



HAL
open science

Quantitative modeling of metastasis: cancer at the organism scale

Sébastien Benzekry

► **To cite this version:**

Sébastien Benzekry. Quantitative modeling of metastasis: cancer at the organism scale. Doctoral. United States. 2021. hal-03261886

HAL Id: hal-03261886

<https://inria.hal.science/hal-03261886v1>

Submitted on 16 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantitative modeling of metastasis: cancer at the organism scale

S. Benzekry
Head of the Inria-Inserm team COMPO
Marseille, France

SMB meeting, 2021 June 14th

COMPO: COMPUtational pharmacology and clinical Oncology



MATHEMATICAL MODELING

Mechanistic modeling

D. Barbolosi
S. Benzekry

Pharmacometrics

D. Barbolosi, S. Benzekry
F. Gattacceca

Statistics Data science

S. Benzekry



PHARMACOLOGY

Experimental

J. Ciccolini
R. Fanciullino
A. Rodallec

Clinical

J. Ciccolini
R. Fanciullino
B. Lacarelle



MEDICINE

Pr L. Greillier Dr X. Muracciole
Pr S. Salas

Inria

Aix-Marseille
université

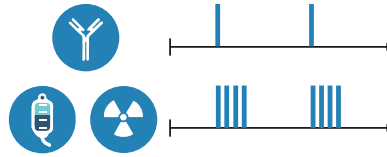


CRCM
Centre de Recherche
en Cancérologie de Marseille

Instituts
thématiques **Inserm**
Institut national
de la santé et de la recherche médicale

INSTITUT PAOLI-CALMETTES
unicancer Marseille

Clinical problem



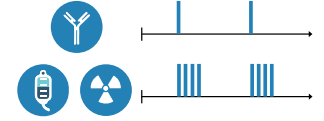
Mathematical models

$$\begin{cases} \partial_t \rho(t, v) + \partial_v (g(v) \rho(t, v)) = 0 \\ g(V_0) \rho(t, V_0) = d(V_p(t)) \left(+ \int_{V_0}^{+\infty} d(v) \rho(t, v) dv \right) \\ \rho(0, v) = \rho^0 \end{cases}$$

Data

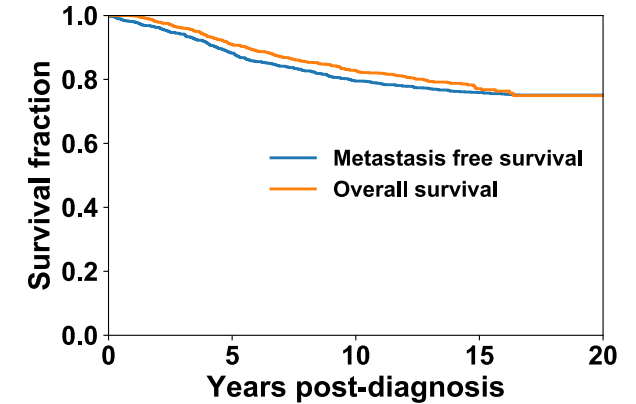


Clinical problem



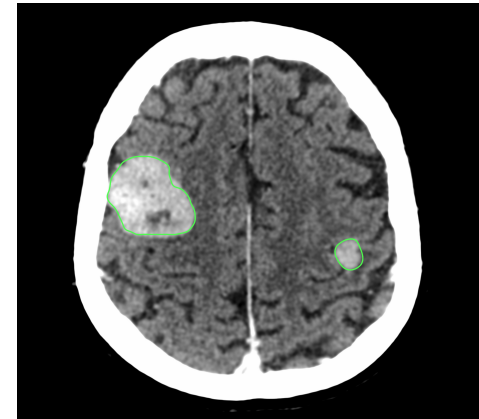
Early-stage breast cancer

- 94% of cases are local or regional at diagnosis but 30% will relapse
- Estimation of the metastatic risk is key to individualize adjuvant therapy
- Reduce the number of chemo cycles for patients with low risk

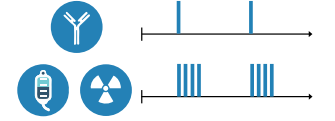


Brain metastases in non-small cell lung cancer (NSCLC)

- Decide whether to use whole brain radiation therapy or just (stereotactic) surgery

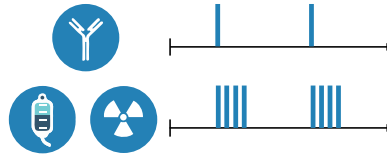


Objectives



- Use a **mechanistic model** to predict metastasis
- Combine with machine learning algorithm to **select features**
- Benchmark predictive power to standard survival methods and **machine learning** algorithms

Clinical problem



Mathematical models

$$\begin{cases} \partial_t \rho(t, v) + \partial_v (g(v) \rho(t, v)) = 0 \\ g(V_0) \rho(t, V_0) = d(V_p(t)) \left(+ \int_{V_0}^{+\infty} d(v) \rho(t, v) dv \right) \\ \rho(0, v) = \rho^0 \end{cases}$$

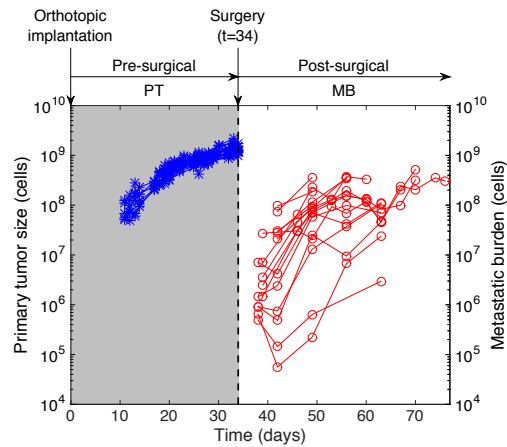
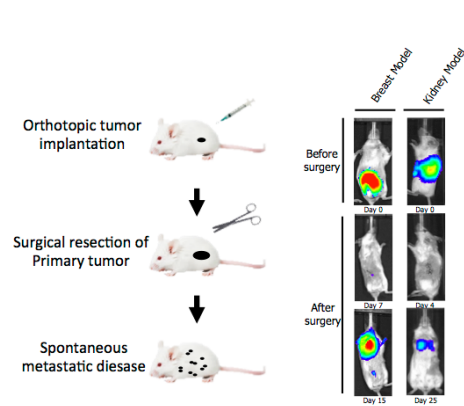
Data





Data

Experimental



n > 400 animals with different treatments and schedules

SMARTc wet lab
collaboration with J. Ebos, Roswell Park Cancer Institute, Buffalo, NY

Clinical

- Databases of metastatic relapse in breast cancer patients with no adjuvant therapy (n=642, p=21 and n=167, p=9)

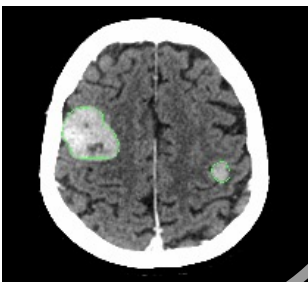
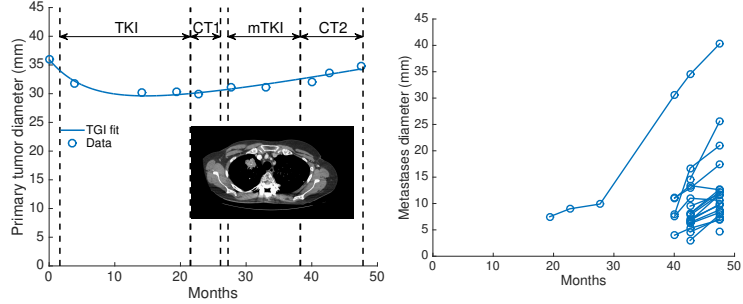
p

metastasis_status	ER	PR	Ki67	HER2	HER2_Intensity	CK56	EGFR	YIM	ALB1
Pre-surgery	20	0	0	0	0	0	0	0	0
Post-surgery	40	95	8	0	0	0	0	0	0
Active glioma	87	10	26	0	0	0	0	0	0
Post-surgery	100	100	8	0	0	0	0	0	0
Pre-surgery	0	0	16	82	+++	0	0	0	0
Post-surgery	100	95	15	0	0	0	0	0	0
Active glioma	56	100	17	0	0	0	0	0	0
Post-surgery	57	85	23	100	+++	0	0	0	0
Pre-surgery	80	5	20	0	0	0	0	0	0
Post-surgery	0	0	15	100	+++	0	5	0	0
Pre-surgery	100	80	10	0	0	0	0	0	0
Post-surgery	30	0	5	0	0	0	0	0	0
Pre-surgery	0	0	15	40	+++	0	0	0	0
Post-surgery	0	80	8	0	0	0	0	0	0
Pre-surgery	0	0	23	0	0	0	30	0	0
Post-surgery	0	0	56	0	0	80	60	100	0
Active glioma	50	95	2	1	+	0	0	0	1
Pre-surgery	0	47	5	0	0	0	0	80	0
Post-surgery	65	0	10	0	0	0	0	60	0
Pre-surgery	100	50	11	0	0	0	0	0	0
Post-surgery	20	100	0	0	0	0	0	0	0
Pre-surgery	100	0	5	0	0	0	0	0	0
Post-surgery	0	0	6	0	0	0	0	0	0
Active glioma	100	5	5	0	0	0	0	0	0
Post-surgery	0	0	6	0	0	0	0	0	0
Pre-surgery	80	100	5	0	0	0	0	0	0
Post-surgery	100	85	25	0	0	0	0	0	0
Pre-surgery	10	45	11	13	+	0	0	0	0
Post-surgery	66	1	2	40	+	0	0	0	0

