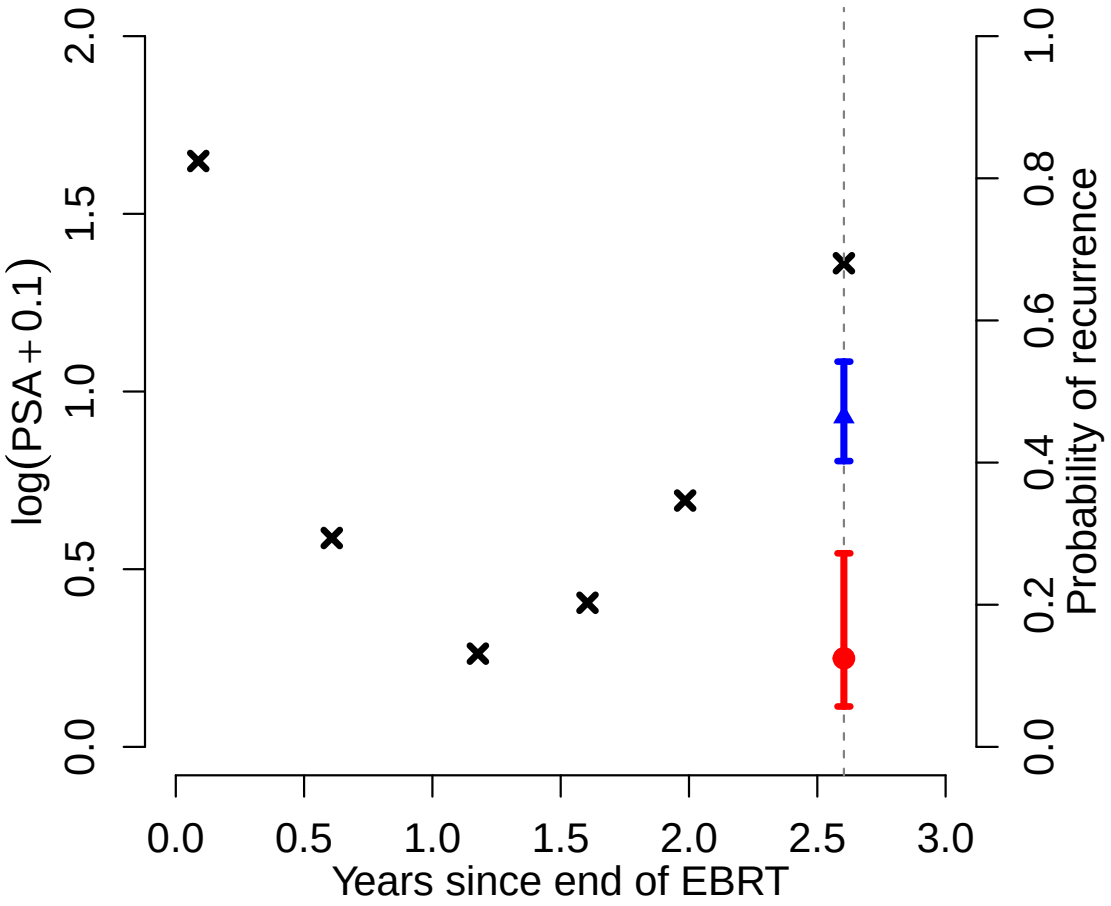
For a man with a recurrence at 2.7 years

x PSA measures

Predicted probability
of recurrence
in the next 3 years
with 95%CI :

▲ **in absence of HT**

● **if immediate initiation of HT**



Individual dynamic prediction in presence of competing events, multiple events

Multiple sources of cancer progression :

local recurrence, metastatic recurrence, death, ...

