Evaluation of the predictive capacity of a biomarker

Bassirou Mboup
(ISUP Université Paris VI)
Paul Blanche
(Université Bretagne Sud)
Aurélien Latouche
(Institut Curie & Cnam)

GDR STATISTIQUE ET SANTE, 10/05/2017

Goal and Definition

Goal :To assess the predictive capacity of a biomarker.

- The **prognostic** value of a marker corresponds to its ability to predict the clinical course of the disease in the **absence of treatment**.
- The **predictive** value of a marker corresponds to its ability to predict the response to **a given treatment**.
  - Genomic test that stratify patients to low/med/high risk of recurrence to guide the treatment with chemo
  - Proliferation Marker (Ki67) to guide the trt with chemo

Evaluating marker predictive performance

- In practice, only a significant interaction test between the treatment assignment and the marker value is used to establish the predictive capacity of the marker.
- A strong interaction is a necessary condition for a predictive marker. However, this not a sufficient condition (Janes et al., 2011).
- Janes et al. (2014) proposed a formal and comprehensive approach to the evaluation of predictive marker with binary endpoint.

Some notations

- D is a binary indicator of the clinical outcome.
- $Treat$ the treatment indicator.
  - $Treat = 1$ if treated
  - $Treat = 0$ if no treated
- We note that
  - $\rho_0 = \mathrm{P}(D = 1 | Treat = 0)$
  - $\rho_1 = \mathrm{P}(D = 1 | Treat = 1)$
- Y a continuous marker.
- We define the risk as the conditional probability of having the event given treatment and biomarker.

$$risk(Y) = P(D = 1 | Treat, Y)$$

Standard decisions in the Janes et al. (2014) approach

The marginal effect of treatment is defined as

$$\rho_0 - \rho_1 = \mathrm{P}(D = 1 | \mathit{Treat} = 0) - \mathrm{P}(D = 1 | \mathit{Treat} = 1)$$

- if $\rho_0 - \rho_1 > 0$ then the standard strategy is to treat all patients.
- if $\rho_0 - \rho_1 \leq 0$ then the standard strategy is to treat nobody.

Treatment rule

- Let $\Delta(Y) = P(D = 1 | \mathit{Treat} = 0, Y) - P(D = 1 | \mathit{Treat} = 1, Y)$ denote the absolute treatment effect given marker value Y.
- The treatment assignment is : do not treat if $\Delta(Y) < 0$.
- We refer to subject with
  - $\Delta(Y) < 0$ as "marker-negatives"
  - $\Delta(Y) > 0$ as "marker-positives".

Estimation and Inference

The observed data $(Y_i, D_i, Treat_i), i = 1, ..., n$

- The risk model (glm)

