

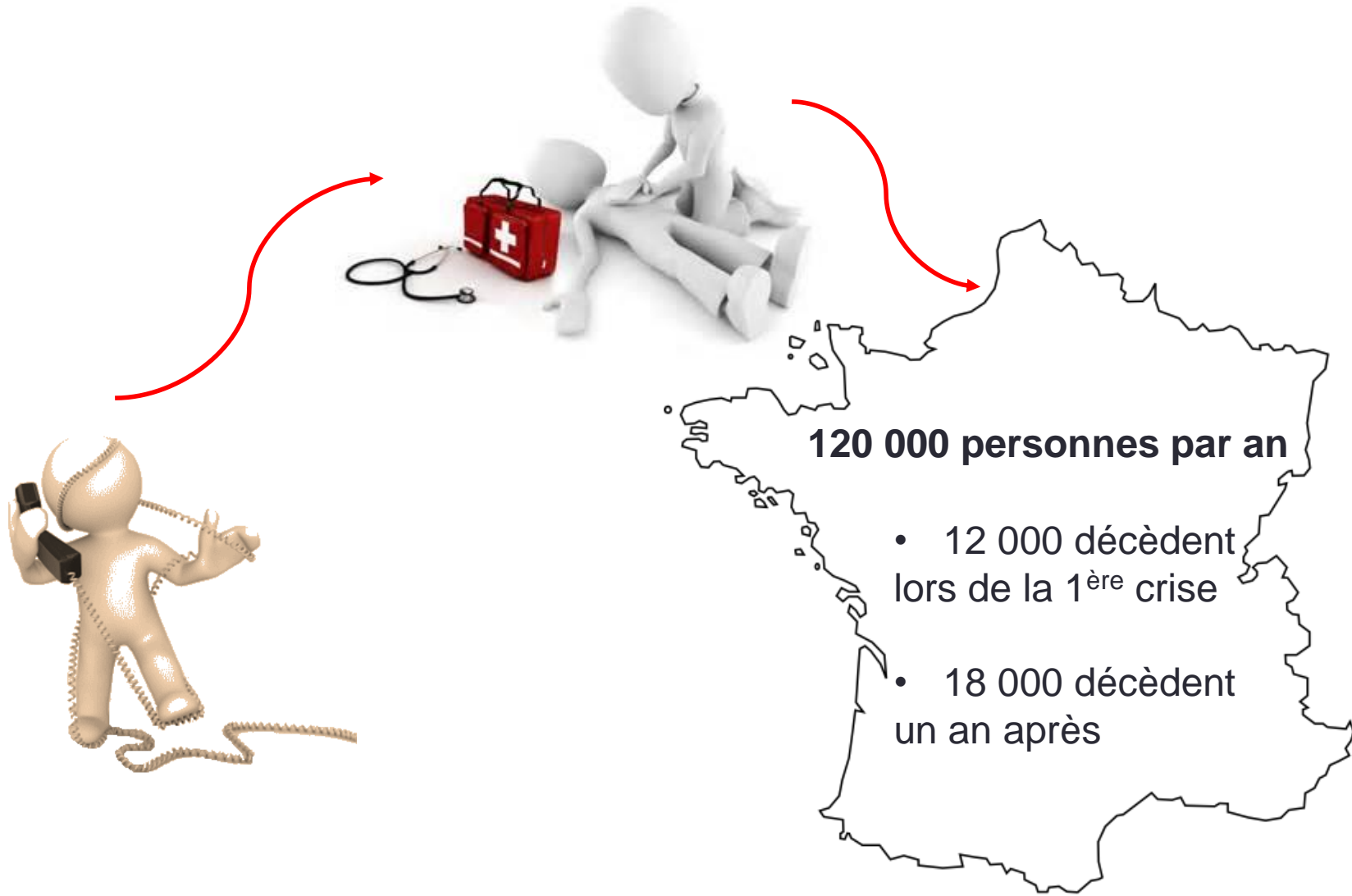
INFARCTUS DU MYOCARDE : QUELLES SONT LES TRAJECTOIRES DE SOINS PRONOSTIQUES DU DÉCÈS À L'HÔPITAL ?