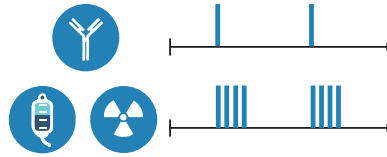
n

data_metastatic_relapse	data_death_or_loss
1999-02-01 00:00:00	1999-01-06 00:00:00
1999-02-01 00:00:00	1999-01-21 00:00:00
1999-02-01 00:00:00	2004-06-11 00:00:00
1999-02-01 00:00:00	2008-09-21 00:00:00
1999-02-01 00:00:00	2006-04-01 00:00:00
1999-12-31 00:00:00	2006-11-21 00:00:00
1999-12-31 00:00:00	2009-11-01 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-12-01 00:00:00
2008-09-15 00:00:00	1990-10-20 00:00:00
2008-09-15 00:00:00	2008-11-04 00:00:00
2008-09-15 00:00:00	2008-11-03 00:00:00
2008-09-15 00:00:00	2008-09-11 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00
2008-09-15 00:00:00	2008-09-15 00:00:00

- Number and sizes of brain metastasis in individual NSCLC patients (n=31)



Clinical problem



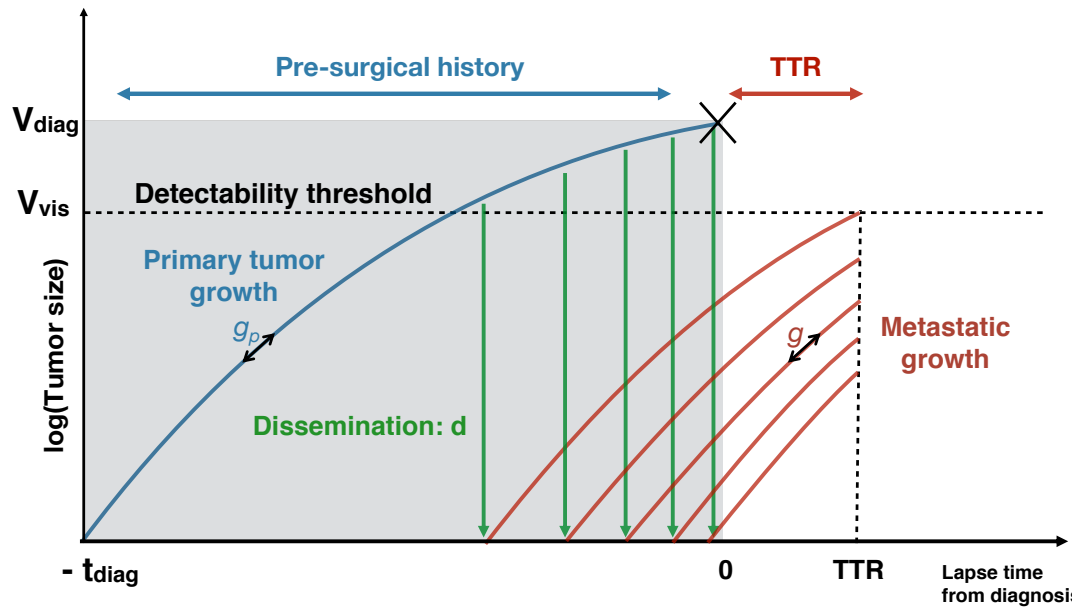
Mathematical models

$$\begin{cases} \partial_t \rho(t, v) + \partial_v (g(v) \rho(t, v)) = 0 \\ g(V_0) \rho(t, V_0) = d(V_p(t)) \left(+ \int_{V_0}^{+\infty} d(v) \rho(t, v) dv \right) \\ \rho(0, v) = \rho^0 \end{cases}$$

Data



Mathematical models



Growth rates of **primary** and **secondary** tumors g_p and g

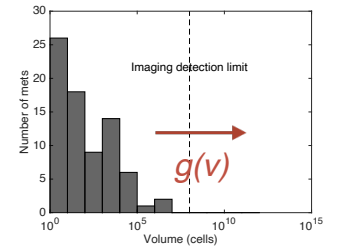
$$\frac{dV_p}{dt} = g_p(t, V_p)$$

ex: Gompertz

Dissemination rate $d(V_p) = \mu V_p^\gamma$

Size distribution of the metastases $\rho(t, v)$ [Iwata et al., 2000]

$$\begin{cases} \partial_t \rho(t, v) + \partial_v (g(v) \rho(t, v)) = 0 \\ g(V_0) \rho(t, V_0) = d(V_p(t)) \left(+ \int_{V_0}^{+\infty} d(v) \rho(t, v) dv \right) \\ \rho(0, v) = \rho^0 \end{cases}$$



Metastatic burden (total number of metastatic cells)

$$M(t) = \int_{V_0}^{+\infty} v \rho(t, v) dv = \int_0^t d(V_p(t-s)) V(s) ds$$

Number of metastases with size larger than the **visible size** V_{vis}

$$N_{vis}(t) = \int_{V_{vis}}^{+\infty} \rho(t, v) dv = \int_0^{t-\tau_{vis}} d(V_p(t)) dt$$

Time to relapse (TTR) = time from diagnosis to first visible met

$$TTR = \inf \{ t > 0 : N_{vis}(t_{diag} + t) \geq 1 \}$$

Statistical procedure for model calibration: nonlinear mixed effects modeling

- Classical approach considers each **subject independently**

Likelihood maximization



$$y_j^i = M(t_j^i, \theta^i) + \sigma \varepsilon_j^i, \quad \varepsilon_j^i \sim \mathcal{N}(0, 1)$$

Subject $1 \leq i \leq N$, Time t_j

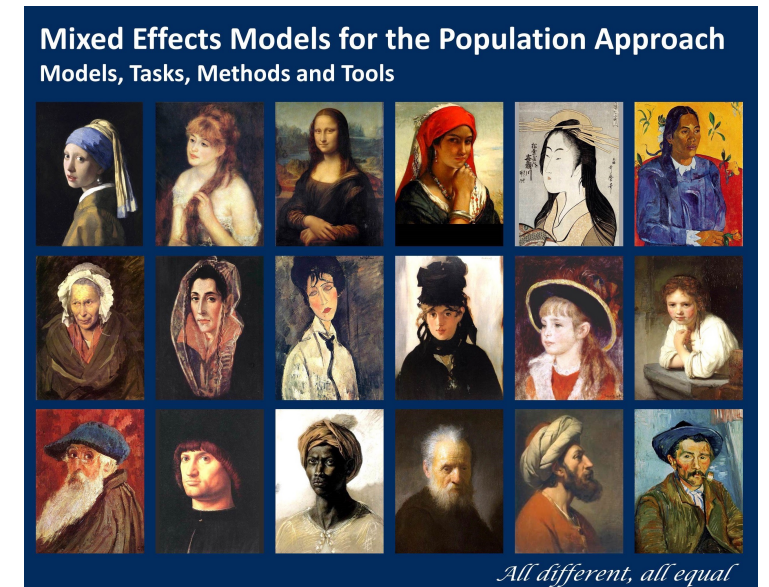
$$\hat{\theta}^i = \min_{\theta^i} \sum (y_j^i - M(t_j^i, \theta^i))^2$$

- Population** approach

$$y_j^i = M(t_j^i, \theta^i) + \sigma \varepsilon_j^i, \quad \varepsilon_j^i \sim \mathcal{N}(0, 1)$$

$$+ \quad \ln(\theta^i) = \ln(\theta_{pop}) + \eta^i, \quad \eta^i \sim \mathcal{N}(0, \omega^2)$$

- Parameters to be estimated = θ_{pop} (p), ω ($\frac{p(p+1)}{2}$) and σ

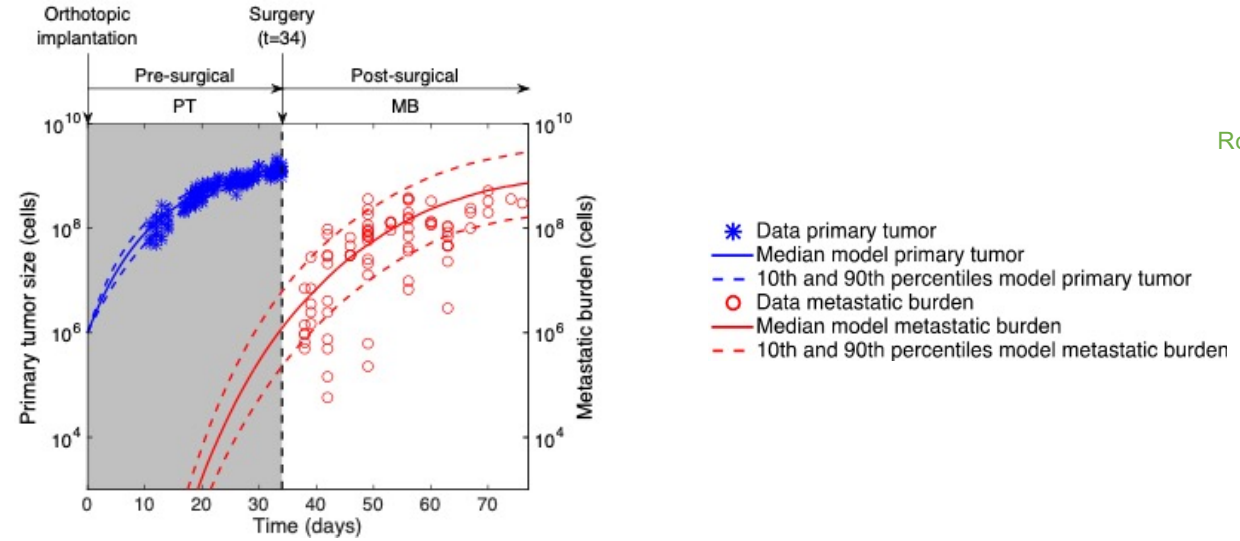
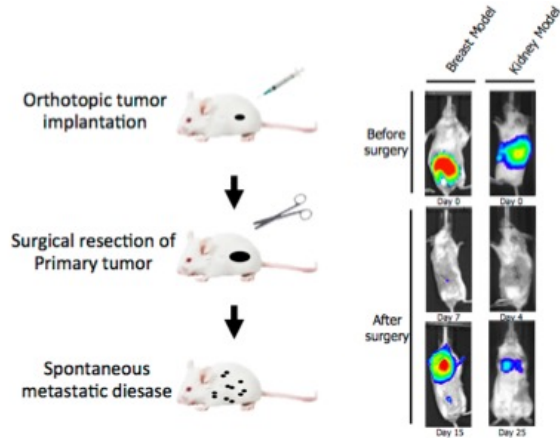




