Tumor size evolution in randomized clinical trials : joint modeling approach and dynamic predictions

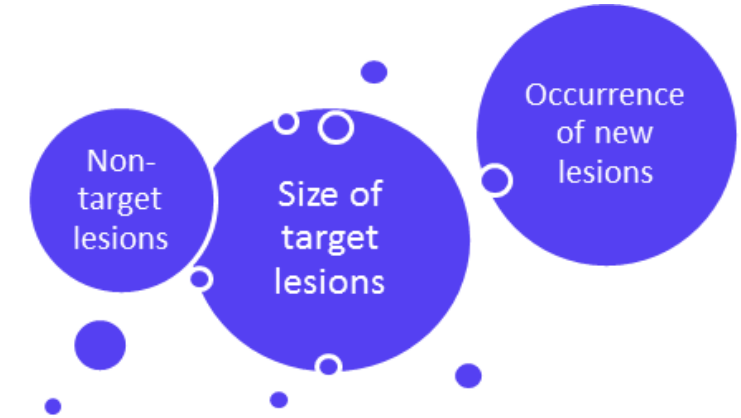
Agnieszka KROL [1] Loïc FERRER [1] Jean-Pierre PIGNON [2]
Cécile PROUST-LIMA [1] Michel DUCREUX [3] Olivier BOUCHÉ [4]
Stefan MICHIELS [2] Virginie RONDEAU [1]

[1] Biostatistics Team, INSERM U1219, Bordeaux,
[2] INSERM U1018 CESP, Gustave Roussy, U. Paris-Sud, Villejuif,
[3] Institut Gustave Roussy, Villejuif,
[4] Centre Hospitalier Universitaire, Reims

8 April 2016, GSO Workshop, Bordeaux

Instituts thématiques Inserm Institut national de la santé et de la recherche médicale
université de BORDEAUX
Réseau doctoral en santé publique

Tumor evaluations in clinical trials



- **WHO criteria** (1980s)
- **RECIST** (Response Evaluation Criteria in Solid Tumours) v1.0 (2000) and v1.1 (2009)

Tumor evaluation criteria

Comparable assessment of tumor change

- Within a trial (between patients)
- Between trials

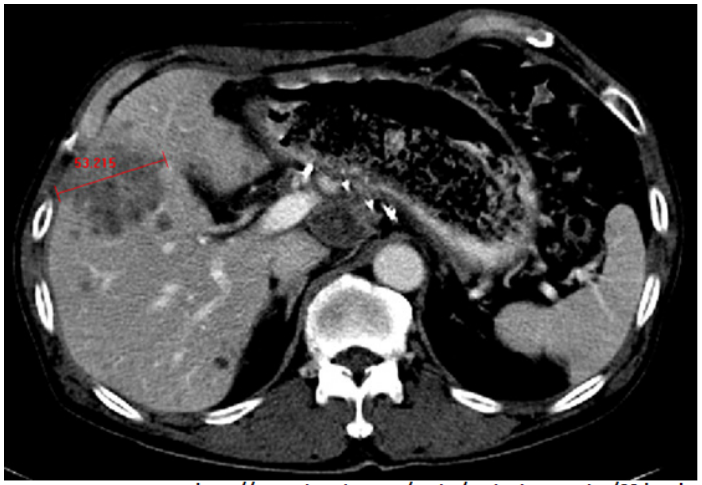
Categorical criteria - RECIST and WHO

- **WHO**
  - Bidimensional size, target lesions determined before treatment
  - Progression : >25% increase of one or more target lesions
  - Appearance of new lesions $\rightarrow$ global progression
- **RECIST** (v1.1)
  - Unidimensional size, max 2 lesions per organ and up to 5 total
  - Progression : >20% increase over smallest sum observed ($>$ 5 mm absolute increase)
  - Appearance of new lesions $\rightarrow$ global progression

4 categories (Complete Response, Partial Response, Progressive Disease, Stable Disease)

$\Rightarrow$ dichotomization : response or no response / progression or no progression

Measurement of lesions


http://www.irrecist.com/recist/recist-in-practice/02.html

- The longest diameters measured in the plane in which the images were acquired
- Measure the longest diameter of a lesion
- Measure the longest perpendicular diameter to it and the burden is their product (WHO criteria)
- Total individual tumor burden is the sum (of the longest diameters - RECIST, of the products - WHO)
- Baseline : no more than 4 weeks before treatment, Follow-up : every 6-8 weeks