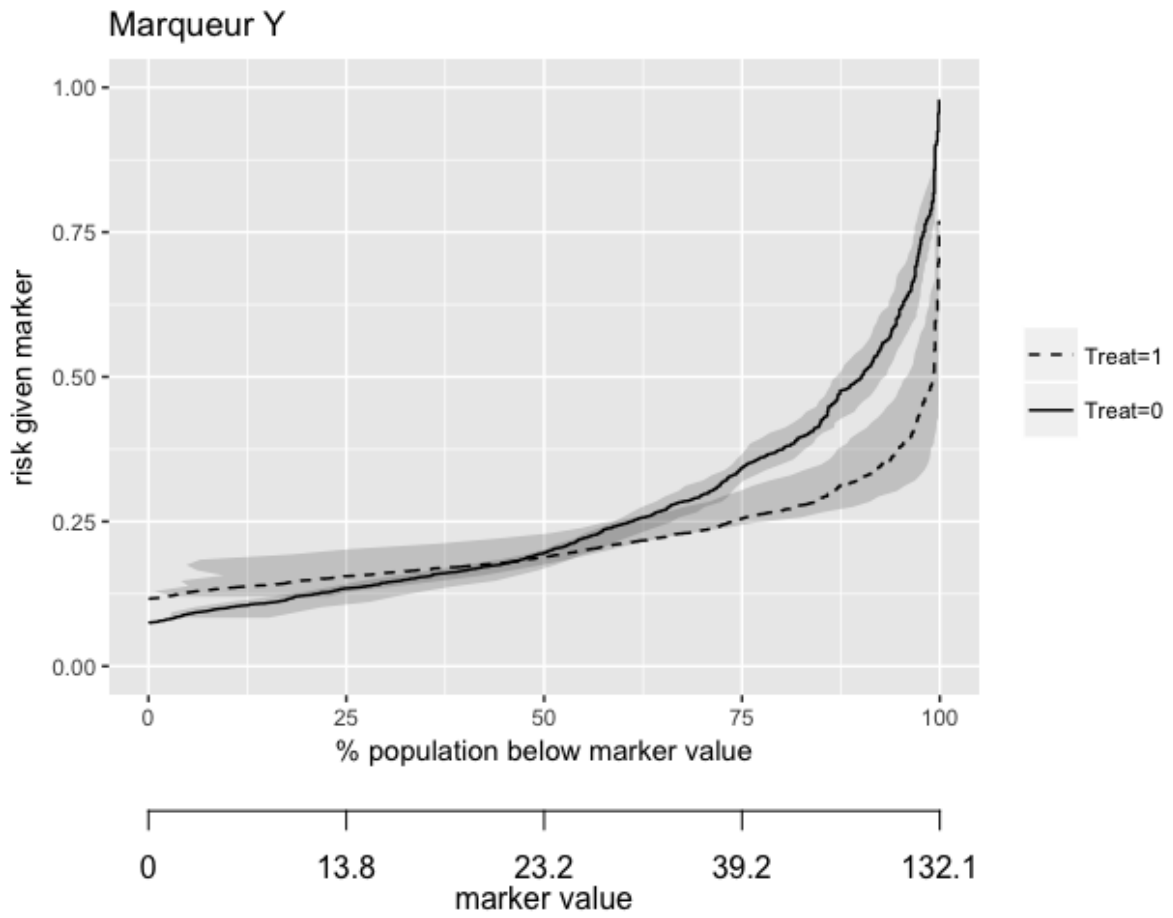
$$g(P(D = 1|Treat, Y) = \beta_0 + \beta_1\ Treat + \beta_2\ Y + \beta_3\ Treat \times Y$$

  with g the logit link

- $risk_0(Y) = \hat{P}(D = 1|Treat = 0, Y) = g^{-1}(\hat{\beta}_0 + \hat{\beta}_2 Y)$
- $risk_1(Y) = \hat{P}(D = 1|Treat = 1, Y) = g^{-1}(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 Y + \hat{\beta}_3 Y)$
- The absolute treatment effect is estimated by $risk_0$ and $risk_1$,