Ebos lab

Roswell Park Cancer Institute

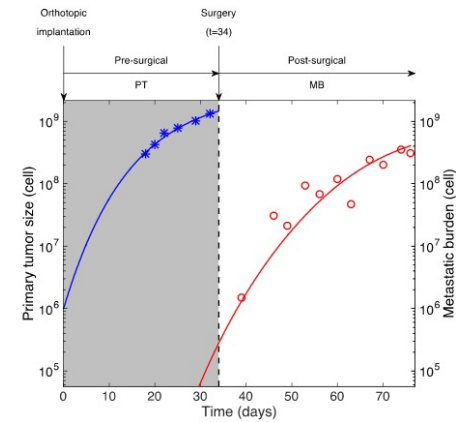
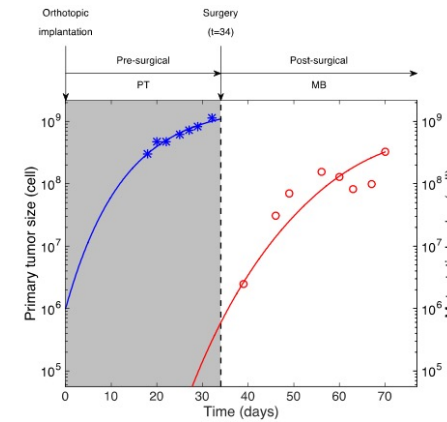
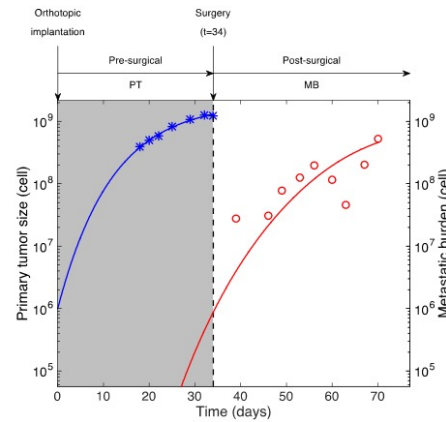
Validation on animal data



Nonlinear **mixed-effects** statistical model for inter-animal variability

$$\ln(\theta^i) = \ln(\theta_{pop}) + \eta^i, \quad \eta_i \sim \mathcal{N}(0, \omega^2)$$

- **Main difficulty:** PDE model
- **Solution:** fast Fourier transform for computation of metastatic burden



⇒ **same growth** for PT and mets: $\alpha_p = \alpha, \beta_p = \beta$

Clinical data of individual breast metastatic relapse

K = 21 features

outcome

n = 642 patients w/o adjuvant

menopausal_status	ER	PR	Ki67	HER2	HER2_intensity	CK56	EGFR	VIM	ALDH1
Post-ménopause	20	0	0	0	0	0	0	0	0
Ménopause	40	95	8	0	0	0	0	0	0
Activité génitale	87	10	26	0	0	0	0	80	0
Post-ménopause	100	100	8	0	0	0	0	0	0
Post-ménopause	0	0	16	82	+++	0	0	0	0
Activité génitale	100	95	12	0	0	0	0	0	1
Activité génitale	56	100	17	0	0	0	0	0	0
Activité génitale	57	85	23	100	+++	0	0	0	0
Post-ménopause	80	5	20	0	0	0	0	0	0
Post-ménopause	0	0	15	100	+++	0	5	0	0
Post-ménopause	100	80	10	0	0	0	0	0	0
Post-ménopause	30	0	5	0	0	0	0	0	0
Post-ménopause	0	0	15	40	+++	0	0	0	0
Ménopause	0	80	8	0	0	0	0	0	0
Post-ménopause	0	0	27	0	0	0	30	0	1
Post-ménopause	0	0	56	0	0	80	60	100	0
Activité génitale	50	92	2	1	+	0	0	0	0
Post-ménopause	0	47	5	0	0	0	0	80	0
Post-ménopause	65	0	10	0	0	0	0	60	0
Post-ménopause	100	50	11	0	0	0	0	0	0
Ménopause	20	100	0	0	0	0	0	0	0
Activité génitale	90	6	5	0	0	0	0	0	0
Post-ménopause	100	3	5	0	0	0	0	0	0
Activité génitale	0	0	6	0	0	0	0	0	0
Ménopause	80	100	5	0	0	0	0	0	0
Post-ménopause	100	85	25	0	0	0	0	0	0
Post-ménopause	10	45	11	13	+++	0	0	0	0
Post-ménopause	66	1	2	40	++	0	0	0	0

metastatic_relapse	date_metastatic_relapse
Yes	04/02/1999
No	
No	
No	
Yes	04/09/1990
Yes	08/02/1993
Yes	15/12/1999
No	
No	
Yes	08/03/1995
No	
Yes	06/04/1990
Yes	02/11/1994
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
No	
Yes	27/10/1999
No	
No	
No	
No	
No	
...	

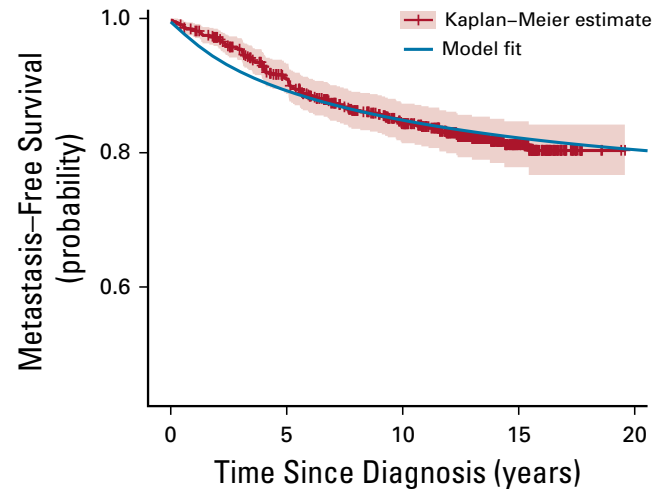
Descriptive power: fit to the data

$$\ln(T^i) = \ln(TTR(V_{diag}^i; \alpha^i, \mu^i)) + \varepsilon^i, \quad \varepsilon^i \sim \mathcal{N}(0, \sigma^2)$$

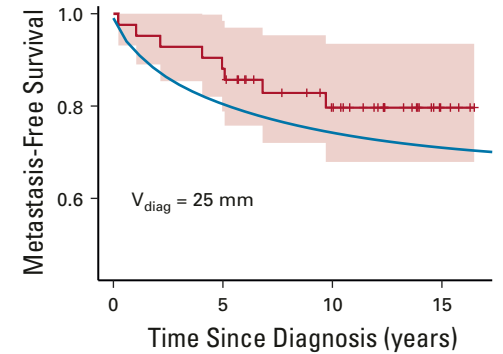
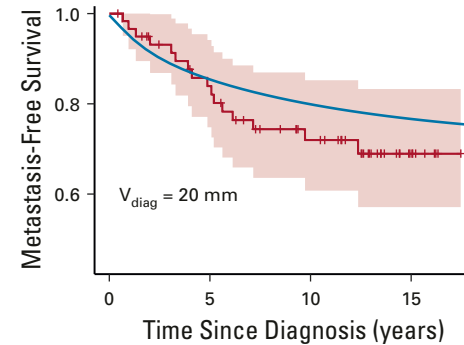
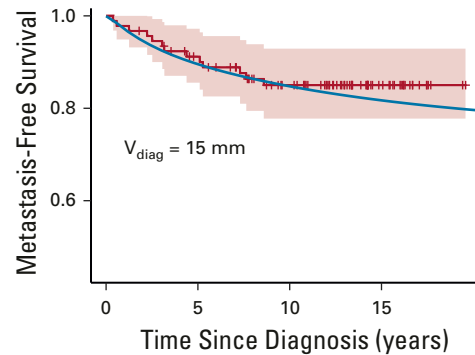
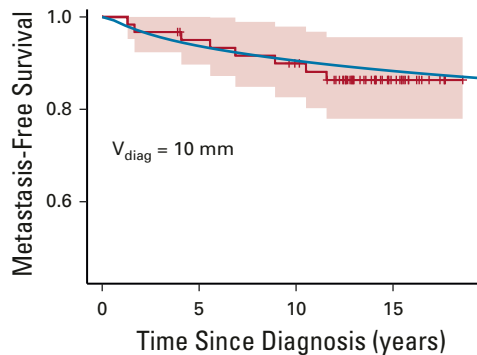
$$S(t|\alpha^i, \mu^i) = \mathbb{P}(T^i > t|\alpha^i, \mu^i)$$

$$\ln(\alpha^i) = \ln(\alpha_{pop}) + \eta_{\alpha}^i, \quad \eta_{\alpha}^i \sim \mathcal{N}(0, \omega_{\alpha}^2)$$

$$\ln(\mu^i) = \ln(\mu_{pop}) + \eta_{\mu}^i, \quad \eta_{\mu}^i \sim \mathcal{N}(0, \omega_{\mu}^2)$$

