

La classification non supervisée des données en vie réelle d'une communauté en ligne de patients permet l'identification de profils de patients lupiques concernant leurs préférences thérapeutiques



Damien Testa¹, Valentin Morisseau¹, Lise Radoszycki¹, Laurent Chiche²

¹ Carenity, Communauté de patients en ligne, 1 rue de Stockholm, Paris, France

² Service de Médecine Interne, Hôpital Européen, Marseille, France



INTRODUCTION

L'adhésion d'un patient à son traitement et l'observance qui en découle peuvent varier selon la pathologie, les contraintes du traitement et le profil du patient. Ainsi, **la sélection d'un traitement** en adéquation avec le profil et donc les préférences du patient pourrait permettre une **meilleure observance**.

OBJECTIF

L'objectif de cette étude est d'identifier des profils de patients lupiques concernant leurs préférences thérapeutiques grâce à une analyse des correspondances multiples (ACM) associée à des méthodes de classification non supervisée.

MÉTHODES

Carenity.com est une communauté de patients en ligne comptant plus de 400 000 membres dans 6 pays (EU5 et USA). Elle permet aux patients atteints d'une maladie chronique et leurs proches de partager leurs expériences, s'informer et participer à des études en ligne.

Caractéristiques de l'étude :

- **Critères d'inclusion** : Patient français adulte atteint de lupus
- **Nombre de répondants** : 268
- **Collecte de données** : Août 2018 – avril 2019

1/ Analyse de Correspondances Multiples

Réalisée à partir des 8 variables qualitatives (**Table 1**) afin d'identifier des profils de patients ayant des préférences homogènes en termes de traitement.

| Variable | Catégorie |
|-----------------------------------|---|
| Âge | Forme galénique du traitement actuel |
| Contrôle du lupus | Forme galénique idéale |
| Capacité à prédire une poussée | Critère le plus important (hors efficacité) |
| Participation à un essai clinique | Sexe |
| Nombre de comorbidités | Type de lupus |

■ Variables de préférences de traitement ■ Variables supplémentaires

2/ Méthodes de classification non supervisée

A partir des résultats de l'ACM, 3 méthodes de classification non supervisée ont été utilisées afin d'identifier des groupes homogènes de patients :

- Classification Ascendante Hiérarchique (CAH)
- Méthode des Kmeans
- Algorithme PAM (k-medoids)

Le modèle de mélange (Mix) qui se base sur l'analyse de la distribution de probabilités des variables a aussi été utilisé.

3/ Sélection de la meilleure méthode

3 indicateurs qui se basent sur l'ACM ont permis de déterminer la méthode et le nombre de clusters les plus performants :

- Mesure de connectivité (à minimiser)
- Indice de Dunn (à maximiser)
- Indice de Silhouette (proche de 1)

4/ Réalisation de la CAH à 3 clusters

Méthode qui sépare tous les individus avant de regrouper en classes les plus similaires à chaque itération jusqu'à reconstituer une classe avec tous les individus.

5/ Comparaison des modèles

Comparaison des classifications obtenues avec chaque méthode à l'aide d'indicateurs de similarité :

- Index de Rand
- Pourcentages d'individus classés dans un cluster similaire

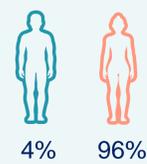
Une cohorte indépendante (n = 92) de patients non français (EU4 et USA) atteints de lupus a été utilisée comme validation externe.

CONCLUSION

Cette étude permet de démontrer l'intérêt des données en vie réelle pour identifier des profils de patients lupiques ayant des préférences similaires en termes de traitement. La personnalisation des soins pourrait favoriser l'adhésion du patient à son traitement ainsi que son observance.

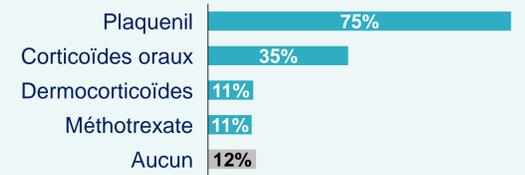
PROFIL DES REpondANTS (n = 268)