J. Pinaire, J. Azé, S. Bringay, P. Landais

LIRMM, EA 2415 Université de Montpellier & CHU de Nîmes

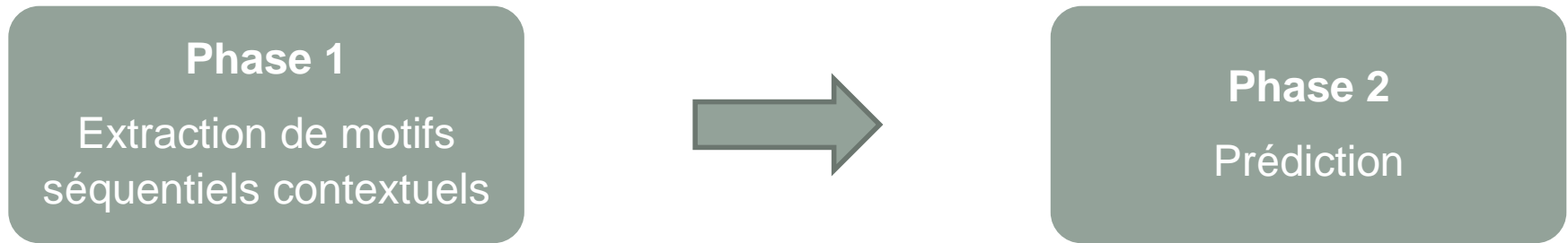
Ingénierie des Connaissances 2017
Caen – 05 juillet 2017

- Contexte et objectifs
- Méthode
- Résultats
- Conclusion et perspectives



Prédire le décès à partir des parcours de soins :

- Déterminer le meilleur modèle
- Identifier les parcours pronostiques du décès



Base nationale
PMSI



Sélection des patients avec
séjours IM

Base nationale

PMSI



Sélection des patients avec
séjours IM

Critères IM

+

De 20 à 99 ans

+

France métropolitaine

+

Du 1^{er} mars 2009 au 31
décembre 2014

=

412 486 patients

Base nationale
PMSI



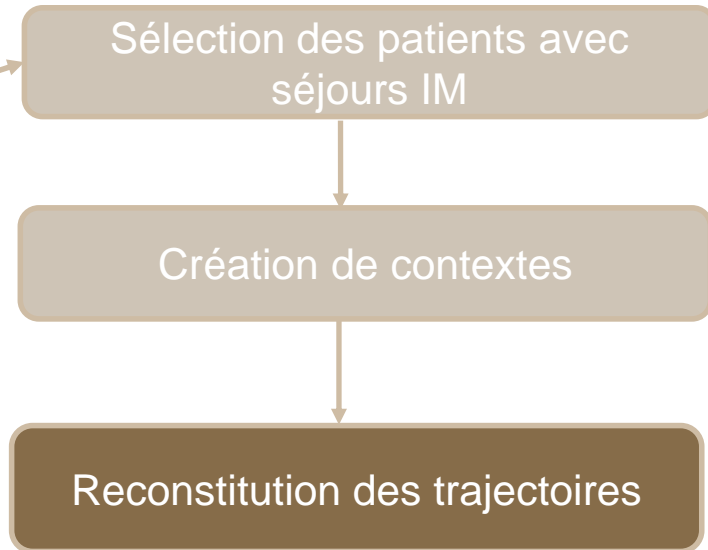
Sélection des patients avec
séjours IM

Création de contextes



Sexe : H - F
Age :
- 45; 45 – 65;
+ 65 ans.
Nb Hospit :
2-5; 5-60;+ 60.

Base nationale
PMSI



Sexe : H - F
Age :
- 45; 45 – 65;
+ 65 ans.
Nb Hospit :
2-5; 5-60;+ 60.