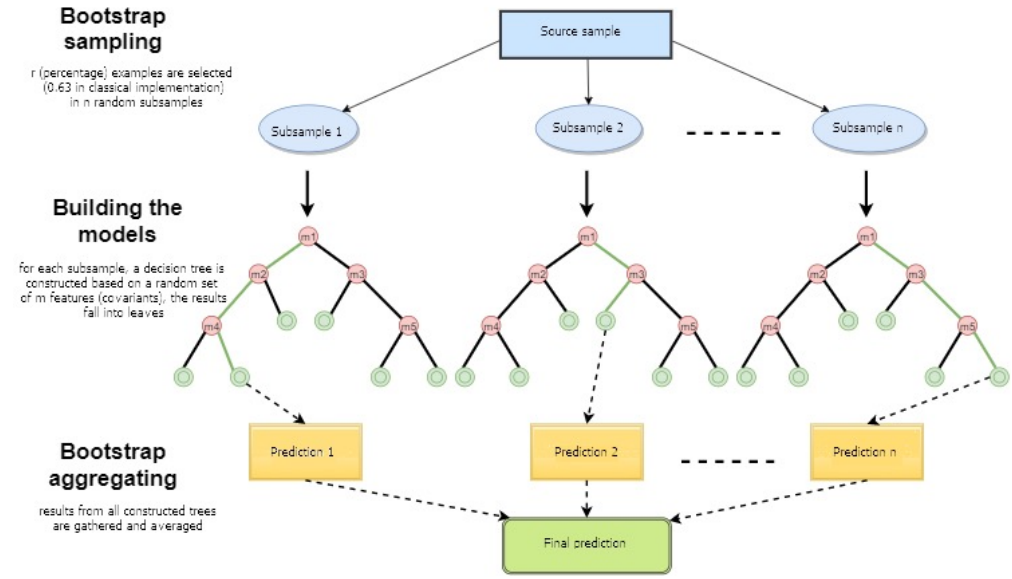
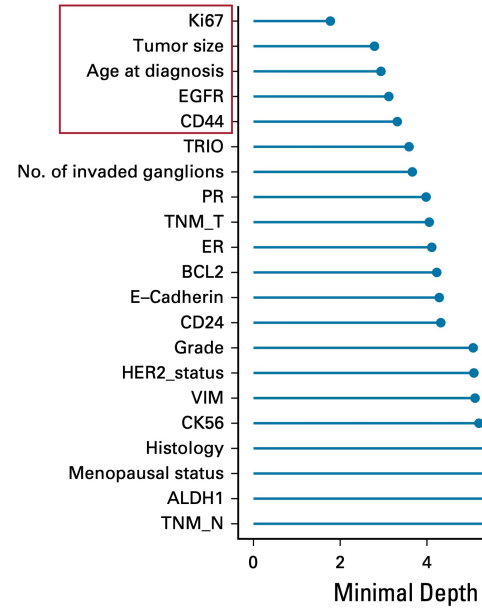
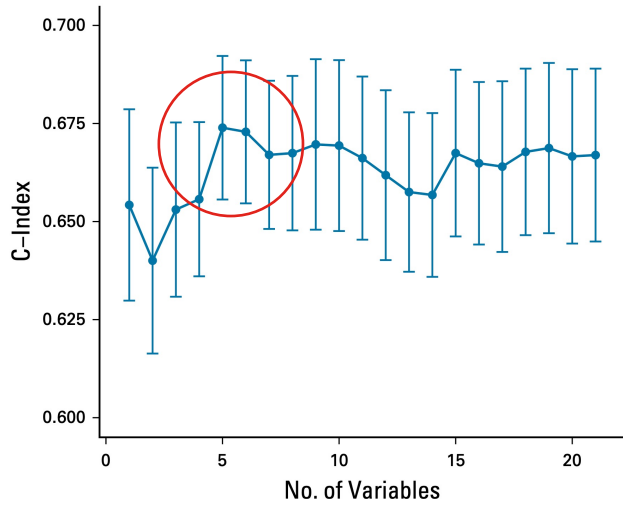


Parameter	Estimate	Relative Standard Error (%)
Model without covariates		
$\log \alpha_{pop}$	-6.34	12.6
$\log \mu_{pop}$	-26.8	3.68
σ	0.542	28.4
ω_{α}	3.37	36.4
ω_{μ}	3.78	15.9

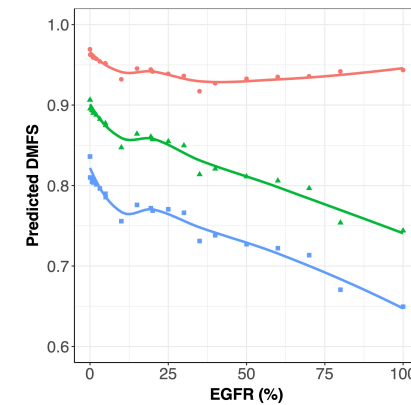
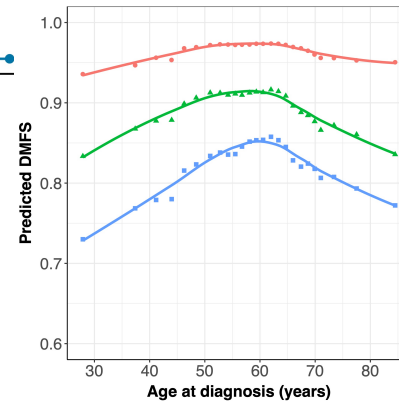
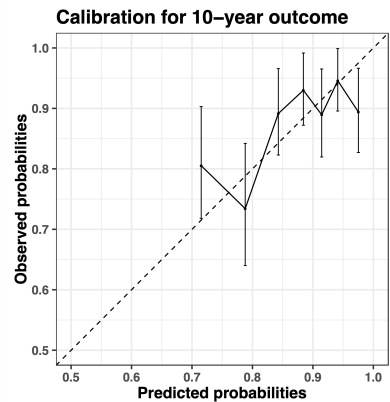
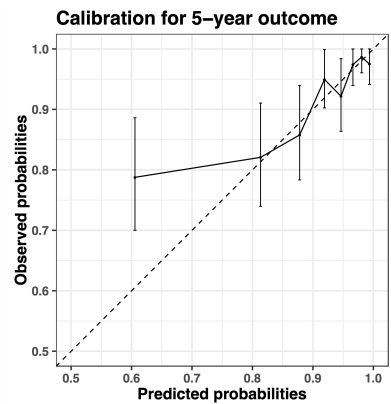


Feature selection: random survival forests

c-index = 0.69 (0.67 – 0.71)



