The TAK-medoids, an algorithm to visualize healthcare pathway pattern on a very large cohort of patients: application on 50,000 patients with COPD

Pauline Lemeille [1], **Jean Louis Pépin** [2], **Alain Palot** [3], **Arnaud Panes** [1], **Aurélie Schmidt** [1], **Anne Josseran** [5], **Florent Lavergne** [5], **Arnaud Prigent** [4], **Hélène Denis** [1]

1. Heva, Lyon; 2. INSERM U1042, Grenoble; 3. Hôpital Saint Joseph, Marseille; 4. Groupe médical de pneumologie, Polyclinique Saint Laurent, Rennes; 5. RESMED, Saint-Priest

ICPE
23-27 August 2023
Halifax, Canada

Background

Innovative **artificial intelligence** methods, especially **Machine Learning based methods**, have shown strong interest to visualize patient's healthcare pathways.

The **Time-sequence Analysis through K-clustering algorithm**, known as **TAK®**, was first introduced in 2019, in order to produce a **visualization of treatment patterns** for an entire cohort.

To improve the interpretation and visualization of medical events sequences and healthcare pathways for **very large cohorts of patients**, the **TAK-medoids**, method derived from meta-TAK [1] is of special interest.

Indeed, to have readable and understandable results for the whole cohort, it is not possible to represent the treatment sequences of every single patient. Therefore, the idea is to regroup patients with a similar sequence.

For diseases affecting large number of patients, such as **COPD**, this double-clustering method based on the TAK algorithm, is thus very interesting to **identify and compare patients with similar healthcare pathways**.

Objective

To apply **artificial intelligence** method, the **TAK-medoids algorithm**, on medical events sequence of a large cohort of **54,545 patients with COPD** to **visualize their healthcare pathway** and to **identify and describe groups with similar pathways**.

Regulatory framework

Specific approvals for this study were obtained from the Comité Ethique et Scientifique pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé (CESREES; ref: 3904033) and from the Commission Nationale Informatique et Liberté (CNIL), the French information technology and personal data protection authority (DR 2021 162 and n° 921198).

Disclosures of interest

This study is sponsored and financed by RESMED. FL and AJ are employees of ResMed.
PL, HD, AS and APan are employees of HEVA and their participation in this study was funded by ResMed.
JLP has received lecture fees or conference traveling grants from ResMed, Philips, AstraZeneca, Jazz Pharmaceuticals, Agiradom and Bioprojet, and has received unrestricted research funding from ResMed, Philips, GlaxoSmithKline, Bioprojet, Fondation de la Recherche Medicale (Foundation for Medical Research), Direction de la Recherche Clinique du CHU de Grenoble (Research Branch Clinic CHU de Grenoble), and fond de dotation "Agir pour les Maladies Chroniques" (endowment fund "Acting for Chronic Diseases").
APal has received consulting fees from ResMed. APr has received investigator fees for clinical trials funded by ResMed.

Glossary

COPD: Chronic Obstructive Pulmonary Disease
NIV: Non invasive ventilation
SNDS: French National Health Insurance Database (Système National des Données de Santé)
TAK: Time-sequence Analysis through K-clustering

References

[1] De Oliveira et al. - DOI: 10.13140/RG.2.2.29680.07680
[2] Tuppin P, Rudant J, Constantinou P, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Rev Epidemiol Sante Publique 2017; 65 Suppl 4: S149-s67.

Conclusion

By modifying the meta-features originally used in the Meta-TAK and by simplifying the resulting visualization, we manage to get a **readable representation of medical events in a cohort of more than 50,000 patients.**

While the Meta-TAK was introduced in order to face the computation complexity issue, the TAK-medoids goes one step further by simplifying the resulting cohort visualization and by being performant on medical events, more occasional and shorter than usual treatment phases the TAK was created for.

Heva ResMed

Methodology

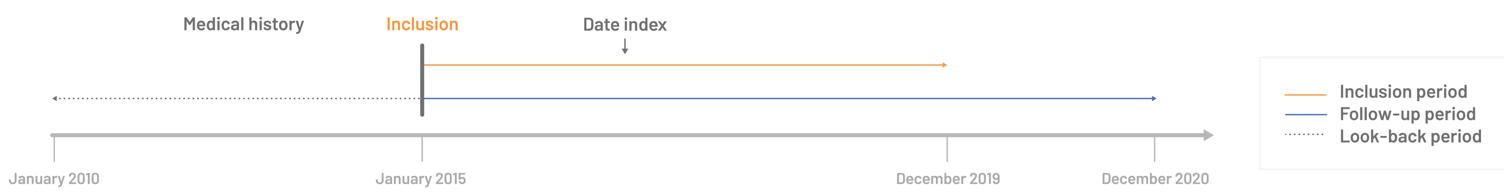
Data sources

Data were extracted from the **French National Health Insurance Database (SNDS)** which covers about 98.8% of the population living in France [2].

