

Préférences des patients vis-à-vis des traitements : comment identifier des profils de patients à partir de données en vie réelle ? Application au cas du lupus.

D. Testa^{1*}, P. Legendre¹, L. Radoszycki¹, R. Gauchoux¹

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

* Auteur correspondant, damien@carecity.com (Damien Testa)

CONTEXTE ET OBJECTIFS

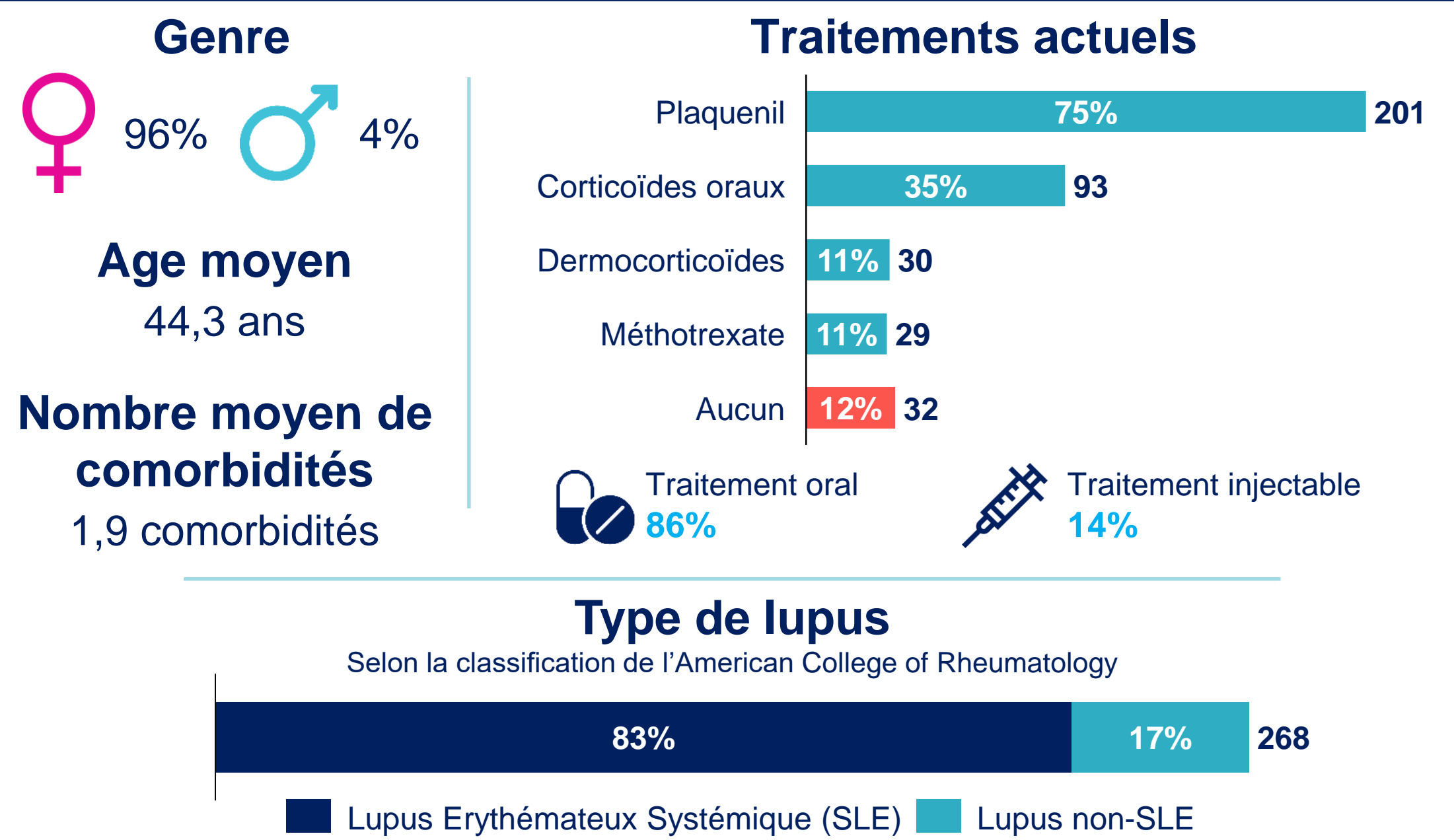
Contexte :

L'adhésion d'un patient à son traitement et l'observance qui en découle peut 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.

Objectifs :

- **Déterminer les attentes des patients** sur la forme galénique et les caractéristiques du traitement.
- **Identifier des profils de patients** partageant des attentes similaires grâce à une Analyse des Correspondances Multiples (ACM).
- **Affiner ces profils et les comparer** avec d'autres techniques de classification non supervisée.