M. Dmitrievsky, <https://www.mqI5.com/en/articles/3856>

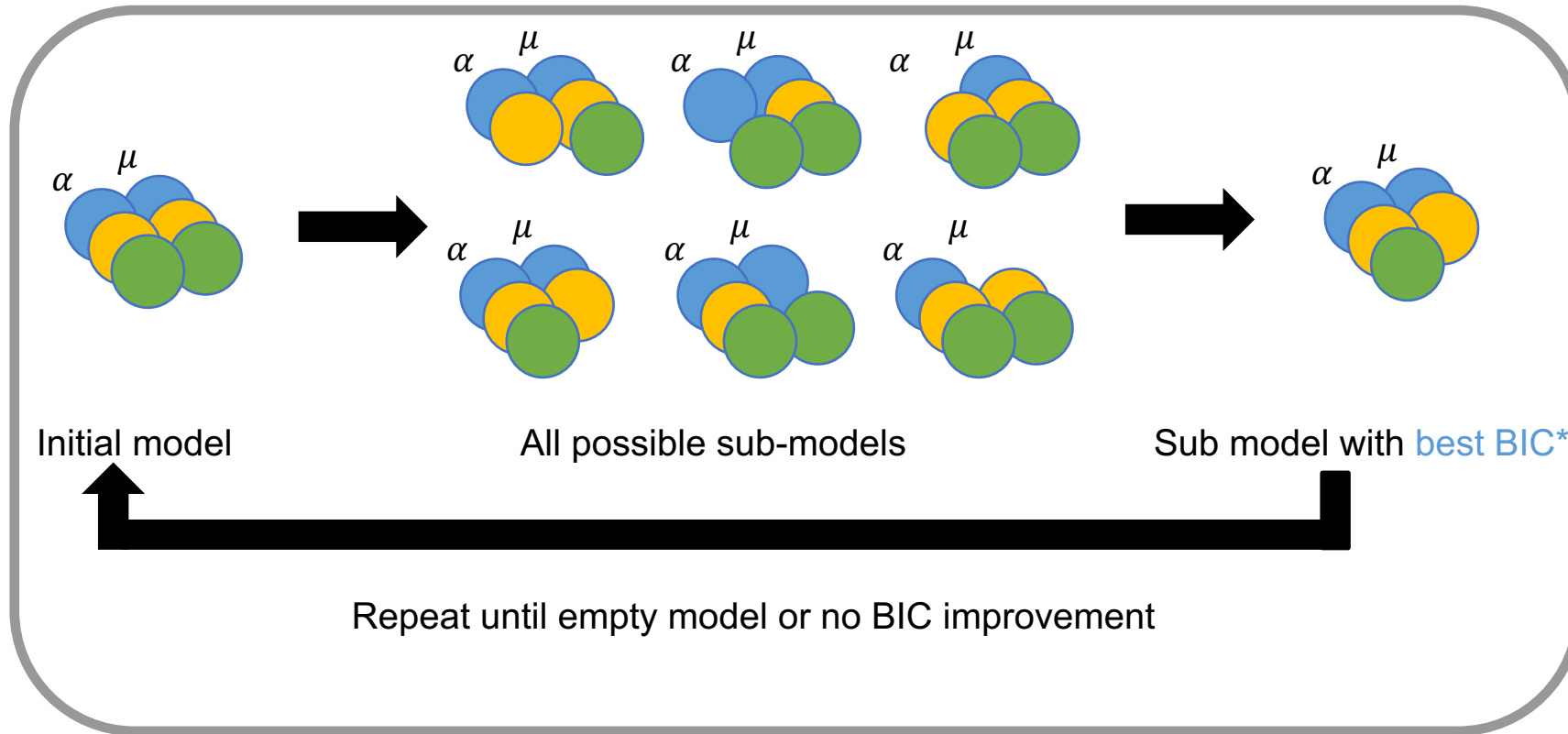


Time
 ● 2 years
 ▲ 5 years
 ■ 10 years

DMFS = Distant Metastasis-Free Survival

⇒ nonlinear effect of covariates and non proportional hazard

Mechanistic selection: backward stepwise selection



$$\ln(\mu^i) = \ln(\mu_{pop}) + \beta_{\mu}^T \mathbf{x}_{\mu}^i + \eta_{\mu}^i, \quad \eta_{\mu}^i \sim \mathcal{N}(0, \omega_{\mu}^2)$$

$$\ln(\alpha^i) = \ln(\alpha_{pop}) + \beta_{\alpha}^T \mathbf{x}_{\alpha}^i + \eta_{\alpha}^i, \quad \eta_{\alpha}^i \sim \mathcal{N}(0, \omega_{\alpha}^2)$$

Parameter	Estimate	Relative Standard Error (%)	P
Model with covariates			
$\log \alpha_{pop}$	-9.01	10.8	
$\beta_{Ki67, \alpha}$	0.093	29.6	.001
$\beta_{CD44, \alpha}$	0.017	57.7	.083
$\log \mu_{pop}$	-25.9	4.4	
$\beta_{EGFR, \mu}$	0.053	38.1	.009
σ	0.606	24	
ω_{α}	2.75	22.1	
ω_{μ}	3.03	20.5	

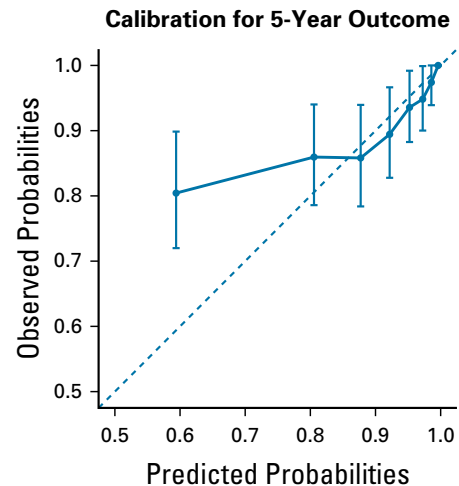
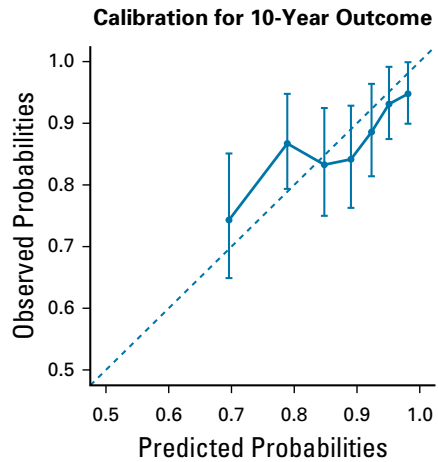
Selected Covariate model:

α ← Ki67 (proliferation marker), CD44 (stem cell marker)

μ ← EGFR (basal marker)

Predictive power

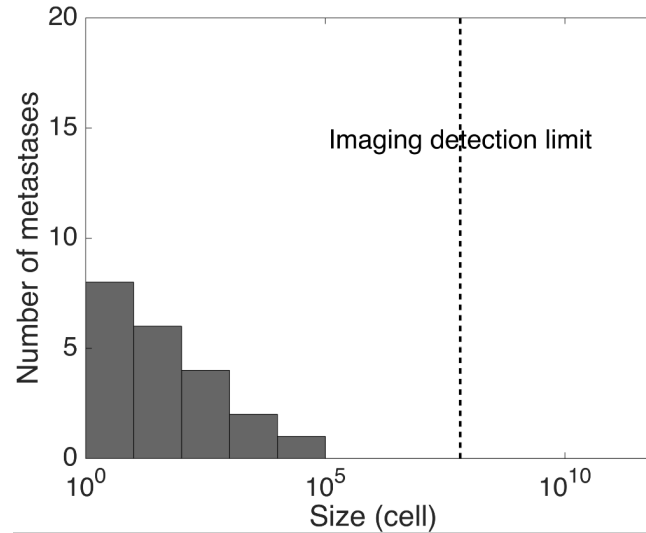
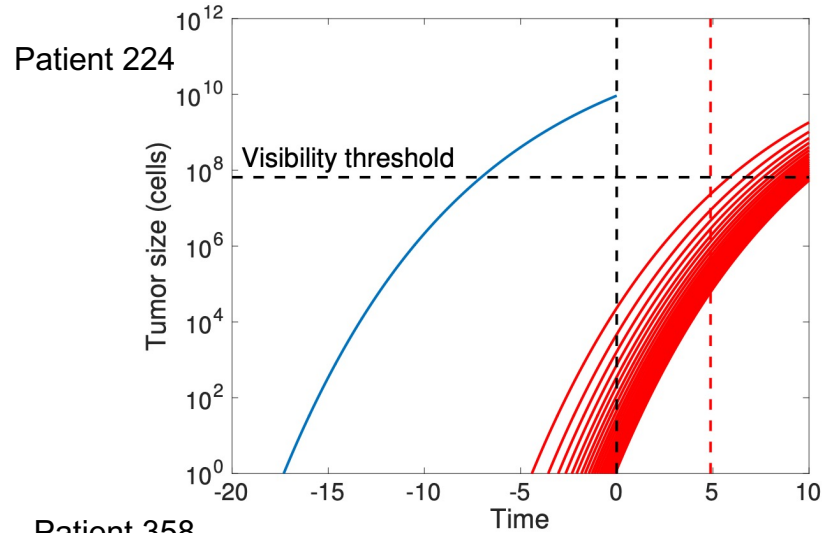
c-index = 0.67
(10-folds cross-validation)



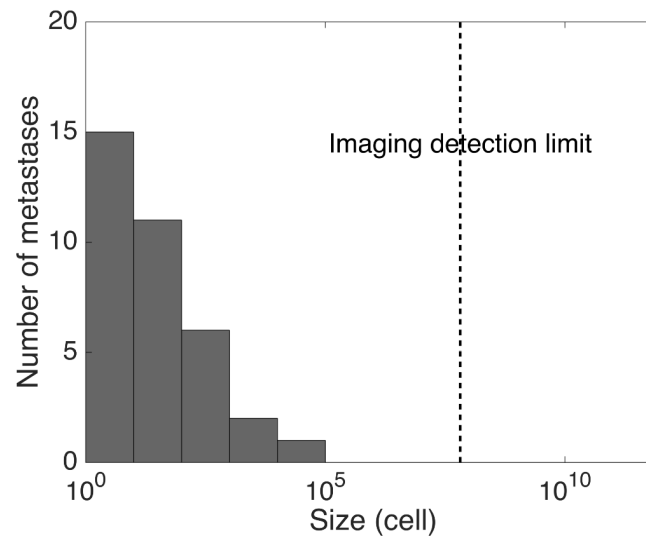
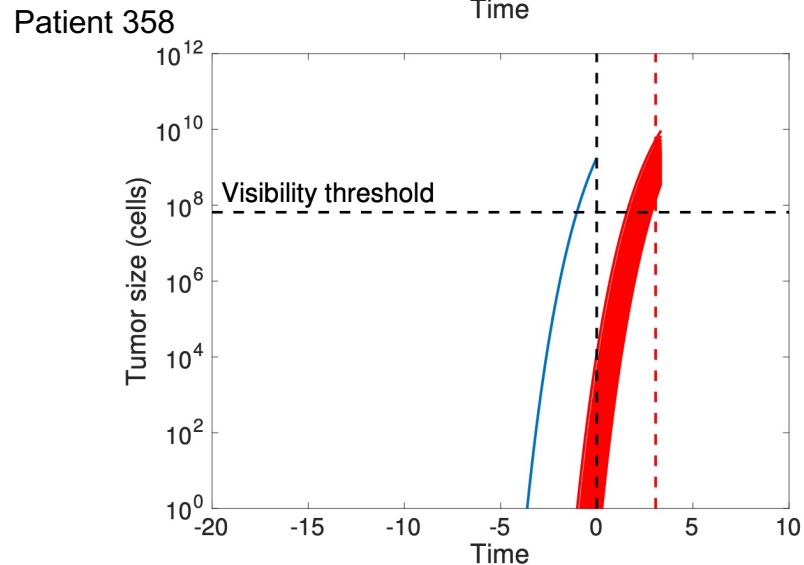
Algorithm	AUROC	Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1
5-year							
Mechanistic model	0.73	0.68	0.75	0.67	0.19	0.96	0.30
Random survival forest	0.73	0.69	0.64	0.70	0.18	0.95	0.28
Random forest	0.75	0.66	0.71	0.66	0.18	0.96	0.28
Logistic regression	0.75	0.83	0.42	0.87	0.24	0.94	0.31
k-Nearest neighbor	0.62	0.91	0.02	1.00	0.41	0.91	0.05
Gradient boosting	0.71	0.90	0.11	0.98	0.36	0.92	0.17
Support vector machine	0.64	0.87	0.09	0.95	0.15	0.91	0.11
Cox	0.71	0.72	0.66	0.73	0.20	0.95	0.31
10-year							
Mechanistic model	0.67	0.67	0.62	0.68	0.30	0.89	0.41
Random survival forest	0.69	0.62	0.71	0.60	0.28	0.90	0.41
Cox	0.65	0.65	0.61	0.65	0.28	0.88	0.39