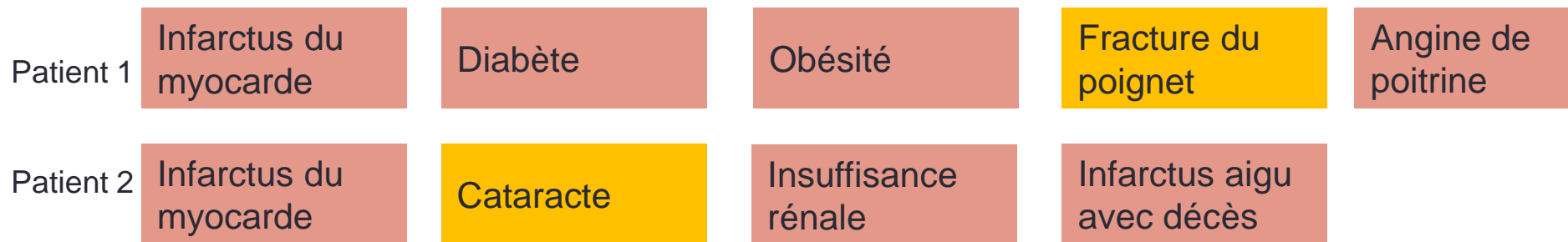
Reconstitution des trajectoires

Légende des différentes catégories de séjours

Séjour lié à la cardiopathologie

Autre type de séjours

I. Reconstitution des séquences des évènements hospitaliers



Tri des évènements

II. Constitution des trajectoires de patients



Base nationale
PMSI




Sélection des patients avec
séjours IM

Création de contextes

Reconstitution des trajectoires

Recherche de motifs séquentiels
contextuels fréquents

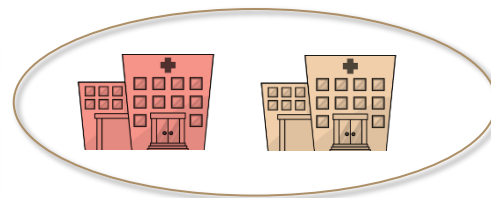
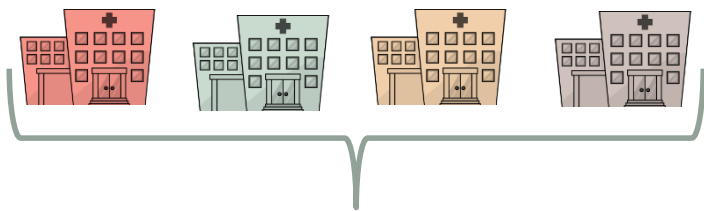
(Rabatel, 2011)



Sexe : H - F
Age :
- 45; 45 – 65;
+ 65 ans.
Nb Hospit :
2-5; 5-60; + 60.

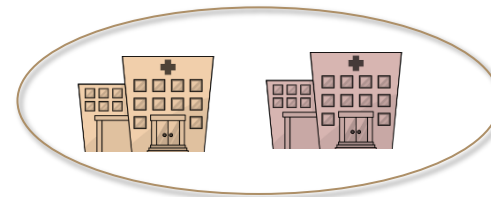
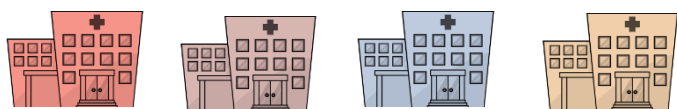


Recherche de motifs séquentiels contextuels fréquents

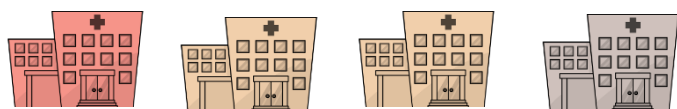
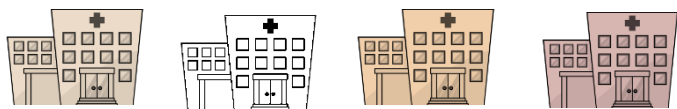
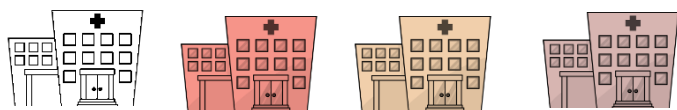


Généralement fréquent : 4/5

Séquence



Fréquent chez les ♀



Base nationale
PMSI



Sélection des patients avec séjours IM

Création de contextes

Reconstitution des trajectoires



Recherche de motifs séquentiels contextuels fréquents

(Rabatel, 2011)

Sélection des motifs maximaux

(Gouda & Zaki, 2005)



Sexe : H - F
Age :
- 45; 45 – 65;
+ 65 ans.
Nb Hospit :
2-5; 5-60; + 60.

