A Machine Learning Algorithm to define disease severity in Chronic Inflammatory Demyelinating Polyneuropathy (CIDP)

MSR68

L. Celico[1], **A. Alves**[1], **D. Karletsos**[1], **S. Iannazzo**[2], **C. Arvin-Berod**[2], **M. De Francesco**[1]
[1]HEOR Value Hub, Brussels, Belgium; [2]ArgenX BVBA, Ghent, Belgium

INTRODUCTION AND OBJECTIVES

- Chronic Inflammatory Demyelinating Polyneuropathy (CIDP) is a rare, serious autoimmune disease which causes demyelination and axonal damage of peripheral nerves. Efgartigimod is the first and only FDA-approved neonatal Fc receptor (FcRn) blocker indicated for the treatment of CIDP.
- A Cost-Effectiveness model (CEM) is required to compare efgartigimod with Standard of Care in CIDP. To this scope, a de novo state-transition CEM is being conceptualized. This analysis aims at identifying clinically relevant health states that form the basis for the state-transition model. Each state should represent a homogeneous health condition, and the specification of states should reflect the biological/theoretical understanding of CIDP.
- To identify the health states, a machine learning (ML) algorithm was developed that classifies CIDP severity based on the functional limitations resulting from the disease.

METHODS

Overview of the analysis

- The analysis was based on data from the ADHERE trial, a global randomized clinical study that investigated the efficacy and safety of efgartigimod on 322 CIDP patients[1].
- The analysis was based on the four steps showed in Table 1 below.

Table 1 – Analysis steps

Step	Description
Step 1 – Select an appropriate measure of functional disability	A measure of functional disability that best captures the functional impairments associated with the disease was selected among the scales used in the ADHERE trial.
Step – 2 Define the ML algorithm	An ad-hoc ML algorithm was created that defines clinically relevant health states based on different levels of functional impairment in CIDP.
Step 3 – Apply the ML algorithm	The ML algorithm was applied to the ADHERE data.
Step 4 – Test the ML algorithm results for statistical significance	The resulting health states were assessed using ANOVA to test whether the differences in disease severity were statistically significant.

Step 1 – Select an appropriate measure of functional disability

The aINCAT score was preferred over other measures available from ADHERE (I-RODS score, MRC Sum Score, TUG and MGS) for the following reasons:

- It is commonly reported in the literature as a measure of functional impairment and presented as an endpoint to the regulatory bodies;
- It is not based on patient-reported outcomes, which ensures a higher degree of objectivity;
- It captures both sensory and motor impairments.

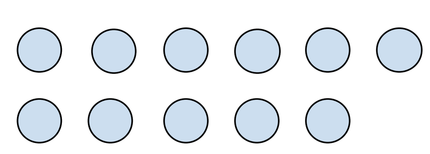
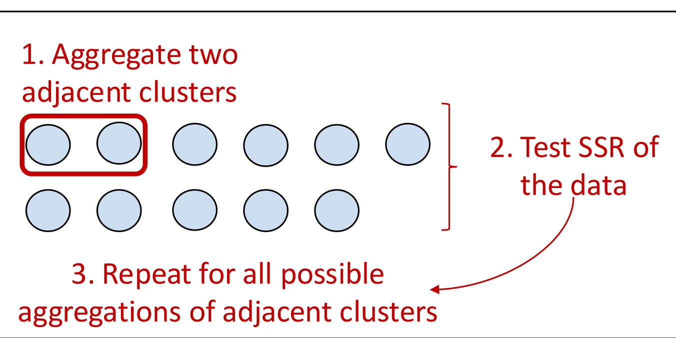
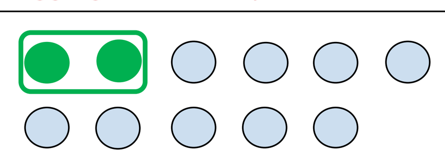
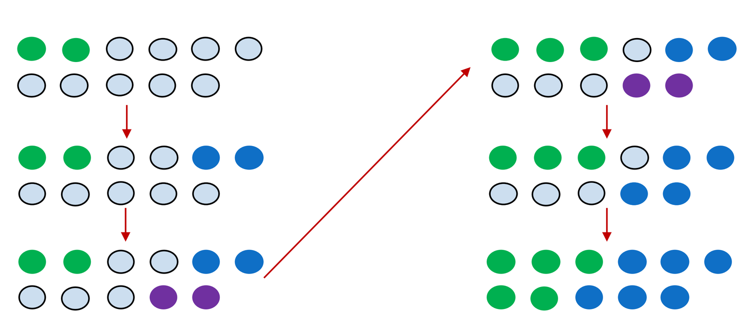
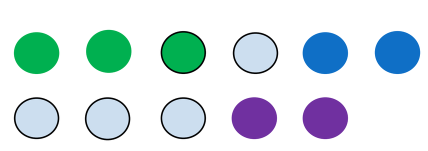
Step 2 – Define the ML algorithm

- The algorithm aggregated different aINCAT scores in clusters (Table 2).
- The EuroQoL 5 Dimensions (EQ- 5D) data based on UK preference weights was used as a measure of the health-related quality of life (HRQoL) to characterize the health condition.
- Starting from 11 clusters corresponding to each aINCAT score, each iteration of the algorithm tested aggregations of the existing clusters and selected the one associated with the lowest increase in within-cluster dissimilarity across the sample. Only clusters of adjacent aINCAT scores were considered for aggregation to limit the number of possible combinations.
- Dissimilarity was assessed using the sum of squares residuals (SSR), and the optimal number of clusters was selected using the Gap statistic[2].