⇒ **Similar predictive power** as classical statistical Cox model or other machine learning algorithms

Predictive simulations of the mechanistic model



	Patient 224	Patient 358
Tumor size (mm)	26	15
Ki67 (%)	40	52
EGFR (%)	0	80
CD44 (%)	0	0
Observed TTR (years)	4.88	3.06
Predicted TTR (years)	6.18	1.97
Prediction error (years)	1.3	1.09
Pre-surgical history (years)	17.3	3.61
Number of mets	21	35



$$\log \hat{\alpha}^i = \log \alpha_{pop} + \beta_{Ki67,\alpha} \cdot Ki67^i + \beta_{CD44,\alpha} \cdot CD44^i$$

$$\log \hat{\mu}^i = \log \mu_{pop} + \beta_{EGFR,\mu} \cdot EGFR^i$$

Second dataset: PAI-1



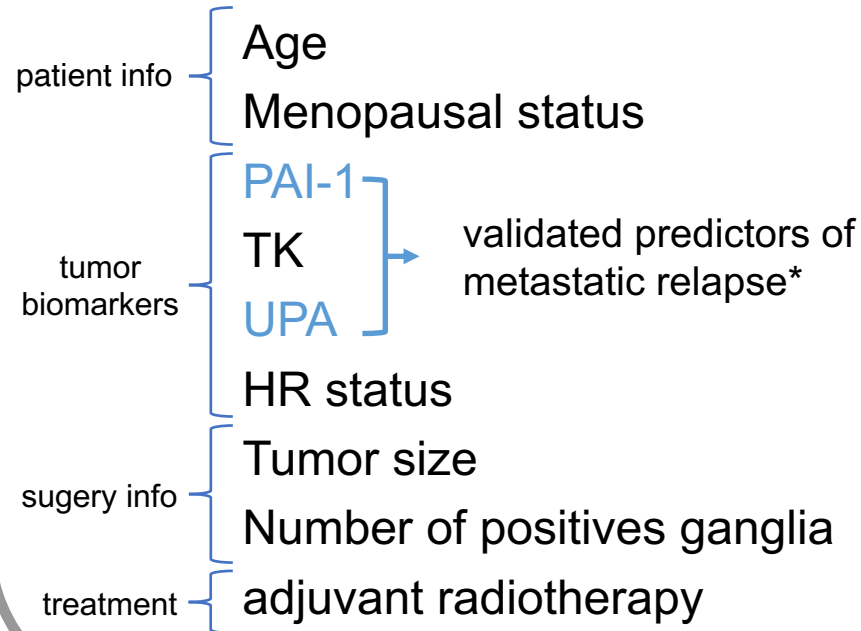
C. Bigarré
X. Muracciole

Data

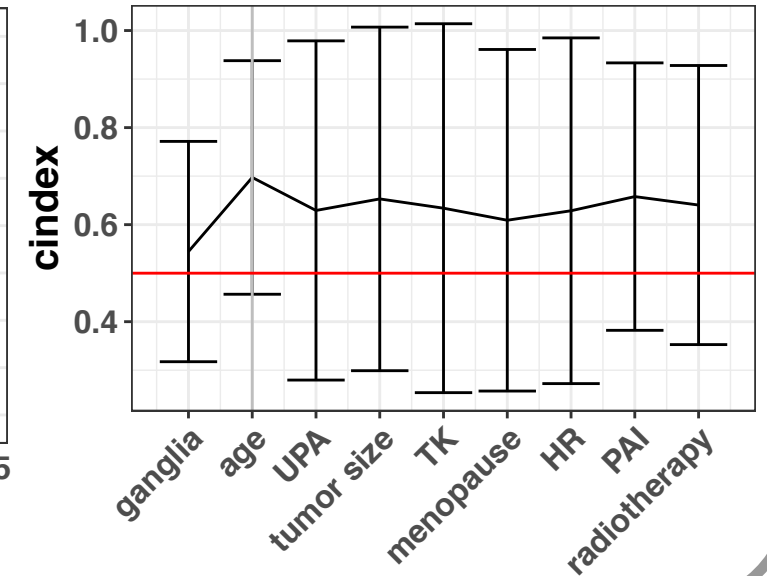
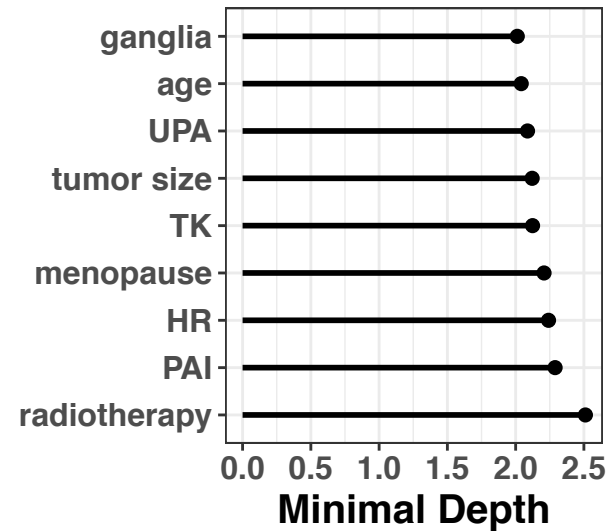
N = 167 patients

Outcome: metastasis free survival

p = 9 covariates



Random survival forest



No clear results for variable importance



All covariates included in the next step

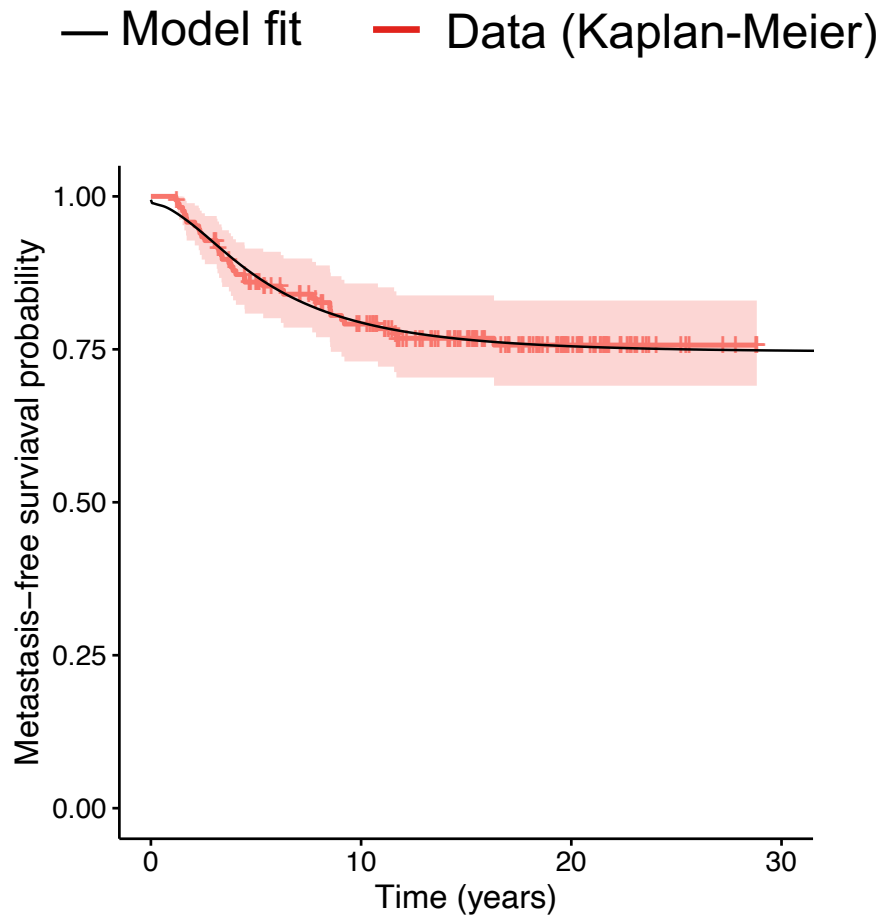
*Duffy, M. J. et al., M. uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Res* **16**, 428 (2014).

