



Impact de l'infiltrat lymphocytaire sur la  
survie sans récurrence des patientes  
atteintes d'un cancer du sein triple-  
négatif.

*Estelle NEVEU*



# SOMMAIRE

- Introduction
- Objectif
- Matériels et Méthodes
- Résultats
- Discussion

# INTRODUCTION

**Cancer du sein:** le plus fréquent chez la femme

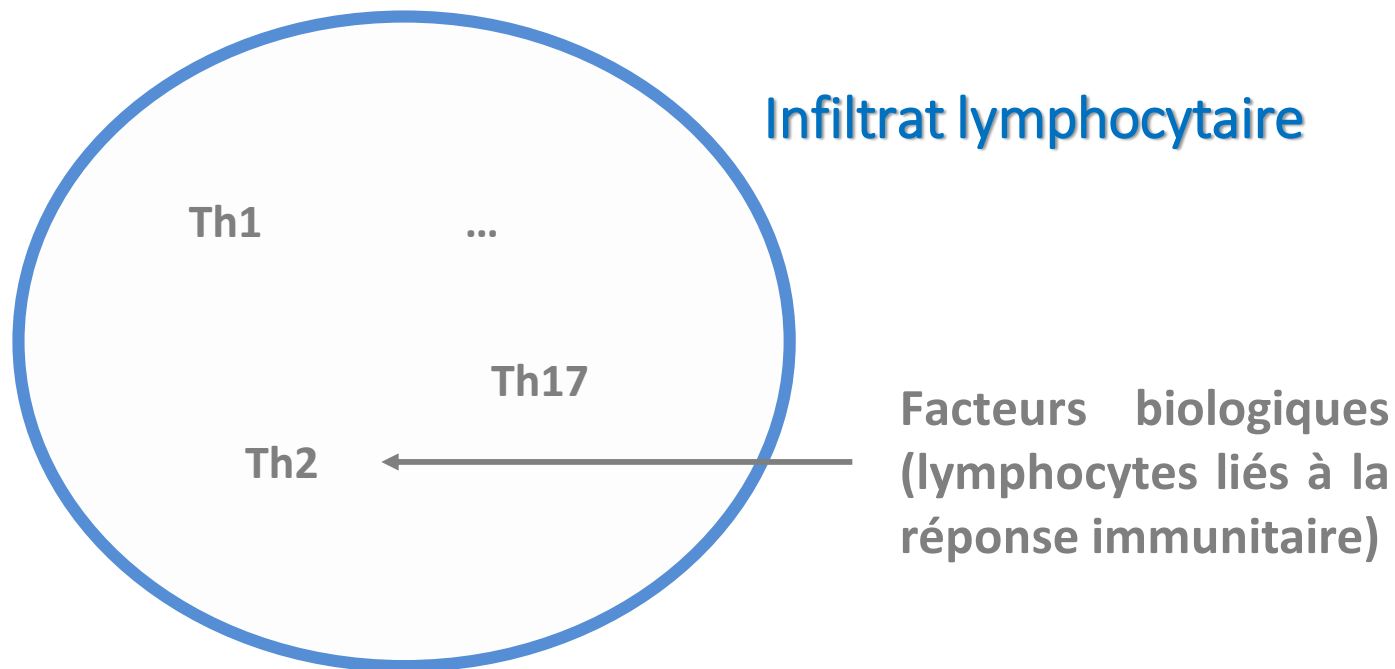
→ > 50 000 nouveaux cas / an en France

---

**Cancer du sein triple-négatif (TNBC):** ~ 15% des cancers du sein

- Absence de récepteurs hormonaux (Œstrogènes et Progestérone) et de récepteurs au facteur de croissance HER2
- Mauvais pronostic en général
- Tumeurs très hétérogènes du point de vue de l'infiltrat lymphocytaire
- Axe de recherche: étudier l'environnement immunitaire des TNBC

Un **lymphocyte\*** est un type de globule blanc présent dans le sang.