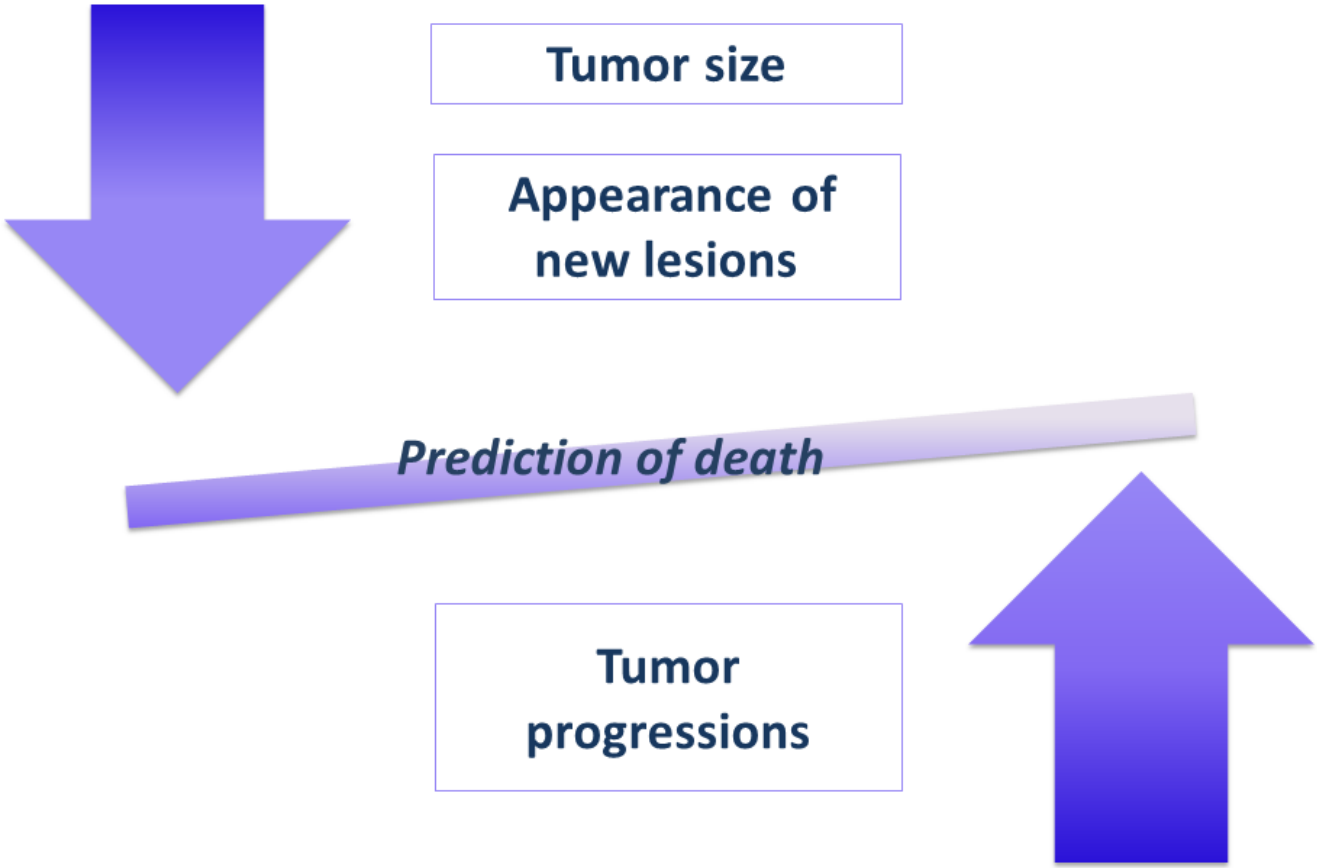
Measurability of lesions

- Measurable tumor lesions - at least one diameter with a minimum size of :
  - 10 mm by CT scan
  - 10 mm caliper measurement by clinical exam
  - 20 mm by chest X-ray

  Lymph nodes : $\geq$15 mm in *short* axis when assessed by CT scan
- Non-measurable tumor lesions
  - small lesions (longest diameter <10 mm)
  - truly non-measurable lesions, eg. leptomeningeal disease, ascites, inflammatory breast disease

Objective

Does the continuous tumor size and/or appearance of new lesions enable better prediction of the OS than times of progression ?