Steps 3 and 4 – Apply the ML algorithm and test the results

- We ran the algorithm on 1000 bootstrap data samples to assess which cluster sets were most frequently selected as optimal and assessed their quality using ANOVA.

Table 2 – Graphical representation of the algorithm

Group the observations in 11 clusters corresponding to the levels on the aINCAT score	
Test possible aggregations of the clusters For each possible aggregation, measure the within cluster dissimilarity in HRQoL using SSR	
Select the aggregation associated with the lowest increase in SSR	
Iterate until reaching two clusters	
Use the Gap Statistics to select the optimal number of clusters	

RESULTS

Clusters most frequently selected as optimal

The dataset included 1,057 observations.

The two most frequently selected Cluster Sets were:

- Cluster Set A, 8.1% of the simulations: aINCAT 0, aINCAT 1-2, aINCAT 3, aINCAT 4, aINCAT 5-6, aINCAT 7-8, aINCAT 9-10
- Cluster Set B, 7.2% of the simulations: aINCAT 0-1, aINCAT 2-3, aINCAT 4, aINCAT 5-6, aINCAT 7-8, aINCAT 9-10.

Table 3 – most frequently selected cluster sets

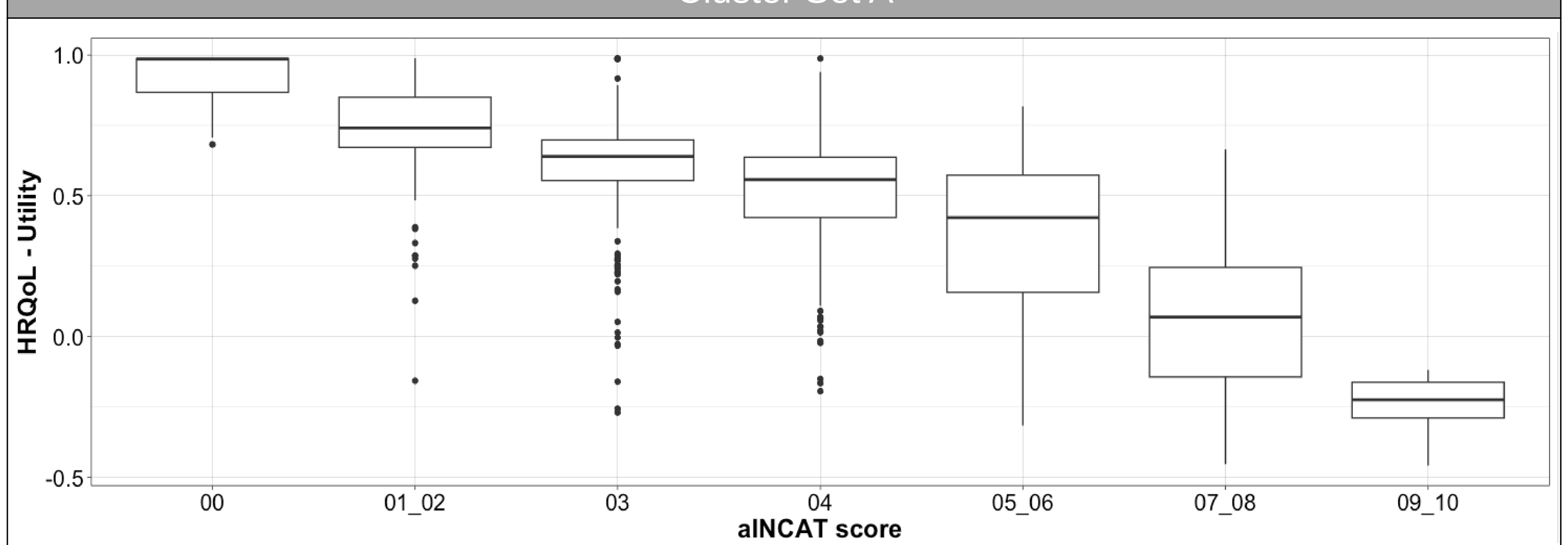
	8.1%	7.2%	6.4%	5.2%
aINCAT 0				
aINCAT 1				
aINCAT 2				
aINCAT 3				
aINCAT 4				
aINCAT 5				
aINCAT 6				
aINCAT 7				
aINCAT 8				
aINCAT 9				
aINCAT 10				