# OBJECTIF

- L'objectif principal est d'investiguer différentes méthodes statistiques permettant d'analyser des données de type censurées comportant un grand nombre de variables.
- Au niveau clinique, cette étude cherche à mettre en évidence les facteurs biologiques qui impactent la survie des patientes atteintes d'un cancer du sein triple-négatif.

# MATÉRIELS ET MÉTHODES

Fusion de deux bases:

- **Base clinique** (alimentée par les médecins du Centre Eugène Marquis)
- **Base biologique** (alimentée par l'équipe INSERM U1242-COSS)

*Ces bases sont renseignées de manière prospective à chaque nouvelle tumeur prélevée.*

# MATÉRIELS ET MÉTHODES

## Modèle de Cox multivarié:

- Modèle de Cox ajusté sur les **variables cliniques** (critère d'Akaike, procédure de sélection Stepwise).
- Modèle de Cox ajusté sur les **variables biologiques + 2 variables cliniques** retenues du modèle précédent (même méthode de sélection).

$$h(t, x) = h_0(t) \exp \left( \sum_k \beta_k x_k(t) \right) = h_0(t) \exp(x' \beta)$$

# MATÉRIELS ET MÉTHODES

## Régression LASSO-Cox:

- La régression pénalisée de type LASSO (Least Absolute Shrinkage and Selection Operator) va sélectionner les variables les plus pertinentes en mettant certains coefficients  $\beta$  à zéro.
- Les variables ayant un coefficient  $\beta$  non nul sont ensuite insérées dans un modèle de Cox.



# MATÉRIELS ET MÉTHODES

## Régression LASSO-Cox:

Soit  $L_p(\beta)$  la vraisemblance partielle des paramètres  $\beta$  du modèle:

$$L_p(\beta) = \prod_{i=1}^n \left\{ \frac{\exp(Z_i(\beta))}{\sum_{j=1}^n Y_j(T_i) \exp(Z_j\beta)} \right\}$$

$Z_i$ : covariables

$T_i$  : date de défaillance

$Y_i(t)$  : 1 si l'individu  $i$  est encore à risque au temps  $t$ , 0 sinon.

On a la log-vraisemblance partielle du modèle:

$$l_p(\beta) = \log(L_p(\beta))$$

# MATÉRIELS ET MÉTHODES

## Régression LASSO-Cox:

Dans le cas du modèle de Cox, on estime  $\beta$  à l'aide de ce critère:

$$\widehat{\beta} = \operatorname{argmin} \left( l_p(\beta) \right) + \lambda \sum_j |\beta_j| \quad (1)$$

- Plus la valeur de  $\lambda$  augmente, plus il y aura de coefficients nuls (et donc une grande sélection de variables).
  - Sous R, la fonction *cv.glmnet* choisit la valeur de  $\lambda$  qui minimise la log-vraisemblance partielle par validation croisée (Leave One Out).
- Puis, la minimisation de l'équation (1) se fait à l'aide d'un algorithme quadratique.

# MATÉRIELS ET MÉTHODES

## Random Forest:

- Arbre de classification adapté aux données censurées (package *rpart* sous R).
- L'idée est d'identifier les variables qui discriminent la population concernant le temps de survie sans récurrence.