Medical decisions differ according to the type of progression

Solution :

Extension of joint models to competing events and/or multiple events

& Individual dynamic cause-specific predictions

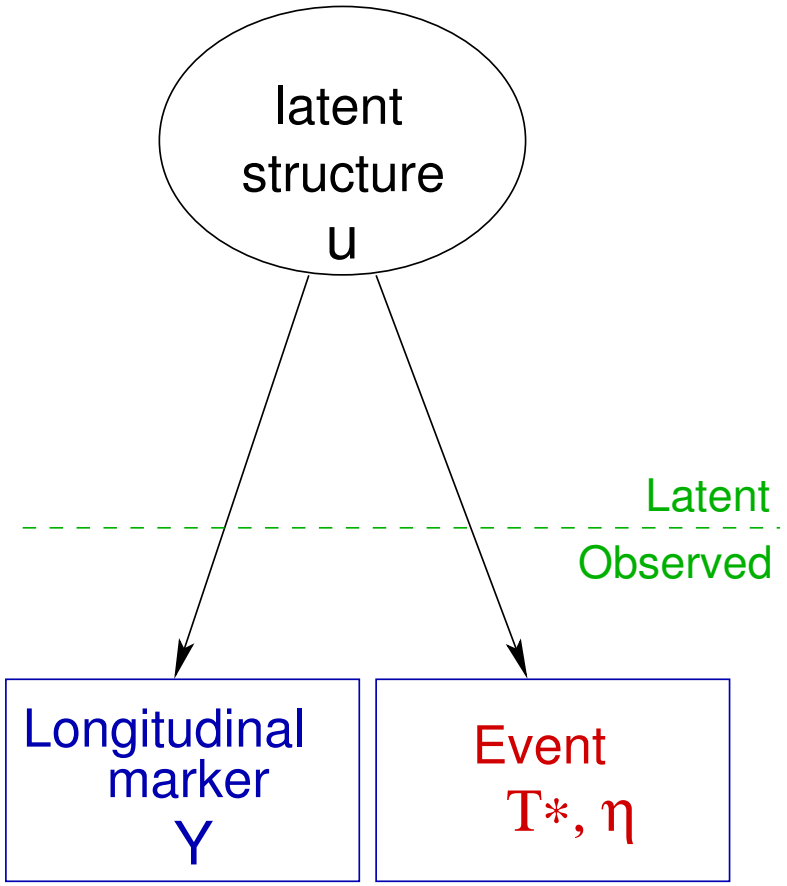
→ what is the probability of having (first) a local recurrence ?

Joint models in presence of competing risks

$\eta =$ cause of event

Predicted probability of
event of type k
given information until s :

$$P_i^{(k)}(s,t) = P(T_i \leq s+t, \eta = k \mid T_i \geq s, Y_i^{(s)}, X_i)$$



→ derived from JLCM or SREM

→ example and evaluation in Paul's talk

Conclusion

Dynamic predictions :

- → derived from any joint model (and landmark analysis)
- → potentially more accurate with systematic updates
- → calculator already online : http ://psacalc.sph.umich.edu
- → caution : should be based on a large amount of data

Predictive accuracy evaluation : essential in prognostic tools development

- → over-optimism to account for in complex models (external data, cross-validation)
- → several available measures (predictive accuracy, discrimination,etc)
- → caution : best goodness-of-fit does not mean best predictive accuracy

Joint models : flexible framework for predictions

- → multiple specifications (between & within JLCM & SREM)
- → extensions to different scenarios (caution : possible indication bias in observational studies)
- → multivariate events, multivariate biomarkers

Acknowledgements and references

Collaborators : Mbéry Sène, Hélène Jacqmin-Gadda, Jeremy MG Taylor, Daniel Commenges, Paul Blanche ...

Grant : French National Institute of Cancer INCa [grant PREDYC #2010-059]

More details on :

- **JLCM and prediction :** Proust-Lima & Taylor (2009). *Biostatistics*, **10**, 535-49
- **JLCM, SREM and prediction :** Proust-Lima, Sène, Taylor & Jacqmin-Gadda (2012). *Statistical Methods in Medical Research*, early view
- **SREM specification and prediction :** Sène, Bellera & Proust-Lima (2013) *Journal Société Française de Statistique*
- **Differential prediction :** Sène, Taylor, Jacqmin-Gadda & Proust-Lima (2013) *submitted*
- **EPOCE :** Commenges, Liquet & Proust-Lima (2012). *Biometrics*, **68(2)**, 380-7
- **Calculator online :** Taylor, Park, Ankerst et al. (2012). *Biometrics*, **69(1)**, 206-13