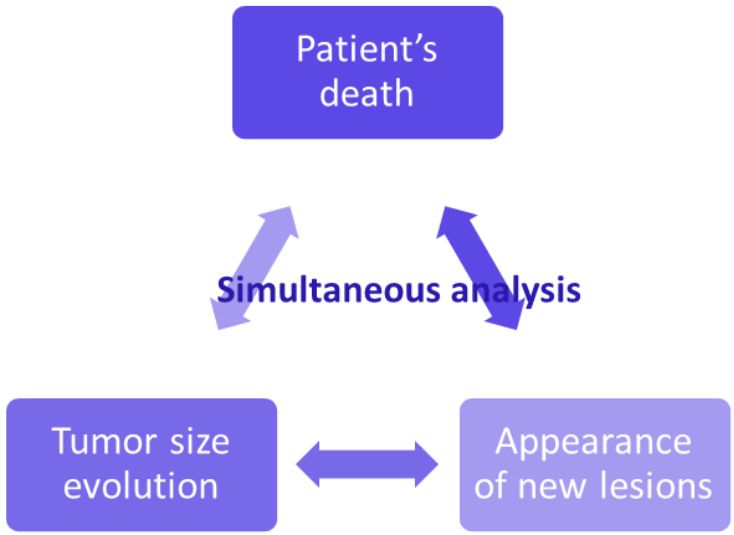


Reference : Król et al. *Biometrics*, 2016.

Observed data

For individual i ($i = 1, \ldots, N$) we observe :

- **Longitudinal biomarker** : $Y_i(t_{ik})$
- **Recurrences** : $T_{ij} = \min(T_{ij}^*, C_i, T_i^*)$ and $\delta_{ij} = \mathbb{1}_{\{T_{ij}^* = T_{ij}\}}$
- **Terminal event** : $T_i = \min(C_i, T_i^*)$ and $\delta_i = \mathbb{1}_{\{T_i^* = T_i\}}$



Joint model for longitudinal data, recurrent events and a terminal event

System of linear mixed-effects model and two hazard functions :

$$\begin{cases} Y_i(t_{ik}) = m_i(t_{ik}) + \epsilon_i(t_{ik}) = \boldsymbol{X}_{i,l}(t_{ik})^\top \beta_l + \boldsymbol{Z}_i(t_{ik})^\top \boldsymbol{b}_i + \epsilon_i(t_{ik}) & \text{(Biomarker)} \\ r_{ij}(t|v_i, \boldsymbol{b}_i) = r_0(t) \exp\left(v_i + \boldsymbol{X}_{ij,r}^\top \beta_r + g(\boldsymbol{b}_i, t)^\top \eta_r\right) & \text{(Recurrences)} \\ \lambda_i(t|v_i, \boldsymbol{b}_i) = \lambda_0(t) \exp\left(\alpha v_i + \boldsymbol{X}_{i,t}^\top \beta_t + h(\boldsymbol{b}_i, t)^\top \eta_t\right) & \text{(Death)} \end{cases}$$

- $u_i = (\boldsymbol{b}_i^T, v_i)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$ with $\mathbf{B} = \begin{pmatrix} \mathbf{B_1} & \mathbf{0} \\ \mathbf{0} & \sigma_v^2 \end{pmatrix}$
- measurement errors *iid*, $\epsilon_i(t_{ik}) \sim \mathcal{N}(0, \sigma_\epsilon^2)$
- $g(\boldsymbol{b}_i, t)$ and $h(\boldsymbol{b}_i, t)$ - link functions
- $r_0(t)$, $\lambda_0(t)$ - baseline hazard functions

