Identification of clusters of patients with similar baseline characteristics in the FOCAL-MS2 study evaluating patient's autonomy

Roche

Ambition Autonomie

Cécile Donzé (1); Géraud Paillot (2); Claude Mekies (3); Mikael Cohen (4); Lucie Brechenmacher (5); Alexandre Civet (5); David Pau (5); Svetlana Le Ralle (5); Amélie Lambert (5); Catherine Mouzawak (6); Patrick Vermersch (7)

(1) Hôpital saint Philibert, Groupement des Hôpitaux de l'Institut Catholique de Lille, Faculté de médecine et de maïeutique de Lille, Lomme, France ; (2) Association Aventure Hustive, Saint-Malo, France ; (3) RAMSAY Clinique des Cèdres, Neurologie, Toulouse, France; (4) Université Nice Cote d'Azur, UR2CA-URRIS CRCSEP CHU Nice Pasteur, Service de Neurologie, Nice, 06002, France ; (5) Affaires médicales, Roche, Boulogne-Billancourt, France ; (6) Structure régionale neuro SEP SYNAPSE, Le Vésinet, France ; (7) Univ. Lille, INSERM UMR1172 LilNCog, CHU Lille, FHU Precise, Lille, France

Background/Introduction

Multiple sclerosis (MS) is an inflammatory and degenerative demyelinating disease of the human central nervous system. Autonomy is recognized as one of the key contributor of health according to the WHO's (World Health Organisation) definition (1), but there currently is a lack of validated MS-specific tools to assess patient autonomy.

Objectives

The Multiple Sclerosis Autonomy Scale (MSAS) questionnaire has been developed in collaboration with MS patients and healthcare professionals. This exploratory research focuses on identifying clusters of patients with similar characteristics at inclusion in the study.

Methods

This French longitudinal prospective observational study included MS patients from January 2024 to May 2024 in 33 sites.
The MSAS questionnaire containing 10 dimensions in a 35-items short form (1 item of importance accorded to each dimension and 25 items dispatched within all dimensions) has to be completed by patients at inclusion, D15, D30 and D365.

Clustering algorithm:

From the Qluster algorithm (2), the following steps were applied:
Partitioning Around Medoids (PAM) clustering method was used.
Number of clusters Selection/validation: K-Means inertia (Elbow Method), maximizing silhouette Coefficient and Calinski-Harabasz index while minimizing the Davies-Bouldin index.
Stability assessment: bootstrap method (random of 50 samples with replacement and noise added in each sample) and Jaccard index.
Descriptive statistics were computed from variables included or not in the clustering step. Missing data were not imputed.

Results

From the 210 patients included in the study from January 2024 to May 2024, 199 completed the MSAS questionnaire at baseline: 132 (66.3%) with relapsing remitting form of MS (RRMS), 23 (11.5%) with primary progressive (PPMS) and 44 (22.1%) with secondary progressive (SPMS).

Patient's characteristics:
74.4% of patients were women, mean age was 49.7±11.5 years, 57 (28.6%) of patients were living alone, 76 (38.2%) were active, and 86 (43.2%) required assistance for walking.

Clusters identification:
6 clusters were identified: silhouette mean score of 0.29, Calinski-Harabasz Index was 61.64 and the Davies-Bouldin Index was 1.22 (Figure 1). For the bootstrap approach, clusters all had a Jaccard similarity statistic >90% over 50 iterations, which characterizes high stability to resampling with replacement (Figure 2).