A-B-C
A-B



A-B-C

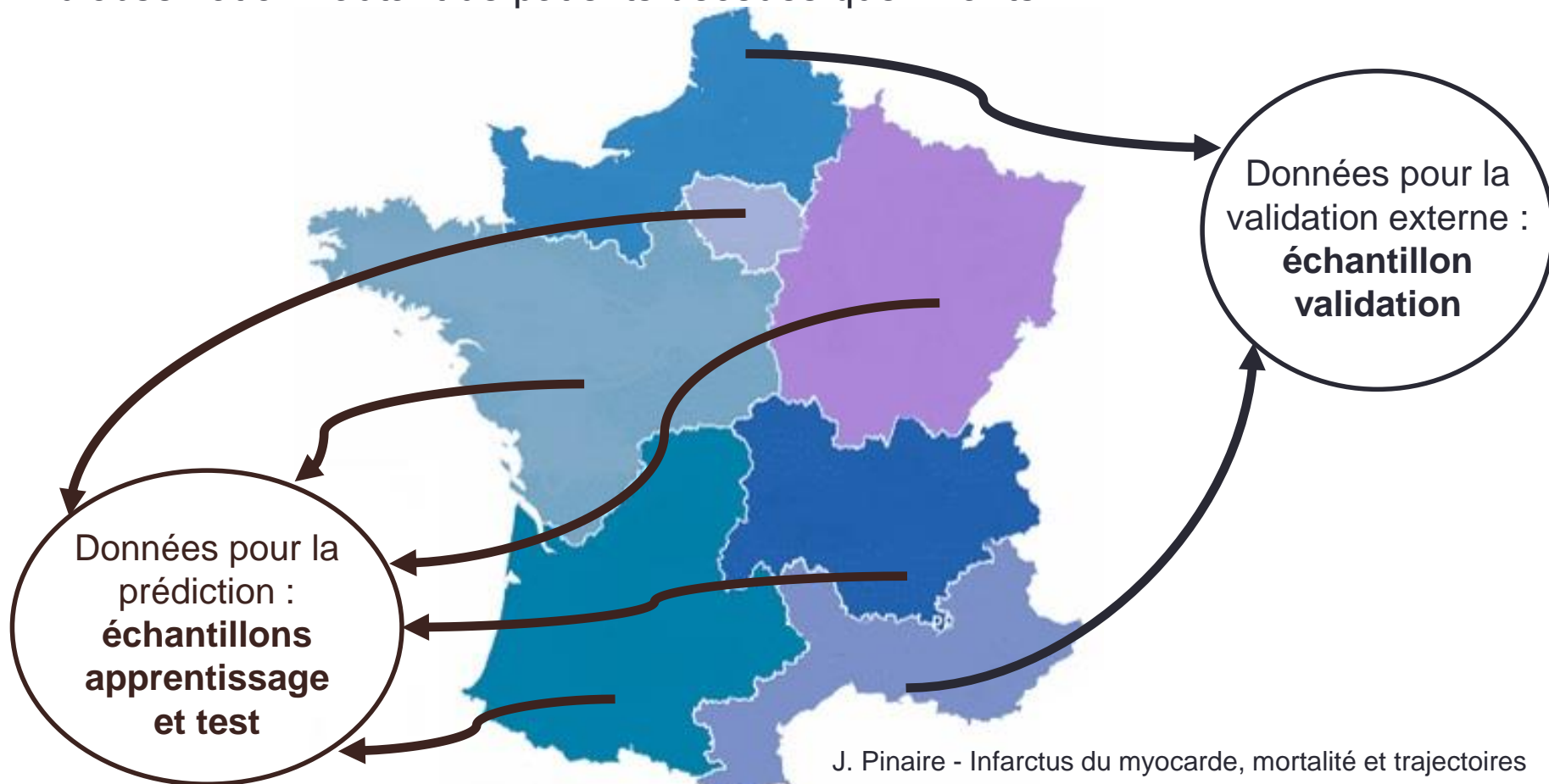


Sélection des données



Sélection des données

- Trajectoires ayant au moins 4 évènements (5 199 patients)
- Echantillons équilibrés selon l'état final du patient à la fin de la période d'observation : autant de patients décédés que vivants