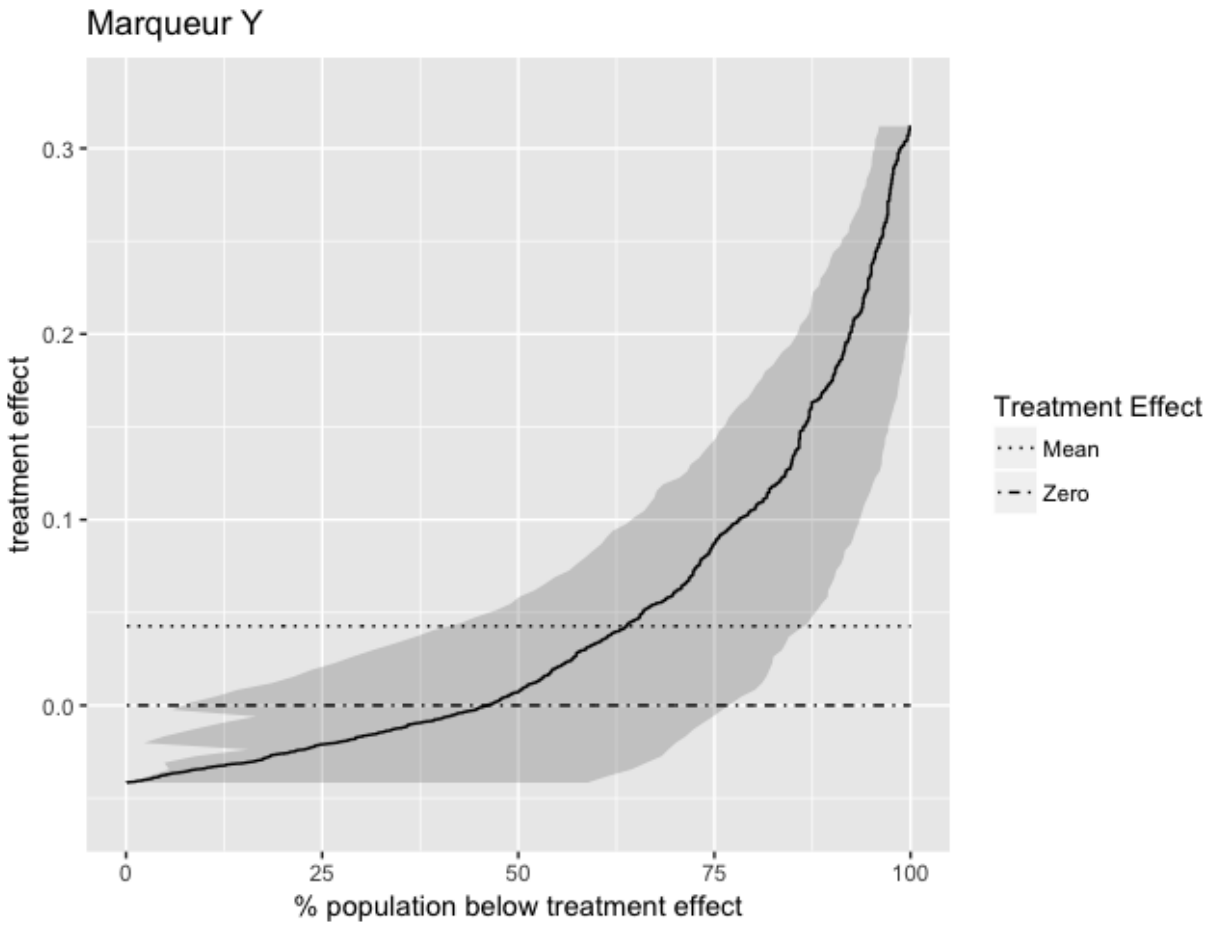
$$\hat{\Delta}(Y) = risk_0(Y) - risk_1(Y)$$

Marker-by-treatment Predictiveness curves with $\rho_0 - \rho_1 > 0$



The proportion of subjects with negative treatment effect who can avoid treatment, 46% for Y.

Distribution of the treatment effect : $\widehat{\Delta}(Y)$



Distribution of treatment effect : $\widehat{\Delta}(Y) = risk_0(Y) - risk_1(Y)$

Proposed extension to right censored data (1/2)

Objective : estimating the absolute treatment effect at time $t^\star$ in the presence of censoring.

- The dynamic prediction of risk can defined by :

$$risk(t, Y) = \mathrm{P}(T \leq t | Treat, Y)$$

with T be the observed failure time and $D(t) = \mathbb{1}_{(T \leq t)}$ the response indicator.

