

# Présélection pour les essais cliniques en oncologie via la science des données : l'étude PreSciOUS

Marie ANSOBORLO<sup>1</sup>, Thibault DHALLUIN<sup>1</sup>, Julien HERBERT<sup>1</sup>, Christine SALPETRIER<sup>1</sup>,  
Christophe GABORIT<sup>1</sup>, Guillaume BOUZILLE<sup>2</sup>, Marc CUGGIA<sup>2</sup>, Leslie GRAMMATICO-GUILLON<sup>1</sup>.

<sup>1</sup>Service d'information médicale, d'épidémiologie et d'économie de la santé (SIMEES), CHU Tours, F-37000 Tours, France.

<sup>2</sup>Université de Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France.

## Contexte

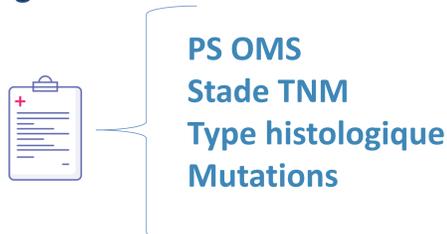
L'incidence des cancers broncho-pulmonaires (CBP) progresse, restant une cause majeure de mortalité.

Détecter les caractéristiques des patients à partir des **comptes rendus de réunions de concertation pluridisciplinaire (CR-RCP)** (1),



## Matériel et Méthode

Des CR-RCP ont été extraits 5 critères d'éligibilité :



Par des algorithmes de TAL :

- Expressions régulières (regEx)
- Apprentissage supervisé (modèles de classifieur bayésien naïf - CBN) (2)

Evalués par les paramètres suivants :

$$\text{Rappel} = \frac{\text{Vrais positifs}}{\text{Faux négatifs} + \text{Vrais positifs}}$$

$$\text{Précision} = \frac{\text{Vrais positifs}}{\text{Faux positifs} + \text{Vrais positifs}}$$

$$\text{F-mesure} = \frac{2 * \text{précision} * \text{rappel}}{\text{précision} + \text{rappel}}$$

## Résultats

- Variation du taux de renseignement des **informations** dans les CR-RCP selon les critères recherchés :
  - PS OMS **systématiquement renseigné**
  - Mutations **majoritairement absentes**

Caractéristiques rares = difficiles à recueillir automatiquement.

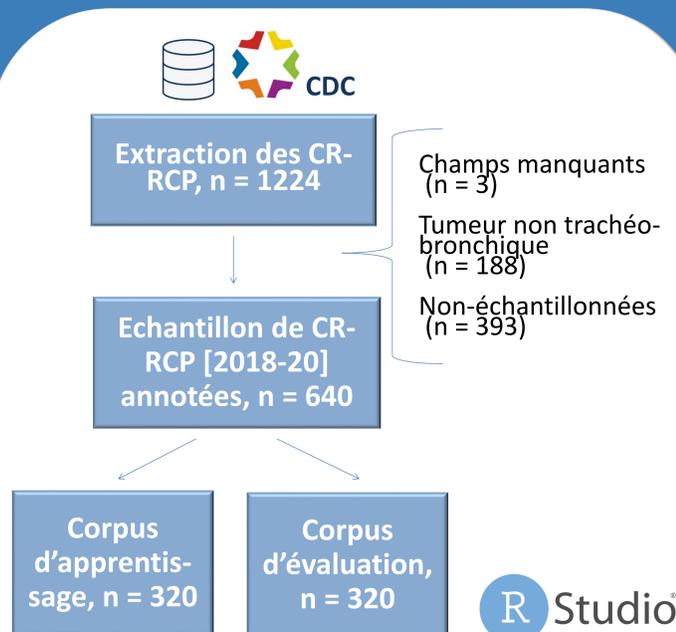


Figure.1 Flow chart des CR-RCP

Table.1 Annotation des CR-RCP selon les informations recherchées

OMS	TNM	Histologie	ALK	EGFR
100 %	55 %	85 %	38 %	33 %

Table.2 Paramètres de performances des techniques de TAL pour extraire les informations

	Rappel	Précision	F-mesure
CBN	64 %	75 %	68 %
RegEx	92 %	98 %	93 %

## Discussion

algorithmes TAL à base de règles = faciles à implémenter + à interpréter

*Au moins aussi performant que des outils évalués dans la littérature (3).*

- Un annotateur unique dans un centre unique restant **insuffisant**, entrainer l'outil sur des données issues d'entrepôts d'autres établissements serait nécessaire pour améliorer la **validité externe**.

### Conclusion :

- Appairer les CR-RCP aux protocoles pourrait **faciliter l'inclusion des patients** dans des essais cliniques en pneumo-oncologie.



## Références

1 - Grundmeier et al. Research subject enrollment by primary care pediatricians using an electronic health record. AMIA Annu Symp Proc. 2007

2 - Nguyen et al. Symbolic rule-based classification of lung cancer stages from free-text pathology reports. J Am Med Inform Assoc. 2010

3 - McCowan et al. Collection of Cancer Stage Data by Classifying Free-text Medical Reports. J Am Med Inform Assoc. 2007

## Contacts

[marie.ansoborlo@etu.univ-tours.fr](mailto:marie.ansoborlo@etu.univ-tours.fr)