Estimation

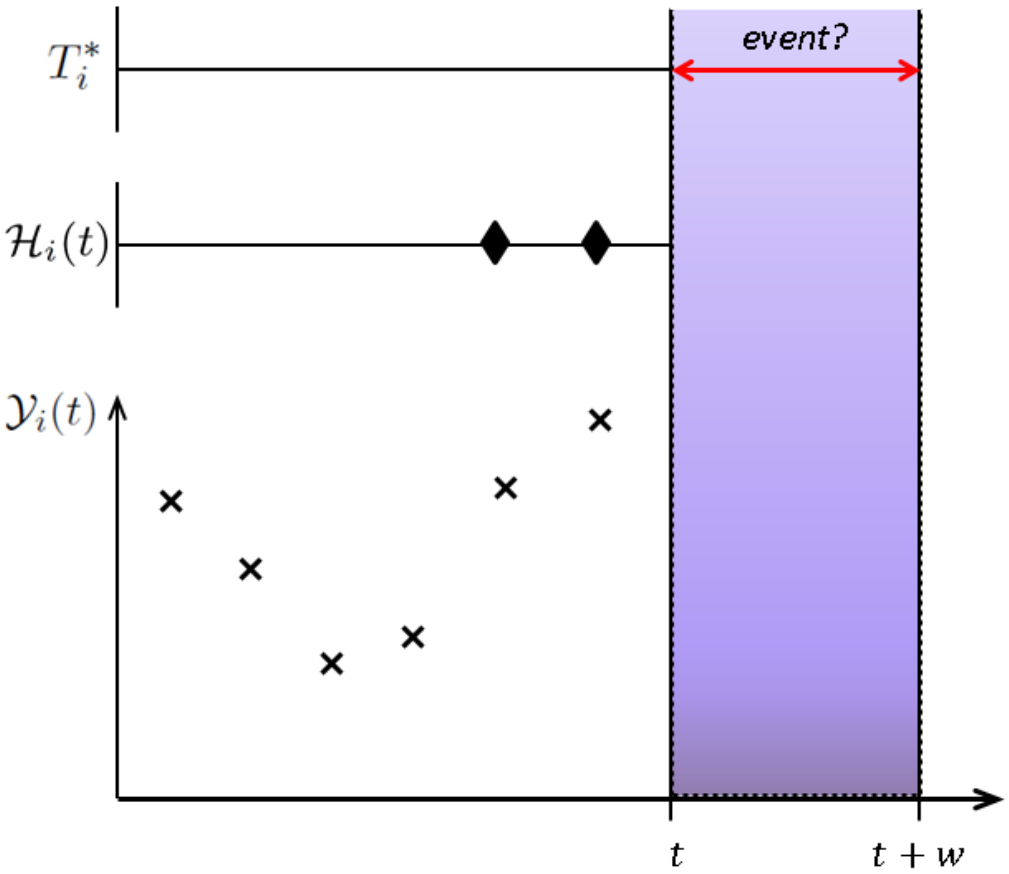
- Joint marginal likelihood

$$L_i(\boldsymbol{\theta}) = \int_{\boldsymbol{u}_i} \prod_{k=1}^{n_i} f_{Y|\boldsymbol{u}_i}(Y_i(t_{ik})|\boldsymbol{u}_i;\boldsymbol{\theta}) \prod_{j=1}^{r_i} f_{T^r|\boldsymbol{u}_i}(T_{ij},\delta_{ij}|\boldsymbol{u}_i;\boldsymbol{\theta}) \cdot f_{T^t|\boldsymbol{u}_i}(T_i,\delta_i|\boldsymbol{u}_i;\boldsymbol{\theta}) f_{\boldsymbol{u}_i}(\boldsymbol{u}_i;\boldsymbol{\theta}) \mathrm{d}\boldsymbol{u}_i$$

  - n_i - number of biomarker measurements of individual i,
    r_i - number of recurrent events of individual i
  - Parameters to estimate $\boldsymbol{\theta} = (\boldsymbol{\beta}_l^\top, \boldsymbol{\beta}_r^\top, \boldsymbol{\beta}_t^\top, \boldsymbol{\eta}_r^\top, \boldsymbol{\eta}_t^\top, \alpha, r_0(\cdot), \lambda_0(\cdot), \mathbf{B}, \sigma_\epsilon)^\top$

- Penalized maximum likelihood estimation using Marquardt algorithm
- Baseline hazard functions approximation using splines : smooth estimation
- Integrals approximated using Gauss-Hermite quadrature

Dynamic predictions



- $\mathcal{H}_i(t)$ - history of recurrences of individual i until t
  $\mathcal{Y}_i(t)$ - history of the biomarker of individual i until t
- Predicted probability of the terminal event T_i^* in a horizon $[t, t+w]$

$$\mathbb{P}(T_i^* \leq t+w | T_i^* > t, \mathcal{F}_i(t), \boldsymbol{X}_i; \boldsymbol{\theta})$$