- The time-dependent absolute treatment effect is :

$$\Delta(t, Y) = \mathrm{P}(D(t) = 1 | Treat = 0, Y) - \mathrm{P}(D(t) = 1 | Treat = 1, Y)$$

☞ **A biomarker can be predictive up to a given time $t^\star$ and not after $t^\star$**

Proposed extension to right censored data (2/2)

- Zheng et al. (2006) proposed a time-dependent logistic regression model for prognostic accuracy with multiple biomarkers

$$g(P(T \leq t|M_i)) = \alpha(t) + \beta_1(t)M_1 + \cdots + \beta_p(t)M_p$$

- We use this model to predict the overall risk :

$$g(P(T \leq t|\mathit{Treat}, Y)) = \alpha(t) + \beta_1(t)\,\mathit{Treat} + \beta_2(t)\,Y + \beta_3(t)\,\mathit{Treat} \times Y$$

- We consider a simplified model,

$$g(P(T \leq t|\mathit{Treat}, Y)) = \underbrace{\alpha(t)}_{} + \beta_1\,\mathit{Treat} + \beta_2\,Y + \beta_3\,\mathit{Treat} \times Y$$

Estimation of model parameters (1/2)

To estimate the $\beta = (\beta_1, \beta_2, \beta_3)$ and α under this model, we partition the time axis into J non-overlapping intervals with cut-off points :

- $t_1 ... t_j ... t_{J-1}$ failure times.
- $D_{ij} = \mathbb{1}(t_{j-1} \leq T_i < t_j)$ response between t_{j-1} and t_j
- $\theta = \{\alpha_1 = \alpha(t_1), ..., \alpha_{J-1} = \alpha(t_{J-1}), \beta\}'$,
- $\mathbf{x_i} = (1, Treat_i, Y_i, Treat_i \times Y_i)$

$$\sum_{j=1}^{J-1} \sum_{i=1}^{n} \frac{\partial \log \frac{\nabla \mu_{ij}(\theta)}{\nabla \mu_{i,j+1}(\theta)}}{\partial \theta} \mathbf{x_i}' \{D_{ij}(t_j) - \mu_i(\alpha_j, \beta)\} = 0 \qquad (1)$$

where

$$\mu_i(\alpha_j, \beta) = \Lambda(\mathbf{x_i}\theta) = \Lambda(\alpha_j + \beta_1 Treat + \beta_2 Y + \beta_3 Treat \times Y)$$

$$\Lambda(x) = \frac{exp(x)}{1 + exp(x)}$$

$$\nabla \mu_{i,j}(\theta) = E(D_{ij} | Treat_i, Y_i, \theta)$$

Estimation of model parameters (2/2)

Suppose for the ith subject, we observe

- $T_i = \min(\tilde{T}_i, C_i)$
- $\delta_i = \mathbb{1}(\tilde{T}_i \leq C_i)$, where C is the censoring time

To account for censoring, we modify the previous equation and consider the following inverse probability of censoring weighting (IPCW) estimating equation :