*Harbeck N. et al., Ten-year analysis of the prospective multicentre Chemo-N0 trial validates ASCO-recommended biomarkers uPA and PAI-1 for therapy decision making in node-negative breast cancer patients, *Eur J Cancer* (2013)

Selected mechanistic model



C. Bigarré



$$\begin{aligned}\log T^i &\sim \log(TTR(V^i, \alpha^i, \mu^i) + \mathcal{N}(0, \sigma^2)) \\ \log(\alpha^i) &\sim \log(\alpha_{pop}) + \beta_{HR,\alpha} \cdot 1_{HR}^i + \mathcal{N}(0, \Omega_\alpha^2) \\ \log(\mu^i) &\sim \log(\mu_{pop}) + \beta_{PAI,\mu} \cdot PAI^i + \mathcal{N}(0, \Omega_\mu^2)\end{aligned}$$

Parameter	Estimate	R.S.E	p-value
α	0.019	18%	
$\beta_{HR,\alpha}$	-0.713	50.1%	0.0459
μ	5.13e-15	381%	
$\beta_{PAI,\mu}$	0.352	31.7%	0.0016
σ	0.377	16.9%	
Ω_α^2	0.308	109%	
Ω_μ^2	14.9	34.1%	

R.S.E = Relative Standard Error

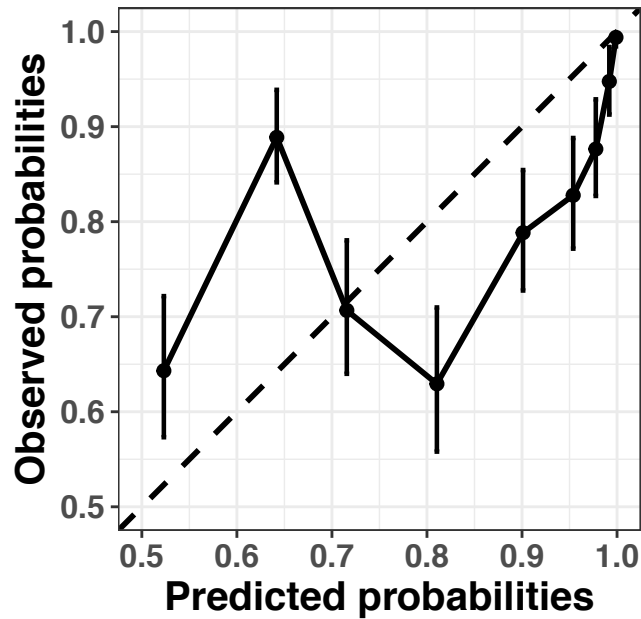


C. Bigarré

Predictive performances

RSF

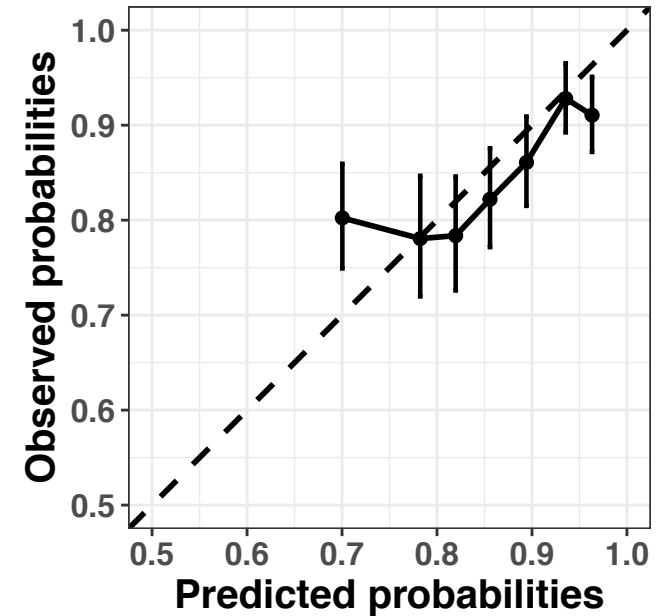
c-index: 0.70 (0.45-0.94)



Calibration at 10 years

Mechanistic model

c-index: 0.70 (0.68-0.72)



Cox

c-index

Full model (p=9)

0.67 (0.33, 1.00)

Tumor size + HR + PAI-1

0.71 (0.45, 0.97)

Conclusions and perspectives

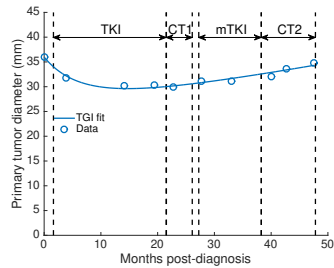
- Similar predictive performances of Cox regression (c-index 0.67 - 0.72), random survival forest (c-index 0.67-0.71) and a novel mechanistic model (c-index 0.63 - 0.70) **for pure prediction**
- Predictive power is confirmed (improved) in an external data set
- Mechanistic modeling provides **biological and clinical insights** that ML does not:
 - Ki67 correlates with proliferation rate α (expected but reassuring), also CD44 or hormonal status
 - EGFR and PAI1 correlate with μ (metastatic potential)
 - prediction of the **invisible metastatic state** at diagnosis \Rightarrow potential for **personalized adjuvant therapy**
- Current/future avenues:
 - Further investigations to **refine the modeling** (dormancy, etc...)
 - Include **treatment**
 - Include (high-dimensional) transcriptomic data

Results: individual clinical data

Methodology

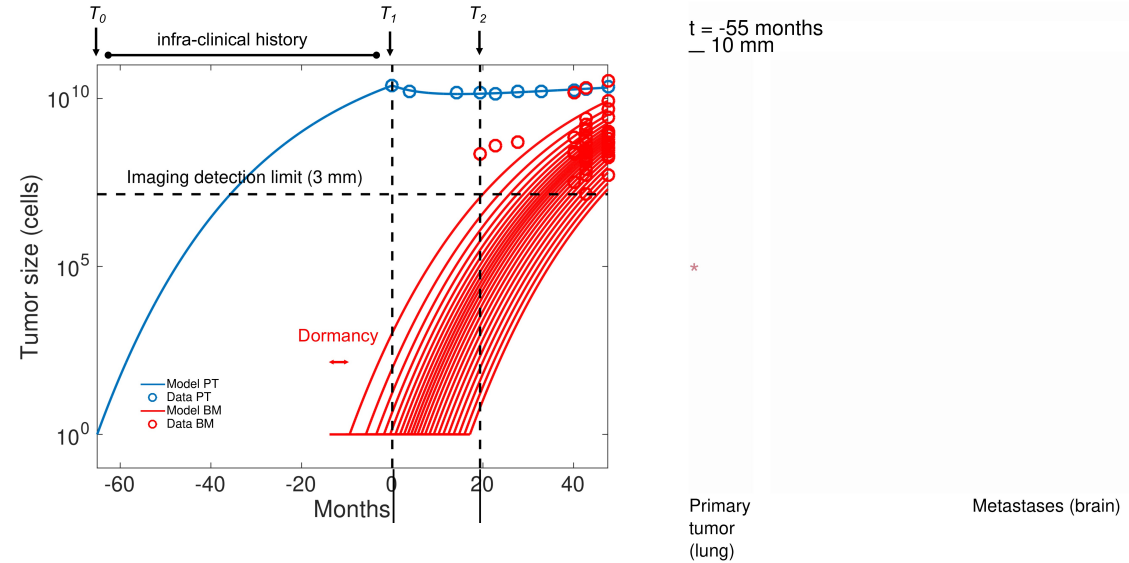
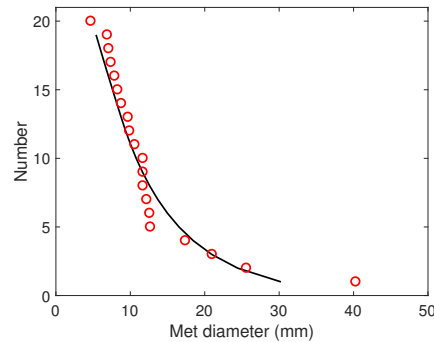
Primary tumor

$$\frac{dS_p}{dt} = \alpha_1 S_p - \kappa e^{-\frac{t-T_d}{t_{res}} \ln 2} S_p, \quad \forall t \geq T_d.$$



Metastases

$$f(t, v) = \int_v^{+\infty} \rho(t, u) du = N(t - \tau(v))$$



- Main difficulties:

no data on PT pre-treatment + fit PDE to discrete data

Results

- Dormancy, estimated to 133 days \pm 4.2%
- Good fit of the entire longitudinal data (47 BM sizes over 6 time points) with only **3 parameters**.
- Onset of BM 14-19 months before diagnosis

Model



Data



Ongoing (collab. P. Schlicke, TU München)

- Extend to larger cohort: **complex** treatment histories
- Include effect of **treatment** $\Rightarrow g(t, v) \Rightarrow$ nontrivial difficulties for param estimation

Differential effects of anti-angiogenic therapies between primary tumor and metastases