Sélection des données



Calcul des scores similarité

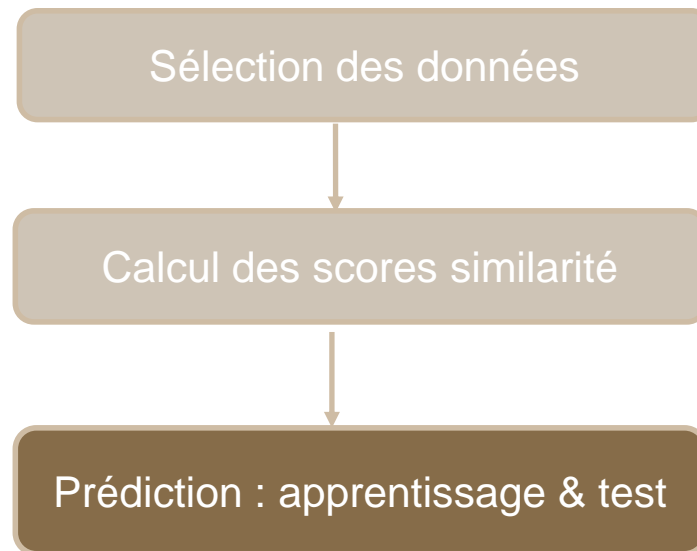
Calcul des scores similarité

Dans un contexte donné, pour chaque patient on calcule la similarité entre chacun des motifs et la trajectoire du patient.

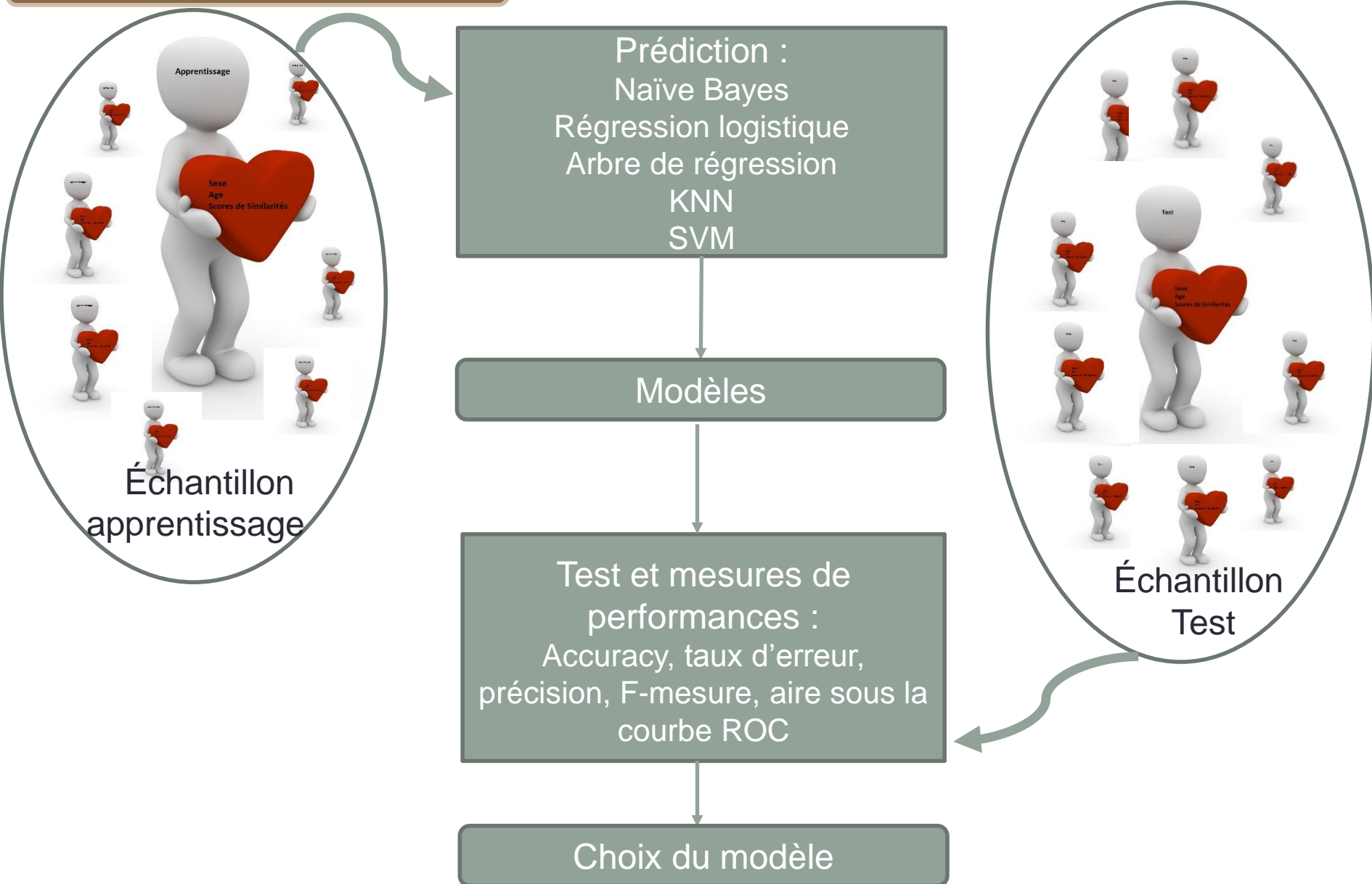
	Motif 1	Motif 2	...	Motif k
TrajPat 1	score			
TrajPat 2				
TrajPat 3				
...				
TrajPat n				