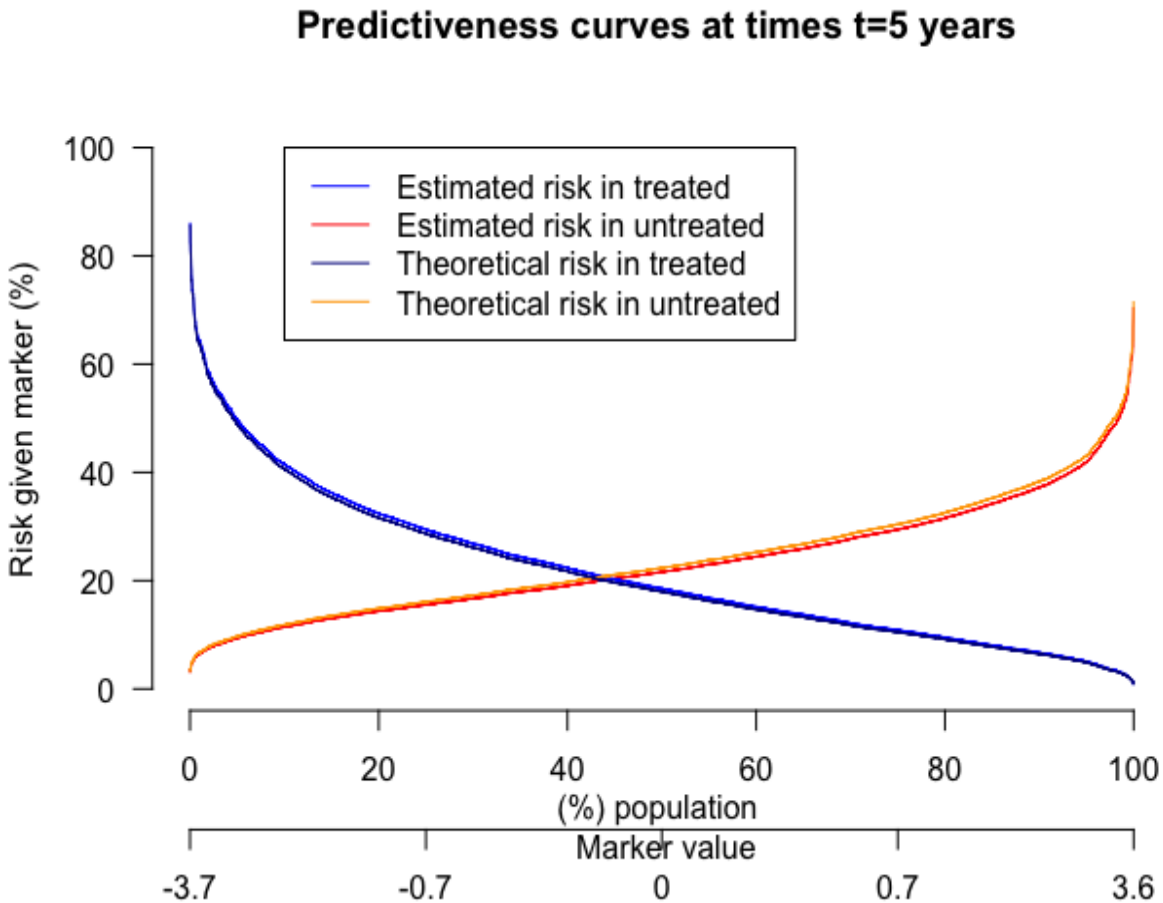
$$\sum_{j=1}^{J-1} \sum_{i=1}^{n} \frac{V_i(t_j)}{\hat{\pi}_i(t_j)} \frac{\partial log \frac{\nabla \mu_{i,j}(\theta)}{\nabla \mu_{i,j+1}(\theta)}}{\partial \theta} \mathbf{x_i}' \{D_i(t_j) - \mu_i(\alpha_j, \beta)\} = 0 \qquad (2)$$

$$V_i(t_j) = \begin{cases} 0 \text{ si } \widetilde{T}_i < t_j \text{ and } \delta_i = 0 \\ 1 \text{ sinon} \end{cases}$$

$\widehat{\pi}_i(t_j)$ is a consistent estimator of the probability of censoring.