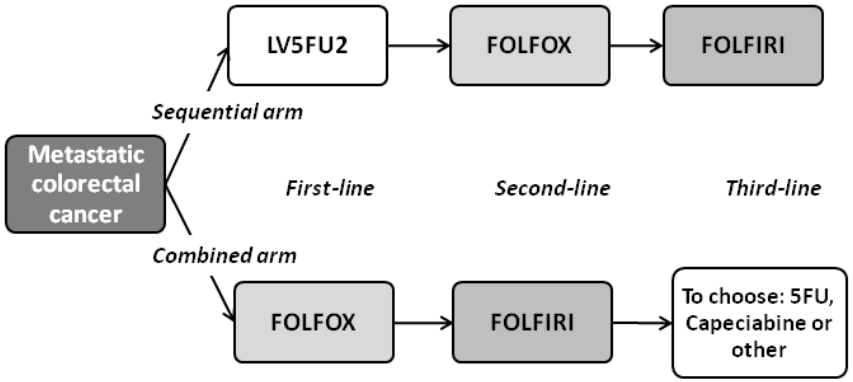
$\mathcal{F}_i(t) = \mathcal{H}_i(t)$,
$\mathcal{F}_i(t) = \mathcal{Y}_i(t)$
or $\mathcal{F}_i(t) = \{\mathcal{H}_i(t), \mathcal{Y}_i(t)\}$

Measures of predictive abilities

- EPOCE (Expected Prognostic Observed Cross-Entropy) *Commenges et al.*, 2012
  - Evaluation of conditional density of the event given the individual history
  - Internal validation : approximate cross-validated estimator CVPOL_a
- Brier score
  - The inverse probability of censoring weighted error estimator (data-based Brier score) *Gerds and Schumacher*, 2006
  - Comparison of predictions and actual observed events
  - Internal validation : *k*-fold cross-validation

Clinical trial FFCD 2000-05

- Follow-up :
  - Phase III randomized multi-center clinical trial (53 centers in France), 407 patients



  - Tumor evaluation every 8 weeks, max 4 target lesions in 2 dimensions
  - Change of line : progression (WHO criteria), unacceptable toxicity, decision of investigator

Ducreux et al., *The Lancet Oncology,* 2011

Clinical trial FFCD 2000-05

- Objectives :
  - Which of **longitudinal biomarker**, **times of appearance of new lesions** or **times of progression** provide the most accurate **prediction** of the **overall survival** ?
  - To identify the prognostic factors on the outcomes of interest
  - To evaluate the treatment effect

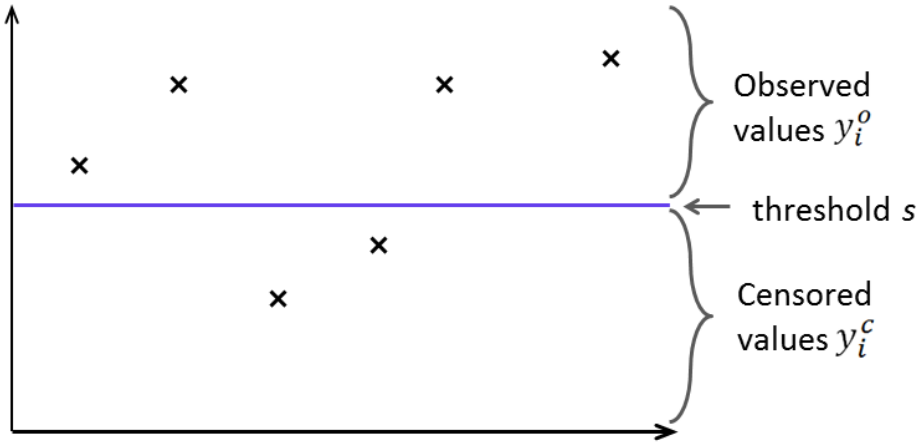
Data

- Biomarker definition : sum of the longest diameters

$$SLD_{ij} = \sum_{k=1}^{n_{ij}} d_{ijk}, \quad j = 0, 1, \ldots, n_i, \quad i = 1, \ldots, 407$$

