Patient treatment preferences: How to identify patient profiles directly from online real-life data? Application to lupus.

Testa B1, Morisseau V1, Radoszycki L1, Fidyk C1, Chiche L2

1Carnety, online patient community, 1 rue de Stockholm, Paris, France
2Internal Medicine Department, Hôpital Européen, Marseille, France

INTRODUCTION
A patient’s view on and subsequent compliance may vary depending on the pathology, treatment constraints, and patient profile. Thus, selecting a treatment that best fits a patient’s profile and preferences could encourage better compliance.

OBJECTIVE
The objective was to identify patient profiles sharing similar expectations through a Multiple Correspondence Analysis (MCA) associated with unsupervised clustering methods.

METHODS
Carnety.com is an online patient community for people with chronic conditions: it has more than 500,000 members in 6 countries in Western Europe and the United States. It allows patients and their caregivers to share experiences, learn, and participate in online studies.

Study characteristics
- Inclusion criteria: adult lupus patients living in France
- Sample size: 268
- Data collection: between August 2018 and April 2019

1/Multiple Correspondence Analysis
MCA was performed using 8 feature variables (Table 1) in order to identify profiles of patients with homogeneous treatment preferences.

2/Unsupervised classification methods
Based on the results of the MCA, three unsupervised classification methods were used to identify homogeneous groups of patients.
- Hierarchical clustering (HAC)
- Kmeans method
- PAM algorithm (k-medoids)

Mixture model (Mix), which is based on the analysis of the probability distribution of the variables, was also investigated.

3/Selection of the most appropriate method
The most efficient method and number of clusters were chosen considering the three following indicators:
- Connectivity measurement (to be minimized)
- Dunn’s Index (to be maximized)
- Silhouette Index (close to 1)

4/Three clusters HAC implementation
A method that separates all the individuals before grouping them into the most similar classes at each iteration until a class is reconstituted with all the individuals.

5/Model comparisons
The classifications obtained with each method were compared using indicators of similarity:
- Rand Index
- Percentages of identical clustering of individuals

An independent cohort (n=92) of non-French lupus patients (from Europe and the US) was used as external validation.

CONCLUSION
This study demonstrates the value of real-life data directly generated through online patient communities to identify patient profiles with similar treatment preferences. Personalization of care based on patient profiles could help to improve compliance.

RESPONDENT PROFILE (n=268)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Current treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plaqueless</td>
<td>75%</td>
</tr>
<tr>
<td>Oral corticoids</td>
<td>35%</td>
</tr>
<tr>
<td>Dermocorticoids</td>
<td>81%</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>21%</td>
</tr>
<tr>
<td>Injectable treatment</td>
<td>14%</td>
</tr>
</tbody>
</table>

Lupus type
According to the American College of Rheumatology classification
- Systemic Lupus Erythematosus (SLE) 93%
- Non-SLE Lupus 17%

RESULTS

1/Identification of three clusters
MCA was performed on a training cohort of 268 French lupus patients who had answered the online survey and three clusters were highlighted.

2/Selection of the HAC method with 3 clusters
HAC with 3 clusters is the most efficient method because:
- It minimizes the measure of connectivity (1 - black Figure 1)
- It maximizes the Dunn Index (1 - black Figure 2)
- It has the Silhouette Index closest to 1 (1 - black Figure 3)

3/Analyses of the three patient profiles identified by HAC
Three clusters (Figure 4) with distinct characteristics are identified.

4/Comparison of the four classification models
The different classification models group patients into similar clusters (Figure 5). These results reinforce the relevance of the cluster analysis performed with HAC. The robustness of these results were confirmed by validation and sensitivity analyses performed on an independent validation cohort of non-French lupus patients (from Europe and the US).

CONCLUSION
This study demonstrates the value of real-life data directly generated through online patient communities to identify patient profiles with similar treatment preferences. Personalization of care based on patient profiles could help to improve compliance.