Health Related Quality of Life

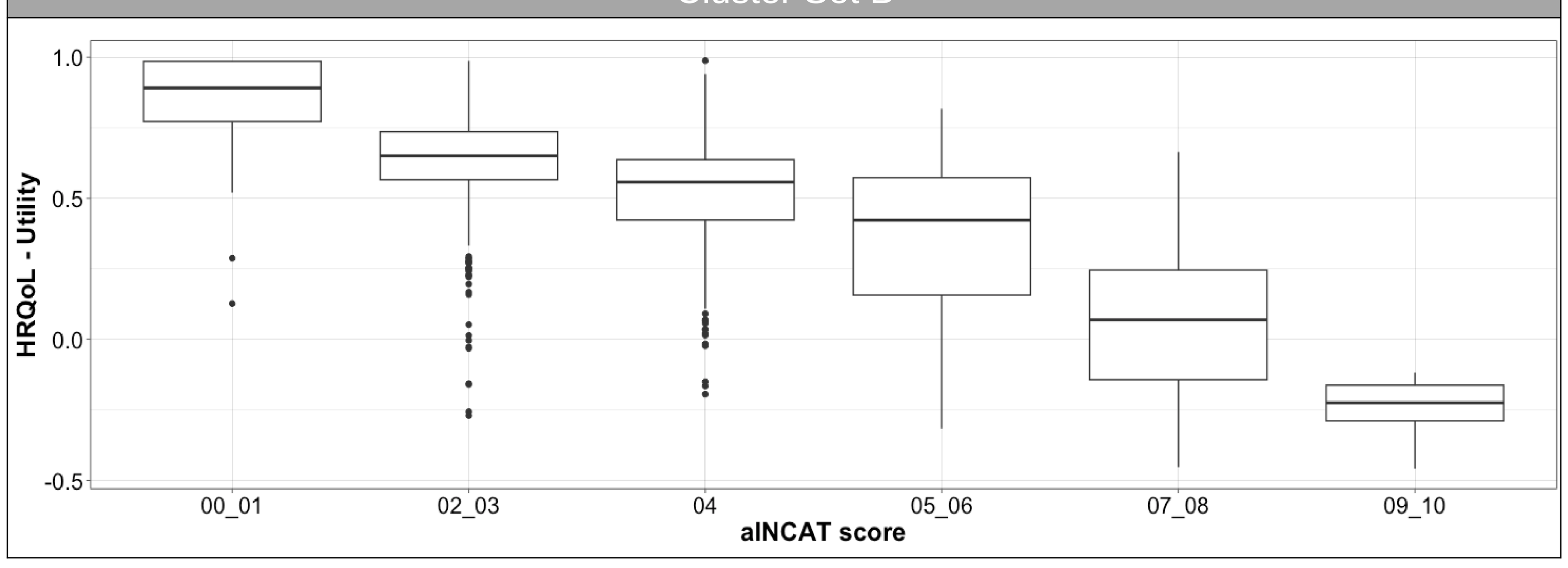
- The health states defined by Cluster Set A and Cluster Set B are well differentiated in terms of HRQoL.
- There is no clear drop in HRQoL utility between one health state and the adjacent ones, which properly reflects different severity levels over a continuous spectrum.
- The p-values from the ANOVA tests were significant ($p<0.001$) for both Cluster Sets.

Table 4 – Health Related Quality of life stratified by cluster in the two most frequently selected Sets

Cluster Set A



Cluster Set B



CONCLUSIONS

- We developed a de-novo ML algorithm to analyze the ADHERE data and identify clinically relevant health states based on different levels of functional impairment in CIDP.
- The algorithm is largely based on the Ward's method[3] and may be subjected to the same limitations. These include sensitivity to outliers, tendency to join clusters with a small number of observations, and bias toward producing clusters with the same number of observations[4].
- Despite these limitations, the algorithm identified Cluster Sets that are well differentiated in terms of HRQoL, with a significant difference across the clusters. This suggests the clusters are appropriate to represent health states that may be the basis for the development of future CEMs. Further validation from clinical experts is needed to confirm this result.

ABBREVIATIONS:
aINCAT: adjusted Inflammatory Neuropathy Cause and Treatment; ANOVA: Analysis of Variance; CEM: Cost-Effectiveness Model; CIDP: Chronic Inflammatory Demyelinating Polyneuropathy; Eq5D: EuroQoL 5 Dimensions; HRQoL: Health-Related Quality of Life; I-RODS: Inflammatory Rasch-built Overall Disability Scale; MGS: Mean Grip Strength; ML: Machine Learning; MRC: Medical Research Council; SSR: Sum of Square Residuals; TUG: Time-Up and Go.

REFERENCES:
1. Allen J, Basta I, Eggers C, et al.. *Neurology*. 2024;102(17_supplement_1):6214. doi:doi:10.1212/WNL-.0000000000206324
2. Tibshirani, R., G. Walther and T. Hastie (2002). Journal of the Royal Statistical Society Series B: Statistical Methodology 63(2): 411-423.
3. Ward Jr, J. H. (1963). Journal of the American Statistical Association 58(301): 236-244.
4. Milligan, G. W. (1980). Psychometrika **45**(3): 325-342.

Presented at ISPOR EUROPE 2024; 17-20 November 2024; Barcelona, Spain.