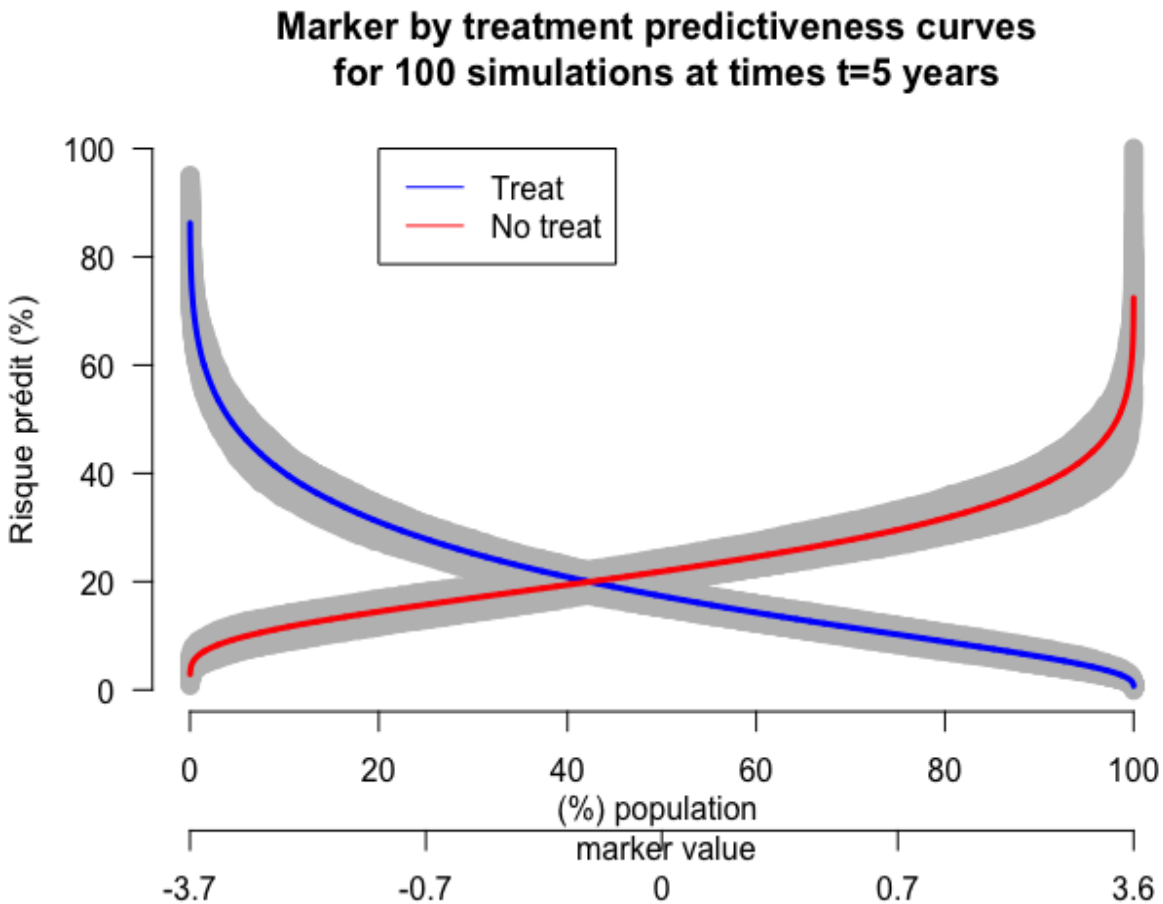
$$\begin{aligned} \pi_i(t_j) &= \mathrm{P}(V_i(t_j) = 1 | Treat, Y) \\ &= \begin{cases} \mathrm{P}(C_i > \widetilde{T}_i | Treat, Y, \widetilde{T}_i) & \text{if } \widetilde{T}_i \leq t_j \text{ and } \delta_i = 1 \\ \mathrm{P}(C_i > t_j | Treat, Y) & \text{if } \widetilde{T}_i > t_j \end{cases} \end{aligned}$$

Time-dependent marker by treatment predictiveness curves



18% of relapses, 46% of survivors and 36% of censorings with a large sample size.