PROFIL DES REpondANTS (n = 268)



METHODE

Carecity.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 adulte atteint de lupus résidant en France
- **Nombre de répondants** : 268
- **Collecte des données** : août 2018 - avril 2019

1/ Analyse des Correspondances Multiples (ACM) :

Réalisée à partir des 8 variables qualitatives afin d'identifier des profils de patients ayant des préférences homogènes en termes de traitement :

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 (bleu) Variables supplémentaires (jaune)

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.

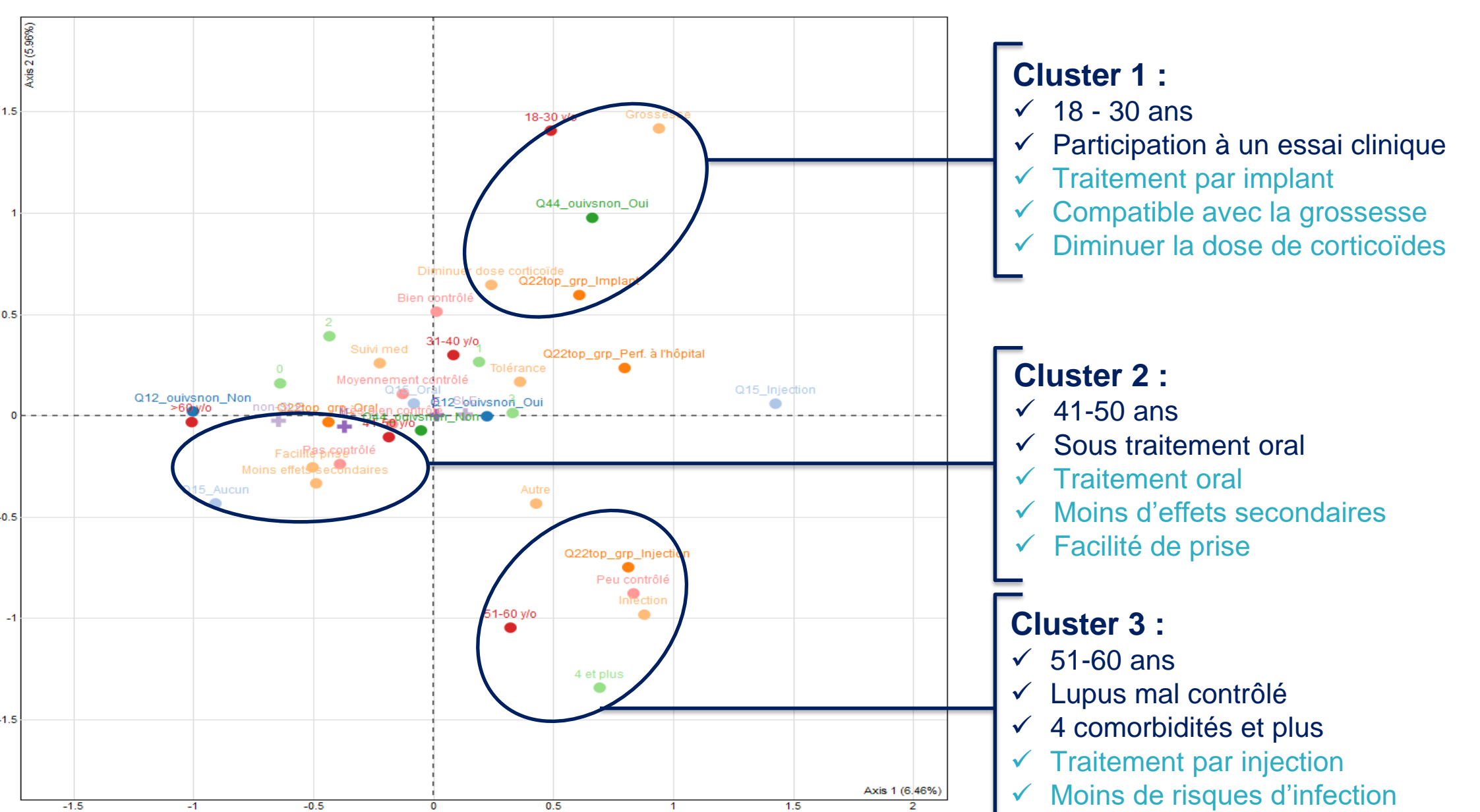
5/ Comparaison des modèles :

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

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

RESULTATS

1/ Identification de 3 clusters :