$n_i \in \{0, 1, ..., 17\}$ - number of visits of individual i, $n_{ij} \in \{1, 2, 3, 4\}$ - number of target lesions measured during visit j, d_{ijk} - max diameter of lesion k measured during visit j of individual i

- Left-censoring



Data : FFCD 2000-05

N=402 patients analyzed. Observed :

- 6.18 tumor size measurements per patient
- 1.05 appearance of new lesions per patient
- 1.82 progression per patient
- 321 deaths
- Overall survival : 16.3 months in the combination (C) arm and 16 months in the sequential (S) arm

Results of the trivariate model

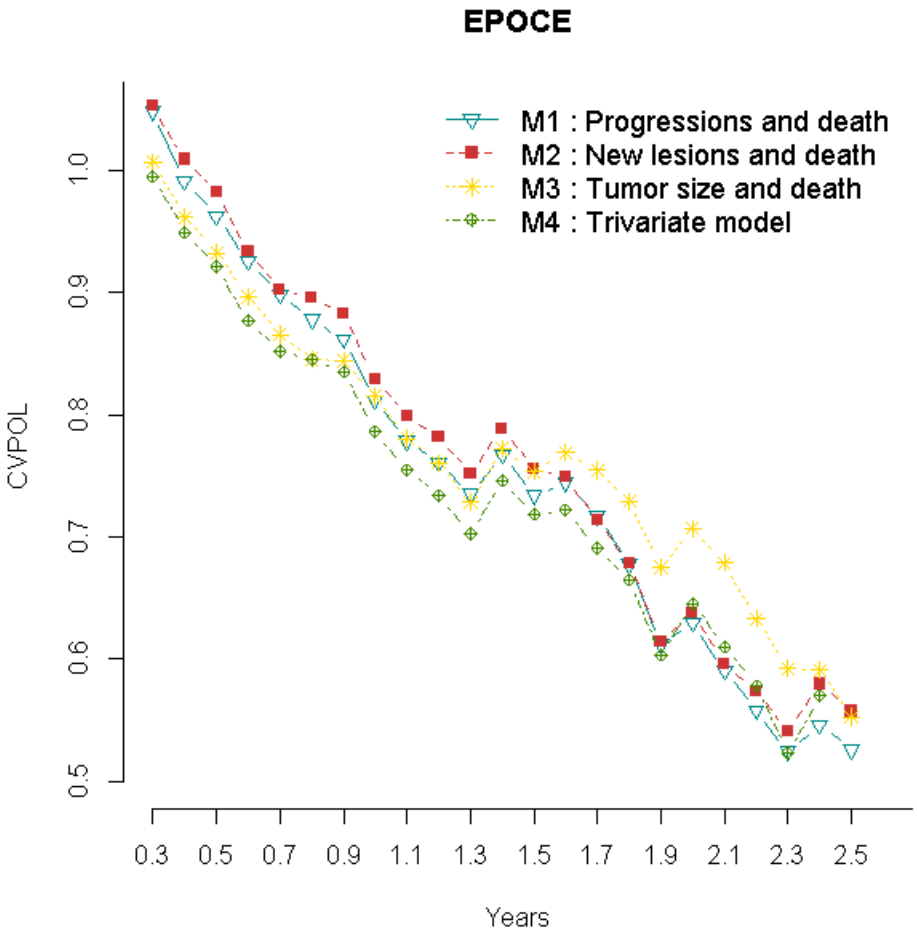
	Biomarker : SLD		New lesions	Death
Covariate	Est. (SE)	p-value	HR (95% CI)	HR (95% CI)
Intercept	2.90 (0.29)	$<$0.001	-	-
Time	$-$0.35 (0.13)	0.006	-	-
Treatement (C/S)	$-$0.20 (0.14)	0.16	0.96 (0.75-1.21)	1.02 (0.64-1.61)
Treatement (C/S)$\times$Time	$-$0.42 (0.15)	0.007	-	-
Age (60-69/$<$60 years)	0.23 (0.18)	0.20	0.75 (0.56-1.02)	1.04 (0.57-1.87)
Age ($\geq$70/$<$60 years)	0.02 (0.16)	0.91	0.82 (0.61-1.09)	1.40 (0.79-2.49)
Sex (Women/Men)	0.27 (0.14)	0.06	0.86 (0.67-1.10)	1.02 (0.63-1.65)
Baseline WHO PS (1/0)	$-$0.14 (0.15)	0.34	1.16 (0.89-1.51)	1.51 (0.85-2.68)
Baseline WHO PS (2/0)	0.45 (0.21)	0.035	2.15 (1.44-3.21)	10.22 (3.68-28.40)
. . . . .				