Replication with "perfect" calibration of the predicted risk model



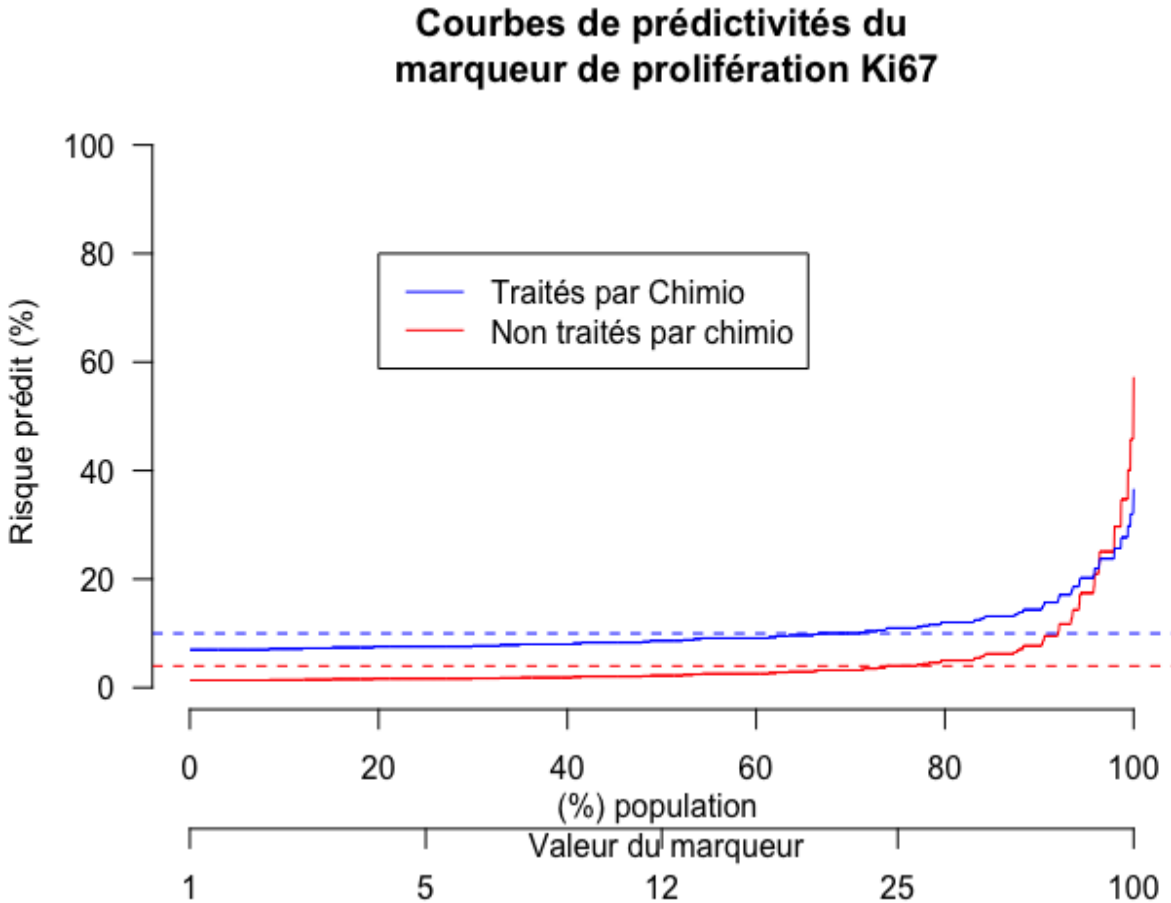
Clinical interest

- The predictiveness curve makes it possible to obtain the cut-off of the marker.
- Determination of sub-group of patient who will benefit or not from the treatment.
- Improve and rationalize the performance of care, avoid unnecessary treatment and improve the quality of life of patient.

As an illustration we consider the case of the proliferation marker Ki67

Illustration : Ki67 as a predictive biomarker ?

Estimating the cut-off for Ki67



Observational study of 3310 women with a non metastatic breast cancer followed from 2000 to 2008 at Institut Curie. The Time-to-event endpoint is Distant Metastasis.

Perspectives

- Predictive biomarkers of response to treatment can potentially improve clinical outcomes and lower medical costs.
- The proposed extension rely on a perfect calibration $\rightarrow$ A test of calibration would be of practical interest
- Hopefully, our request for the MINDACT data will be fulfilled ! ! !
- Extend the package R `TreatmentSelection`.

THANK YOU FOR YOUR ATTENTION

References

Janes, H., Brown, M. D., Huang, Y., and Pepe, M. S. (2014). An approach to evaluating and comparing biomarkers for patient treatment selection. *Int J Biostat*, 10(1) :99–121.

Janes, H., Pepe, M. S., Bossuyt, P. M., and Barlow, W. E. (2011). Measuring the performance of markers for guiding treatment decisions. *Ann. Intern. Med.*, 154(4) :253–259.

Zheng, Y., Cai, T., and Feng, Z. (2006). Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics*, 62(1) :279–287.