Genre



Moyenne d'âge
44,3 ans

Traitement actuel



Oral **86%** Injectable **14%**

Type de lupus

Selon la classification de l'American College of Rheumatology



RÉSULTATS

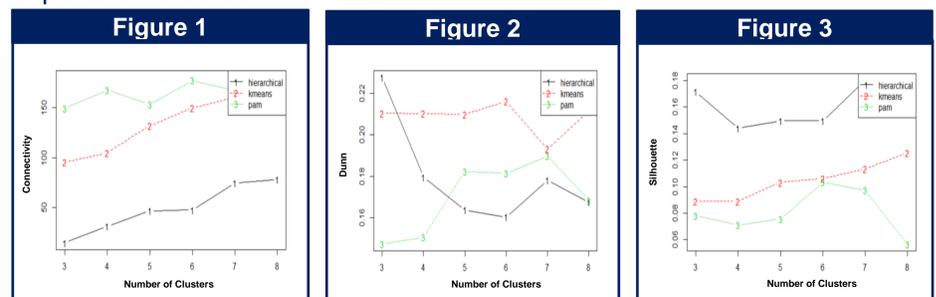
1/ Identification de 3 clusters

L'ACM a été réalisée sur une cohorte de 268 patients français atteints de lupus qui avaient répondu à l'enquête en ligne et a permis de mettre en évidence trois groupes distincts.

2/ Sélection de la méthode CAH avec 3 clusters

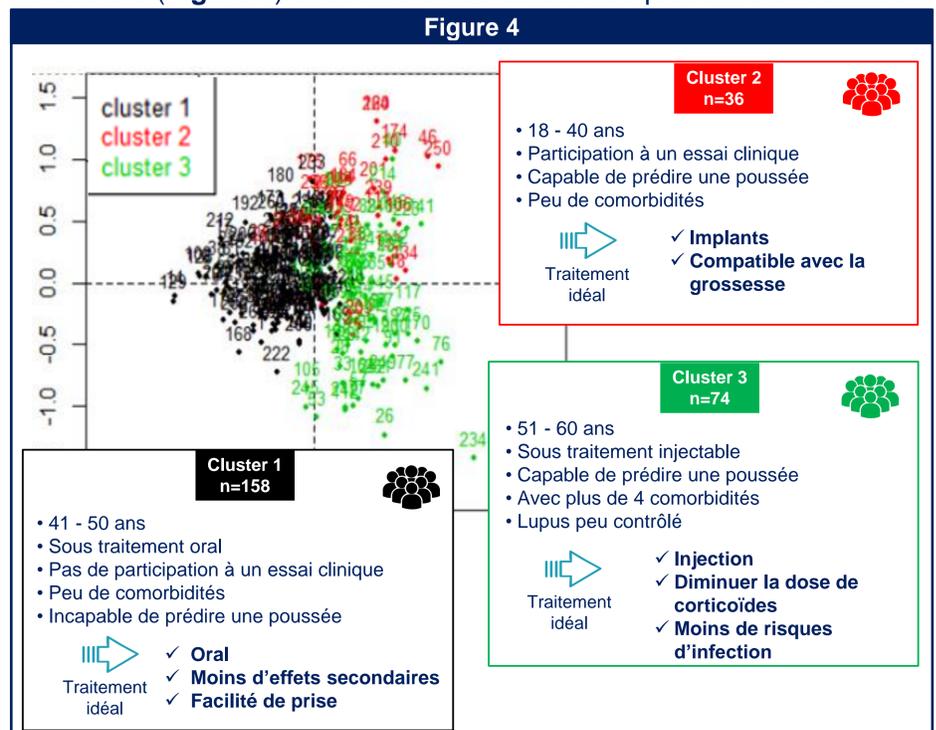
La CAH avec 3 clusters a été sélectionnée car :

- elle minimise la mesure de connectivité (**Figure 1**)
- elle maximise l'indice de Dunn (**Figure 2**)
- elle possède l'indice de Silhouette le plus proche de 1 (**Figure 3**)
- la perte d'inertie confirme le choix des 3 clusters



3/ Analyses des 3 profils de patients identifiés par la CAH

Trois clusters (**Figure 4**) avec différentes caractéristiques ont été identifiés.



4/ Comparaison des 4 modèles de classification

Les différentes méthodes de classification regroupent les patients de l'échantillon dans des clusters similaires (**Figure 5**). Ces résultats renforcent la pertinence de l'analyse des clusters obtenus avec la CAH. La robustesse de ces résultats a été confirmée par des analyses de validation et de sensibilité effectuées sur une cohorte de validation indépendante de patients non français (EU4 et USA) atteints de lupus.

| | Index de Rand | | | | % d'individus classés dans un cluster similaire | | | |
|--------|---------------|--------|------|------|---|--------|------|------|
| | HAC | Kmeans | Mix | PAM | HAC | Kmeans | Mix | PAM |
| HAC | 1 | 0,97 | 0,73 | 0,54 | 100% | 98% | 66% | 42% |
| Kmeans | | 1 | 0,74 | 0,54 | | 100% | 67% | 41% |
| Mix | | | 1 | 0,56 | | | 100% | 46% |
| PAM | | | | 1 | | | | 100% |