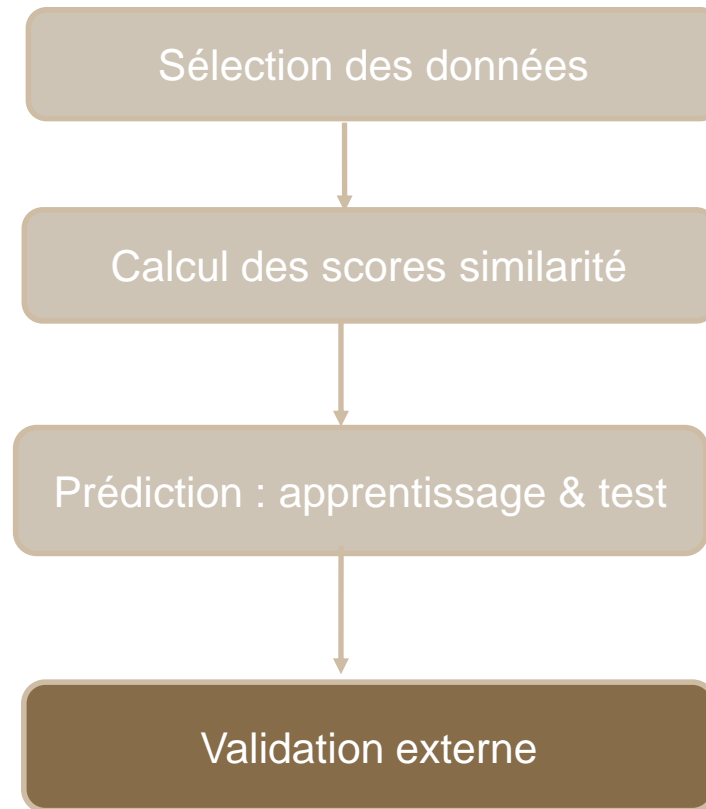
Cette similarité est mesurée par :

- Distances d'édition : LCS, OSA, Levenshtein, Damerau-Levenshtein
- Distances q-gramme, Jaccard, cosinus
- Distances de Jaro et Jaro-Winckler