- Significant decreasing value of SLD with time (-0.35), and decreasing with time more pronounced for the combination arm (-0.40)
- Strong effect of WHO performance status 2 on the risk of death, new lesions and on tumor size
- No significant associations with gender and age
- Significant associations between the processes via the shared random effects (except of the link between the biomarker and recurrent events)

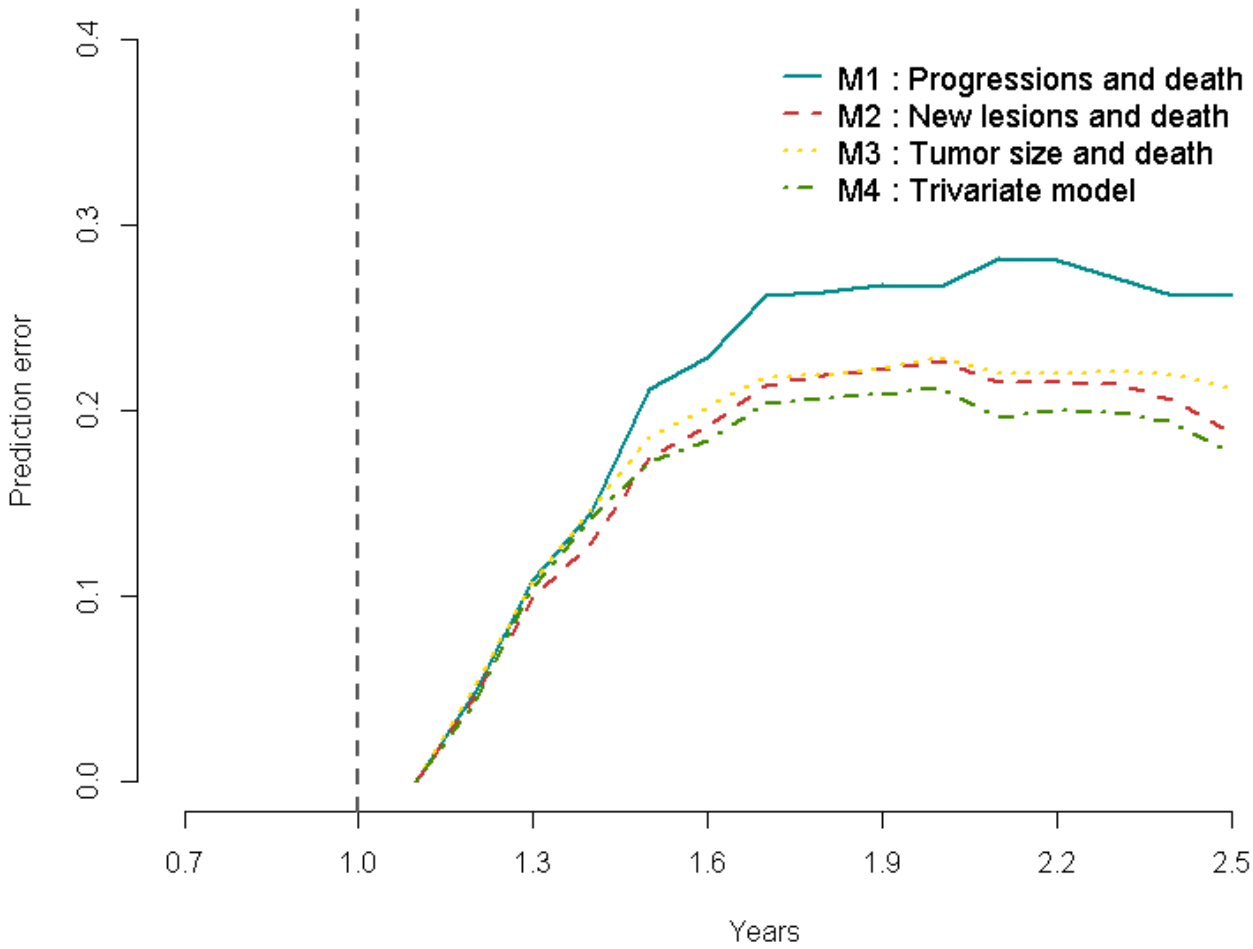
Comparison with the alternative models - predictive ability