The SNDS contains socio-demographic data (age, sex, and area), and information on all health care expenses, including outpatient visits, reimbursed medication, medical procedures, hospital admission diagnoses and procedures, and date of death.

Study population and study period

The study population comprised individuals aged **≥40 years** who had **at least one reimbursement for Non-Invasive Ventilation (NIV)** between **1 January 2015 and 31 December 2019**.



Only incident cases were included (i.e. those with no NIV treatment in the previous 5 years); the index date was the date of first delivery of NIV during the study inclusion period.

The presence of COPD was determined based on **ICD-10 codes (J41 – J44 or J961)** during at least one hospitalization of presence of long-term disease. Data collection for included individuals continued until December 2020.

Outcomes definition

Only **severe exacerbations of COPD** that required hospitalisation were identified. Only hospitalisations for cardiology or respiratory diagnoses were recorded.

Statistical analysis / Data visualisation

Results

Cohort description

A total of 54,545 incident COPD patients treated with NIV were included in the analysis

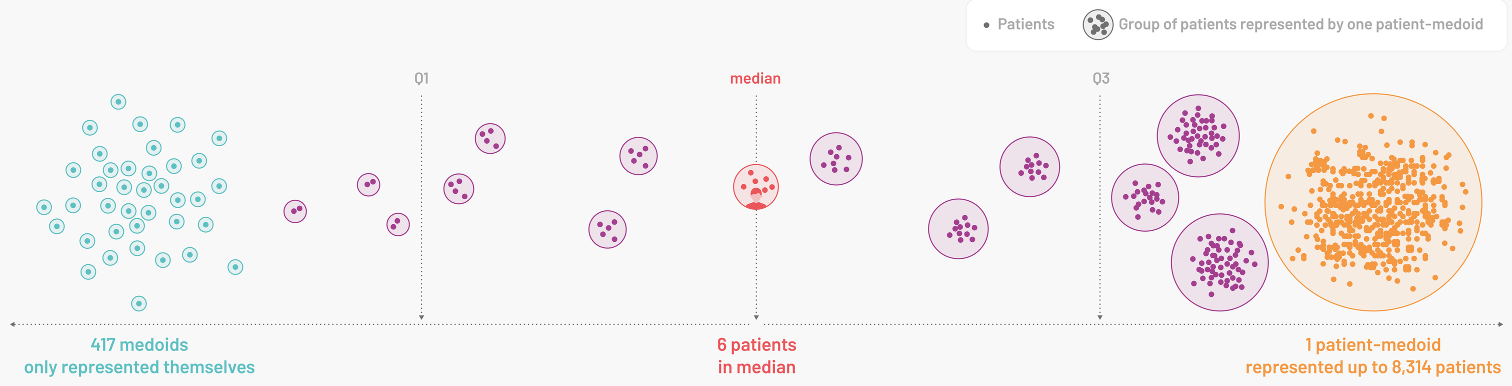
N=150,518
Individuals having at least one reimbursement for home NIV between Jan 1, 2015 and Dec 31, 2019

N=54,545
Individuals aged ≥40 years with confirmed COPD during the pre-NIV period
Median age : 70 (IQR: 62 ; 79)
51% of men

Optimization of the statistical model: Selection of medoids

A total of 50 meta-features were selected corresponding to the number of observations of each pathway event, and the order of occurrence of these events.

The objective is to select the patient-medoids so that they are very different one from another (so that we limitate the number of patient-medoids and so the complexity and the computational time) et all patients represented by the same patient-medoids are very similar one from another.



The optimal number of patient-medoids was 3,000, each of them representing 6 patients in median (**Q1**: 3; **Q3**: 12). One patient-medoid represented up to 8,314 patients while 417 medoids only represented themselves.
The mean homogeneity score for the distance between all patients represented by the same patient-medoid was 95.3% (standard deviation 8%), **Q1** = 94.2%, **Q3** = 99.9%.

Final vizualisation of clusters

Finally, using the hierarchical agglomerative clustering on patient-medoids, we identified and represented 4 main care pathways. Two are represented in this poster as examples:

Cluster A: Early initiation of the NIV

- Concerned 66% of the whole cohort (N= 35,975 patients).
- Represented by 458 (15%) patient-medoids → many patients with homogeneous pathways.

Cluster B: Late initiation of NIV, following several episodes of exacerbations

- Concerned 25% (N=11,375 patients).
- Represented by 1,581 (53%) patient-medoid → more heterogeneous cluster.