Clusters description:
C1(n=30): Patients aged [50-60] years, on sick leave, with a carer and in a wheelchair, with MS duration between [20-25] years.
C2(n=42): Male patients aged [40-50], requiring occasional walking assistance, with MS for [10-15] years.
C3(n=42): Patients aged [50-60] with personal activities and requiring occasional walking assistance, MS duration between [15- 20] years.
C4(n=30): Retired patients with caregiver and walking aid at all times. Patients with MS duration >25 years.
C5(n=19): Patients under 40 years old without activities without caregiver, without walking aid and MS duration < 10 years.
C6(n=36): Professionally active women under 40 years without caregiver, without walking aid and MS duration <;10 years.
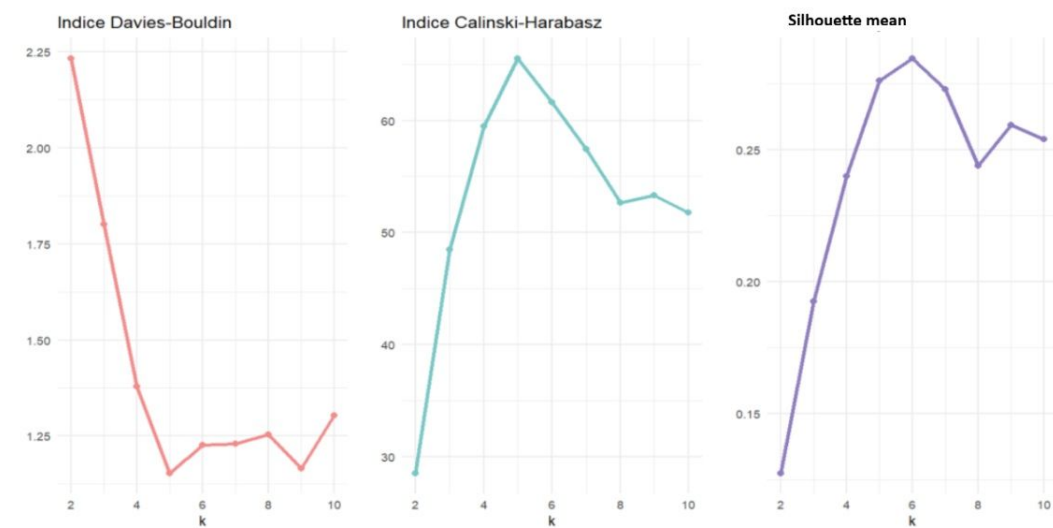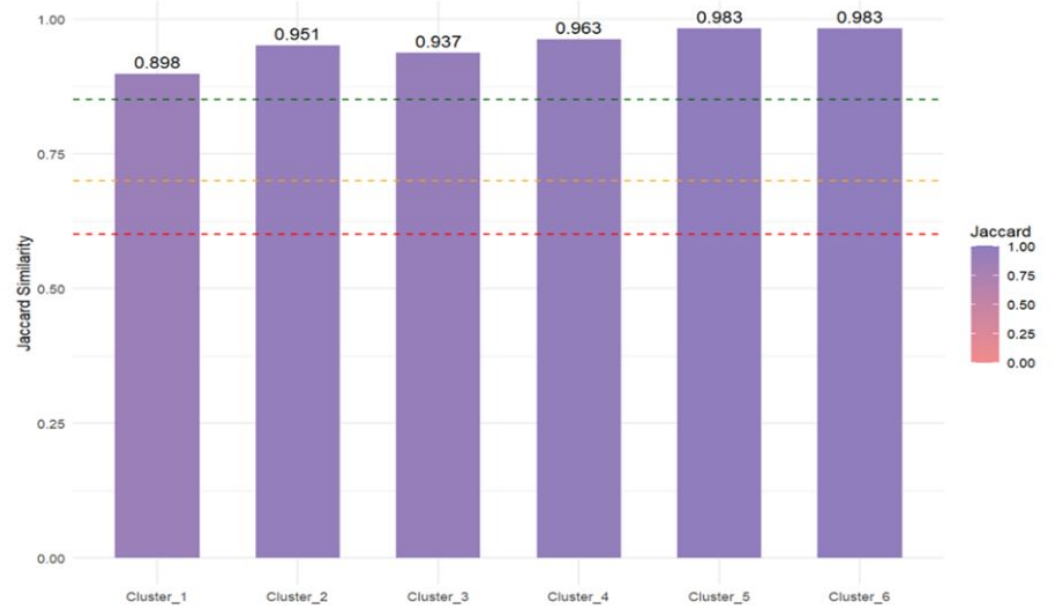
Figure 1: Average silhouette width

Figure 2: Clusters stability evaluated by the Jaccard index

Conclusion

An optimal number of six clusters was identified for the study. The Qluster workflow, by integrating proven methodologies and prioritizing simplicity of implementation and robustness, offers a practical and reliable solution for data scientists in the healthcare domain. This case study underscores its utility for preliminary data mining and the discovery of clinically relevant patient profiles.

References:
(1)   Otorkpa D. World Health Organization(WHO) Definition Of Health - Public Health 2022
(2)   Esnault C, Rollot M, Guilmin P, Zucker JD. "Qluster: An easy-to-implement generic workflow for robust clustering of health data." Front Artif Intell. 2023 Feb 6;5:1055294. doi: 10.3389/frai.2022.1055294. PMID: 36814808; PMCID: PMC9939832