Prédiction : apprentissage & test





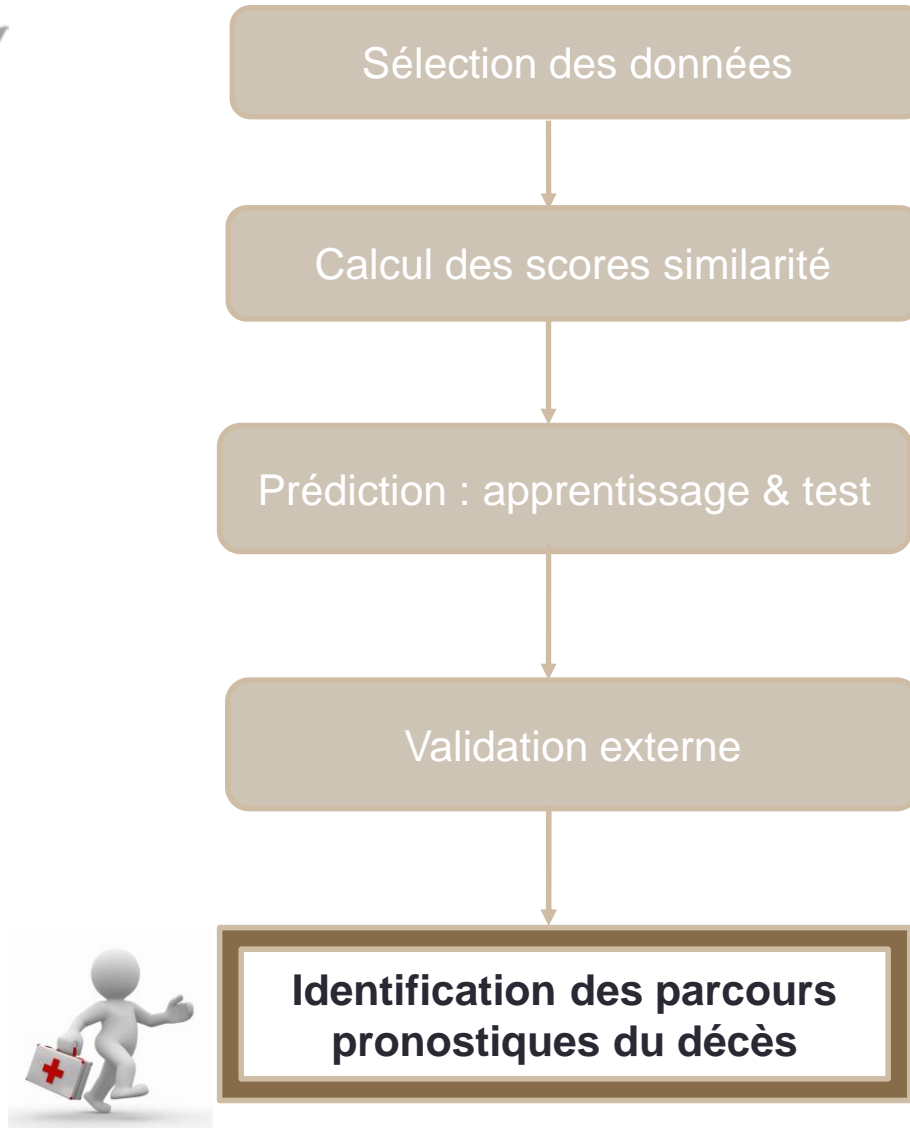
(Collins, 2015)

Validation externe



Modèle sélectionné

Validation et mesures de performances :
aire sous la courbe ROC et
score de Brier



- Choix des modèles :
 - Régression logistique couplée avec une distance d'édition
 - SVM couplée avec une distance q-gramme

- Identification des parcours pronostiques du décès :

- ❖ Contexte 5-60 séjours :

- IM aigu – pose de stent
 - Actes diagnostics par voie vasculaire
- } Facteurs de risque

- Pose de stent – pose de stent – pose de stent
 - Actes thérapeutiques par voie vasculaire
- } Facteurs protecteurs

- ❖ Contexte 3-5 séjours :

- IM aigu
- Actes diagnostics par voie vasculaire
- Actes thérapeutiques par voie vasculaire

- Modélisation du décès :
 - Données : trajectoires des patients
 - Méthode en 2 étapes : extraction de motifs fréquents et prédiction

- Identification de parcours à risque :
 - Evolution de la pathologie
 - Suivi du patient

- Limites de l'étude :
 - Différence de longueur des séquences
 - Qualité des données PMSI

- Amélioration de l'approche proposée :
 - Pondération sur la longueur des séquences
 - Utilisation d'autres mesures (émergence, confiance,...)
- D'autres approches sont possibles :
 - Autres contextes + Classifieur
 - Analyse de concept formelle + Event Calculus
- Système de surveillance à partir des bases PMSI

- Collins G. S., Reitsma J. B., Altman D. G. & Moons K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) : the TRIPOD statement. *BioMed Central Medicine*, 13(1), g7594.
- Gouda K. & Zaki M. J. (2005). GenMax : An Efficient Algorithm for Mining Maximal Frequent Itemsets. *Data Mining and Knowledge Discovery*, 11(3), 223–242.
- Rabatel J. (2011). Extraction de motifs contextuels : Enjeux et applications dans les données séquentielles. PhD thesis, Université Montpellier II.