# MATÉRIELS ET MÉTHODES

## Analyse de sensibilité:

- Rééchantillonnage sur les 3 méthodes: **BOOTSTRAP**

# RÉSULTATS

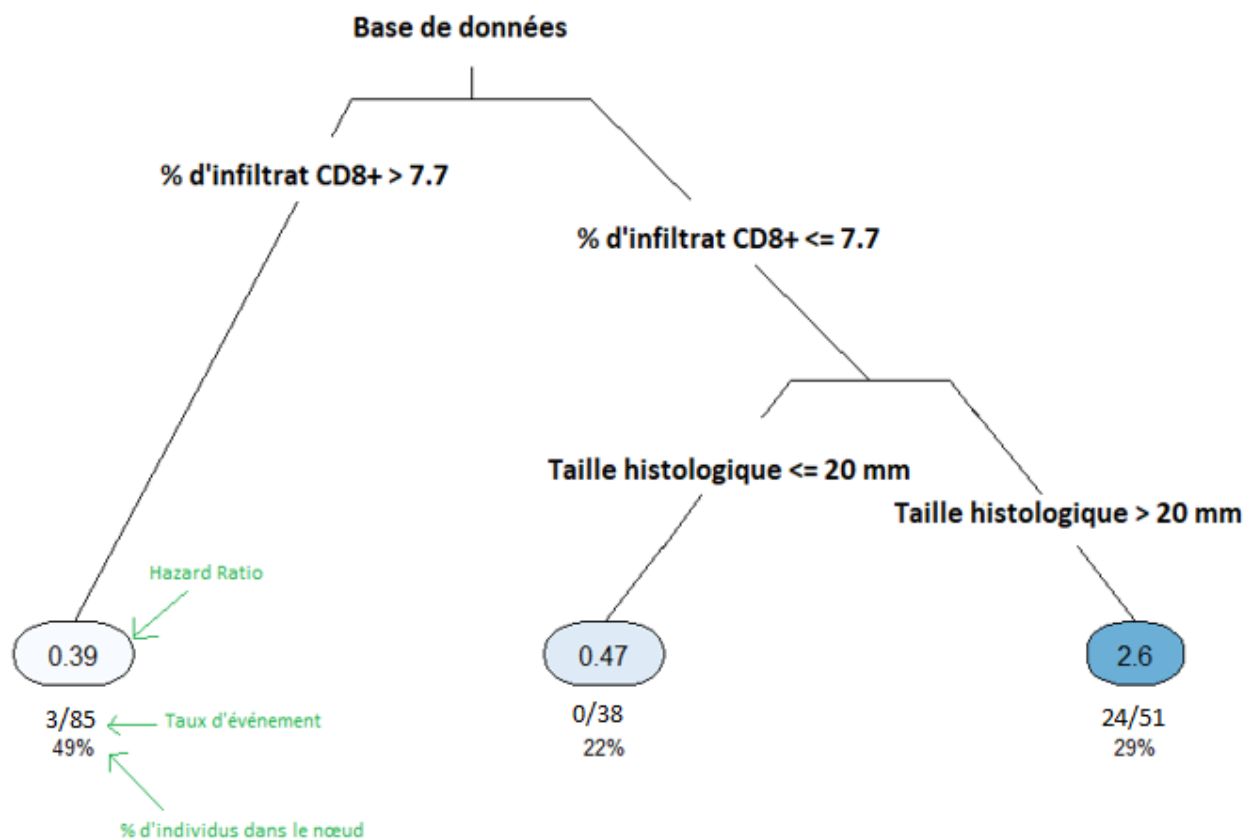
## Régression pénalisée LASSO-Cox:

- Stade  $HR(Stade\ III) = 2.75$
- Taille histologique tumeur  $HR(taille > 20\ mm) = 3.64$
- % d'infiltrat CD8+  $HR(\% > 7.7) = 0.61$
- % Th2  $HR(\% > 8.3) = 0.52$
- % Th1  $HR(\% > 0.4) = 0.70$

# RÉSULTATS

Arbre de classification:

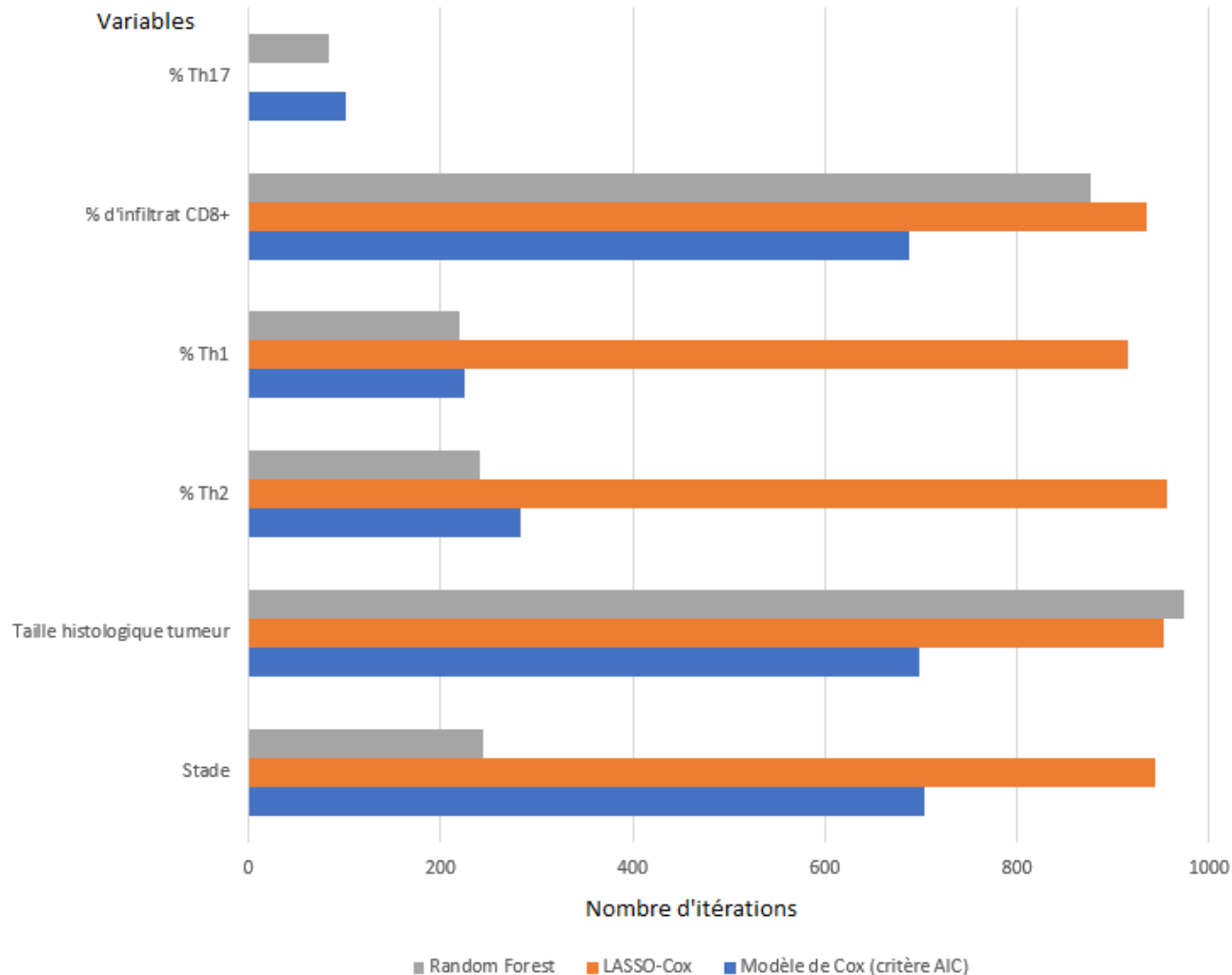
Figure 3: Arbre de classification



# RÉSULTATS

## Analyse de sensibilité - Rééchantillonnage:

Fréquence des variables significatives pour chaque méthode après rééchantillonnage



# DISCUSSION

## Limites

- Taille de la base de données:
  - Règle: 10 événements par variable non respectée
  - *Overfitting, mauvaise précision des coefficients et des tests statistiques sur la significativité*
- Facteurs de confusion non mesurés (Ethnie, ...)
- Critère d'éligibilité: patientes n'ayant pas eu de traitement néo-adjuvant



# DISCUSSION

## Ouverture

- Augmenter la durée de recueil des données/ Étude multicentrique avec les autres CLCC de France.
- Validation des résultats sur une base existante.

Merci pour votre attention