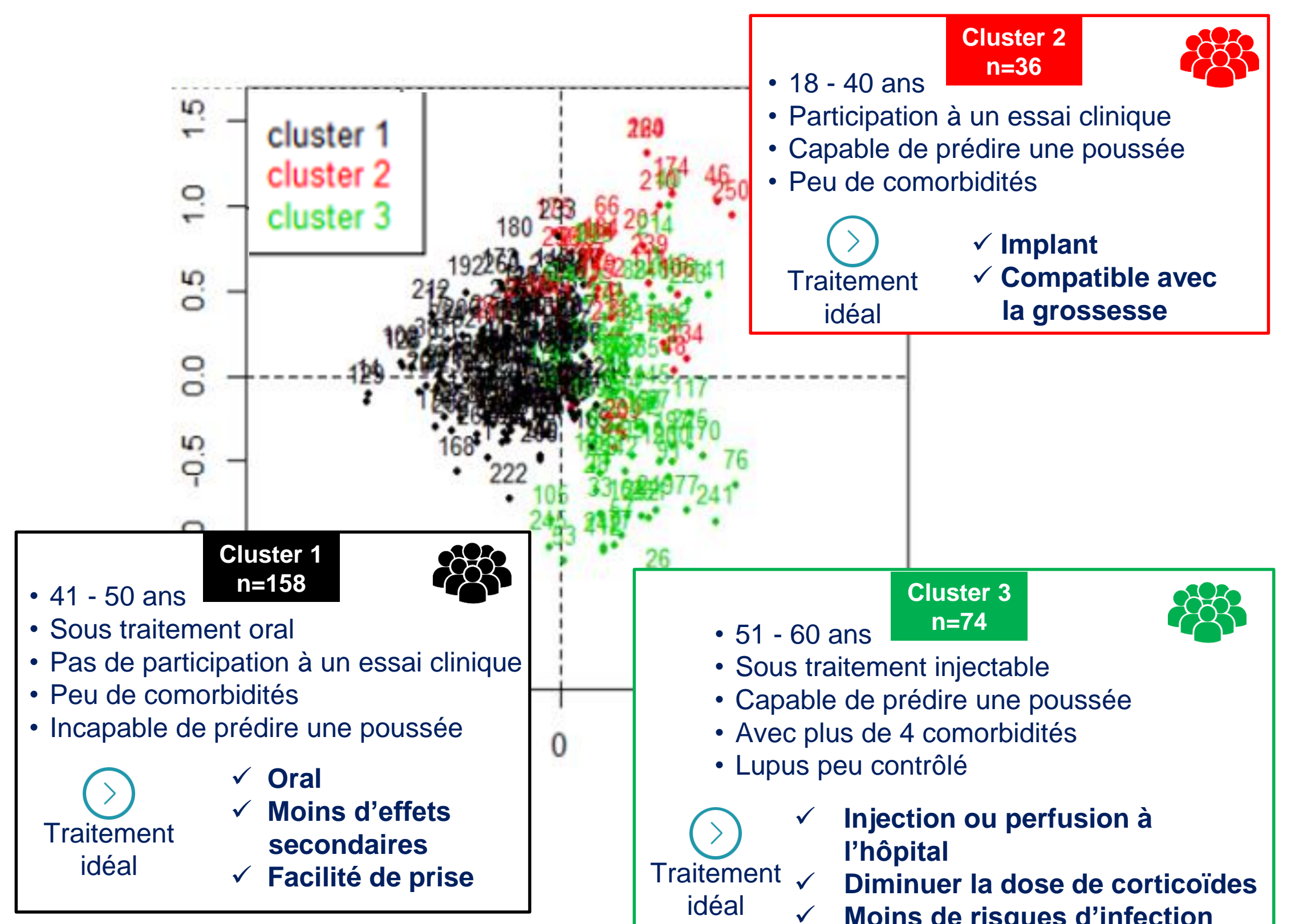
3/ Sélection de la méthode CAH avec 3 cluster :

- La mesure de connectivité sélectionne la CAH avec 3 clusters
- L'indice de Dunn sélectionne la CAH avec 3 ou 7 clusters
- L'indice de Silhouette sélectionne la CAH avec 3 clusters

La CAH avec 3 clusters est la méthode sélectionnée car :

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

4/ Analyses des profils de patients de la CAH avec 3 clusters :



5/ Comparaison des modèles et du modèle de mélange :

	Indice de Rand				% d'individus classés dans un cluster similaire			
	CAH	Kmeans	Mix	PAM	CAH	Kmeans	Mix	PAM
CAH	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%

Les différentes méthodes de classification regroupent les patients de l'échantillon dans des clusters similaires. Ces résultats renforcent la pertinence de l'analyse des clusters obtenus avec la CAH.

CONCLUSION

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