

Clustering of diabetes treatment trajectories

Romane Le Goff, Statistician intern at IQVIA (28/02/2022 – 31/08/2022)

M2 Applied Mathematics, Statistics at Université de Rennes 1 – ENSAI

Supervisors: Geoffray Bizouard and Oriane Bretin

Contact: romane.le-goff@laposte.net

Abstract

Introduction: All medical acts such as prescriptions throughout the patient's follow-up compose his or her care pathway, also called treatment trajectories. In real-life studies, where the number of distinct pathways can be very large, the search for similarities of patient trajectories is essential to better understand the population studied. The objective of the internship was to compare methods to cluster treatment sequences. The underlying goal of clustering sequences lies in the computation of the dissimilarity matrix to feed clustering algorithms (AHC, PAM).

Methods: A cohort of diabetic treatment-naïve patients was followed between 2015 and 2021. Data were extracted from an IQVIA database containing treatments dispensed by French pharmacies and prepared for sequence analysis. Various dissimilarity measures for sequence analysis have been explored and compared, e.g., Optimal Matching. Index plots were mainly used to compare the clustering results. The classes were described according to age, gender, and type of diabetes.

Results: Based on our simulations, dissimilarity measures were sensitive to different aspects of sequences comparison (duration, timing, sequencing). The applied methods seemed to separate patients in classes according to the severity of their therapeutic schemes. There are no statistical criteria that is sufficient to choose the ideal number of clusters. It is hence necessary to test different clustering solutions and select the most relevant ones.

Conclusion: Numerous methods exist to compute dissimilarity between sequences, stressing different aspects of the trajectories. The way treatment sequences are built can influence results, which requires a careful data management step. Sequence analysis is exploratory and requires interactions with experts to interpret results.