- Comparison of the models in terms of the **predictive ability** of the overall survival
  - Joint modelling of times of progression and time of death (M1)
  - Joint modelling of times of appearance of new lesions and time of death (M2)
  - Joint modelling of tumor size (SLD) and time of death (M3)
  - Joint modelling of tumor size (SLD), times of appearance of new lesions and time of death (M4)
- Measures of predictvie ability using internal validation
  - **Brier score** (10-fold cross-validation)
  - **EPOCE** (CVPOL_a - approximated cross-validation)

Results - EPOCE



Results - Brier score

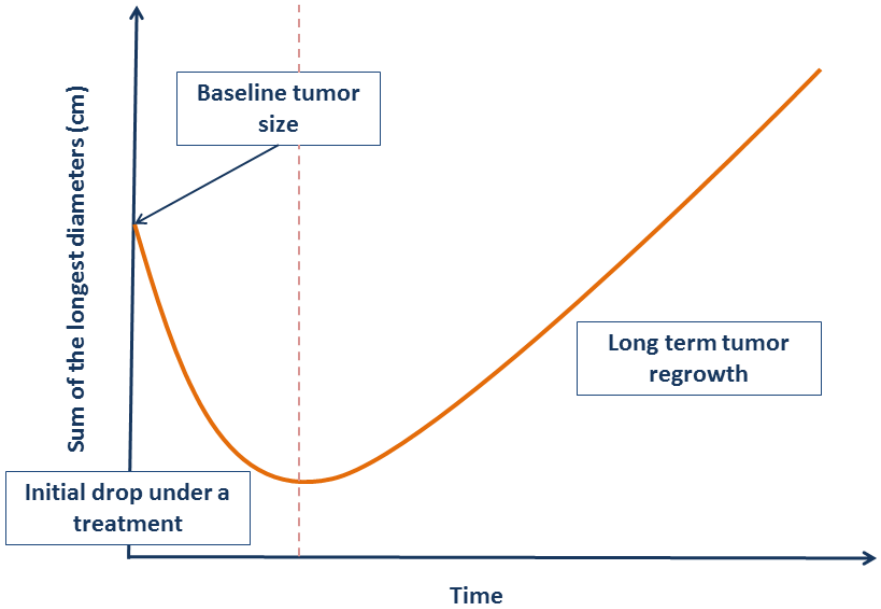


Conclusion

- Advantages of using joint models for simultaneous analysis of prognostic factors
- Comparison of joint models of different types in terms of predictive accuracy
- Proposition of a new trivariate joint model
- FFCD 2000-05 : Improvement of predictive abilities using tumor size and appearance of new lesions
- Implementation of the proposed model into the **R** package `frailtypack` *Rondeau et al.*, 2012

Perspectives

- Incorporation of information on progression of non-target disease
- Application to other clinical trials, in particular to a meta-analysis
- More flexible modeling of the biomarker
  - Parametric approach : two slopes of time
  - Approximation by B-Splines
  - Tumor dynamics modeled using ordinary differential equations *Claret et al.*, 2009



References

1 Claret, L. et al. (2009). Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *Journal of Clinical Oncology* **27**(25) 4103-08.

2 Commenges, D. et al. (2012). Choice of prognostic estimators in Kullback-Leibler risks. *Biometrics*, **68**, 380-7.

3 Ducreux, M. et al. (2011). Sequential versus combination chemotherapy for the treatment of advanced colorectal cancer (FFCD 2000-05) : an open-label, randomised, phase 3 trial. *The Lancet Oncology* **12**, 1032-44.

4 Gerds, T. and Schumacher, M. (2006). Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. *Biometrical Journal* 48, 1029-40.

5 Król, A. et al. (2016). Joint model for left-censored longitudinal data, recurrent events and terminal event : predictive abilities of tumor burden for cancer evolution with application to the FFCD 2000-05 trial. *Biometrics*.

6 Rondeau, V. et al. (2012). frailtypack : an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software* **47**(4) 1-28.

7 Rondeau, V. et al. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation. *Biostatistics*, **8**, 708-21.

8 Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 67, 819-29.

9 Wolfsohn, M. S. and Tsiatsis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330-9.