Cancer Cell
Report

Published online: October 31, 2014

Research Article



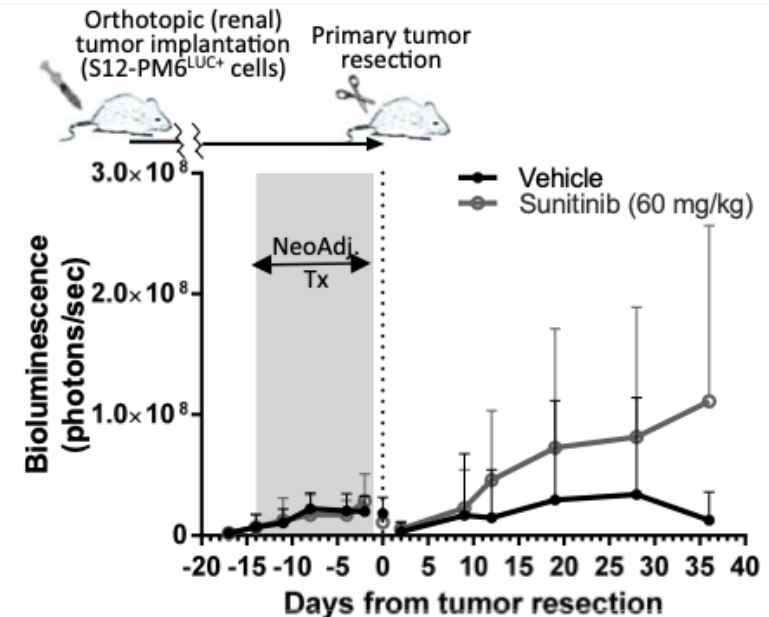
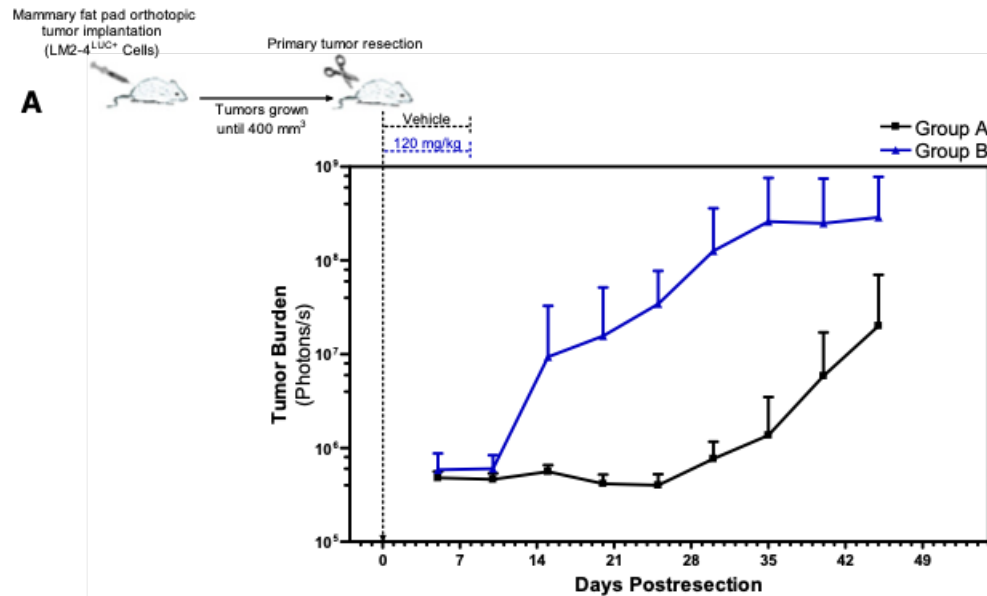
EMBO
Molecular Medicine

Accelerated Metastasis after Short-Term Treatment with a Potent Inhibitor of Tumor Angiogenesis

John M.L. Ebos^{1,2}, Christina R. Lee,¹ William Cruz-Munoz,¹ Georg A. Bjarnason,³ James G. Christensen,⁴ and Robert S. Kerbel^{1,2,*}

Neoadjuvant antiangiogenic therapy reveals contrasts in primary and metastatic tumor efficacy

John M L Ebos^{1,*}, Michalis Mastro¹, Christina R Lee², Amanda Tracz¹, John M Hudson², Kristopher Attwood³, William R Cruz-Munoz², Christopher Jedszko², Peter Burns^{2,4} & Robert S Kerbel^{2,4}

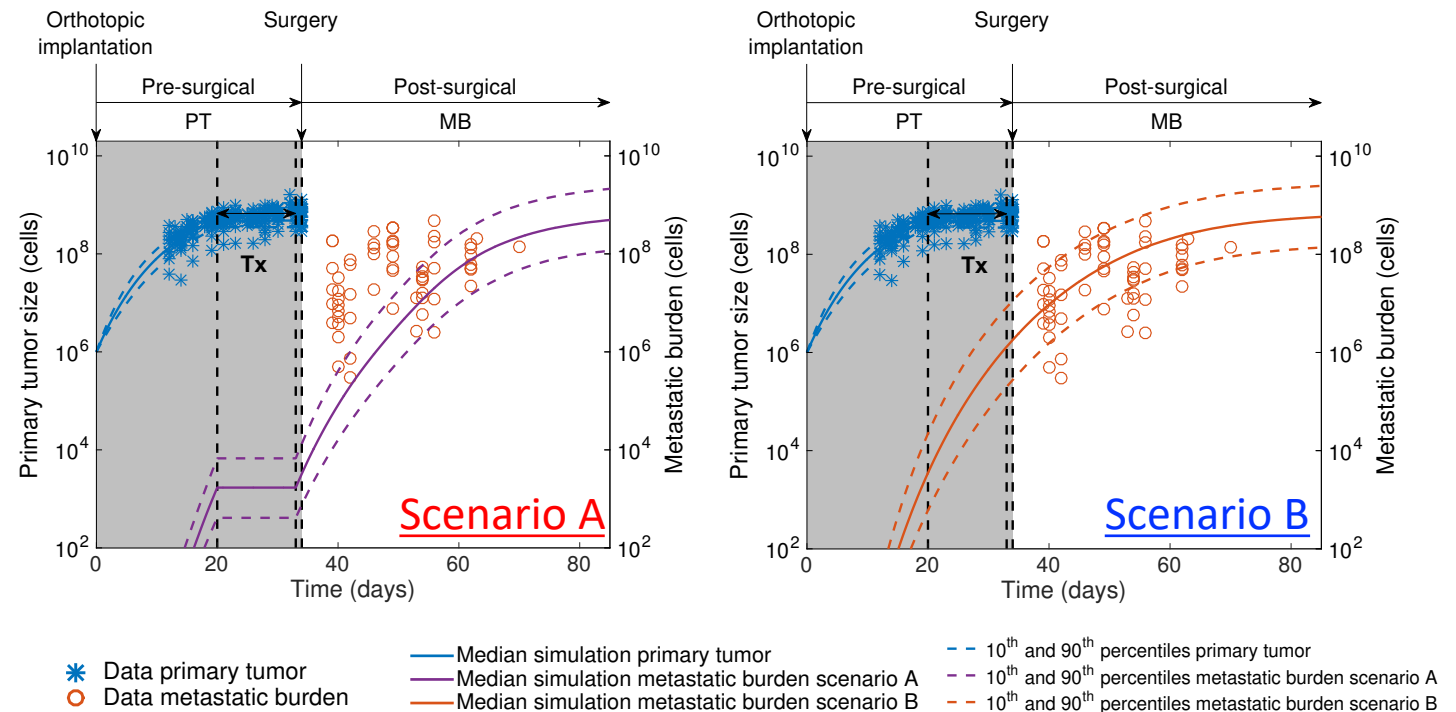
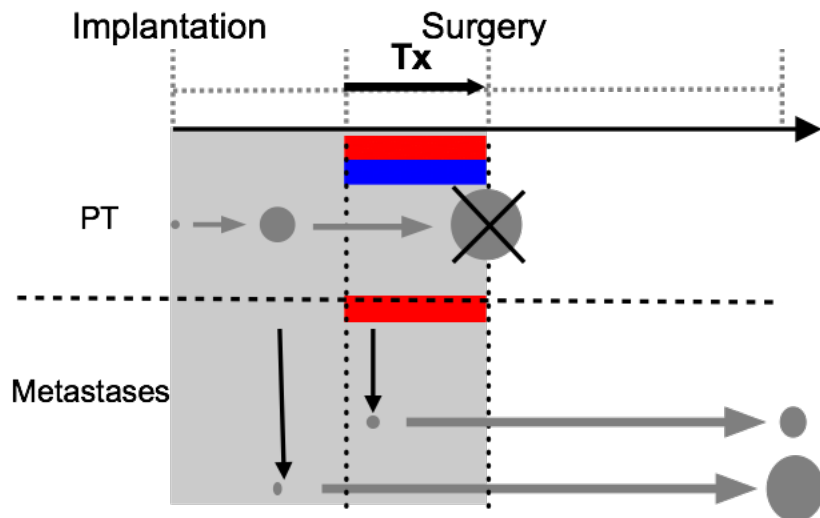


Simulations of the effect of neoadjuvant sunitinib treatment on metastases suggest no effect on growth of metastases

- Parameter values from the previous study on control groups \Rightarrow simulations are pure mechanistic predictions
- In first approximation, the effect of the drug was modeled by setting the tumor growth rate to zero during the phase of treatment

Tx Effects

- Scenario A**
PT and MB growth arrest
- Scenario B**
PT growth arrest only





oui
nide
iou

We have open positions!!

- Full research tenure position

sebastien.benzekry@inria.fr

