Hélène Denis[1], **Marie-Ange Rasendra**[1], **Aurélie Schmidt**[1], **Jean-Louis Pépin**[2], **Anne Josseran**[3], **Florent Lavergne**[3], **Martin Prodel**[1], **Alain Palot**[4], **Arnaud Prigent**[5], **Eleonore Herquelot**[1]

1. Heva, Lyon, France ; 2. Grenoble Alpes University, Inserm U1300, CHU Grenoble Alpes, HP2, Grenoble, France ; 3. ResMed Science Center, Saint-Priest, France ; 4. Hôpital Saint-Joseph, Marseille, France ; 5. Polyclinique Saint-Laurent, Rennes, France

DSVR 2024

20 juin 2024 - Paris, France

546733

Intérêt de l'Intelligence Artificielle et des méthodes de clustering

Exemple sur une cohorte de 50 000 patients atteints d'une BPCO à partir des données du SNDS

INTRODUCTION

Les patients atteints de bronchopneumopathie chronique obstructive (BPCO) et pris en charge par Ventilation Non Invasive (VNI) peuvent avoir des parcours de soins très différents en France. Une précédente étude utilisant des méthodes de machine learning a permis d'identifier différents clusters de prise en charge pré-VNI, ayant un impact sur la mortalité des patients[1].

L'objectif de ce travail est de montrer la variabilité observée sur les groupes identifiés, ainsi que sur les visualisations et interprétations possibles en fonction de différentes méthodes de clustering.

MÉTHODE & RÉSULTATS

Ce travail porte sur une cohorte de patients BPCO > 40 ans traités par VNI identifiés à partir des données du Système National des Données de Santé (SNDS) entre le 01 janvier 2015 et le 31 décembre 2019.

Trois méthodes de clustering ont été testées, prenant en compte le parcours de soins dans l'année précédant la mise sous VNI, dont : les hospitalisations pour atteinte cardiaque ou respiratoire, les hospitalisations en soins intensifs / passages en unité de soins intensifs de cardiologie (USIC) et les exacerbations de la BPCO. La probabilité de Survie Globale (SG) à 24 mois a été évaluée par l'estimateur de Kaplan Meier.

Cartes-patient

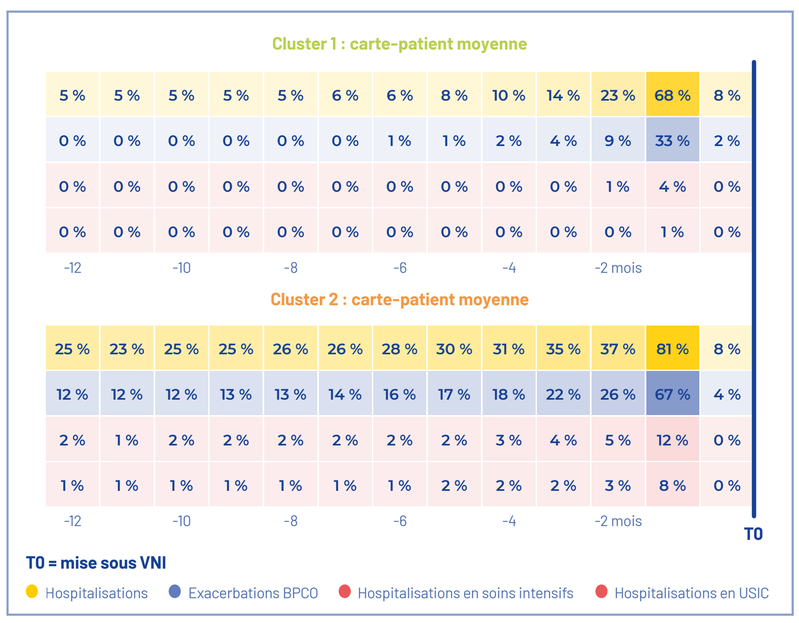
Méthode

La 1ère méthode était basée sur des cartes-patients, utilisées en tant qu'images, sur lesquelles un algorithme de machine learning non supervisé K-means a été appliqué couplé à la distance Dynamic Time Warping (DTW).

Clusters obtenus

- **Cluster 1** (68 % de la cohorte) : patients ayant majoritairement des hospitalisations et des exacerbations juste avant la mise sous VNI, SG à 24 mois : 76 %
- **Cluster 2** (32 %) : patients ayant tendance à avoir des hospitalisations et des exacerbations BPCO plus fréquentes dans l'année précédant la mise sous VNI, SG à 24 mois : 60 %.

Cartes-patients (ou heatmaps) obtenues pour les 2 clusters identifiés par l'algorithme de machine learning non supervisé K-means présentant la fréquence des événements dans l'année précédant l'initiation de la VNI



Guide de lecture : Chaque ligne représente un événement, chaque colonne représente une unité temporelle (le mois ici). Plus l'évènement est fréquent, plus la couleur est foncée. Le pourcentage correspond aux patients ayant au moins un évènement d'intérêt pendant le mois correspondant à la case.

Avantages

Modélisation intéressante lorsqu'il y a beaucoup d'événements ponctuels et concomitants à prendre en compte, sans avoir à les prioriser.

La distance DTW permet de prendre en compte la similarité de deux événements identiques proches dans le temps.

Inconvénients

Le pas de temps mensuel, nécessaire pour des raisons de lisibilité, limite la précision temporelle de l'enchainement des événements.

Visualisation synthétique ne permettant pas de visualiser les séquences de traitement de chaque patient du cluster (contrairement au TAK®).

TAK®

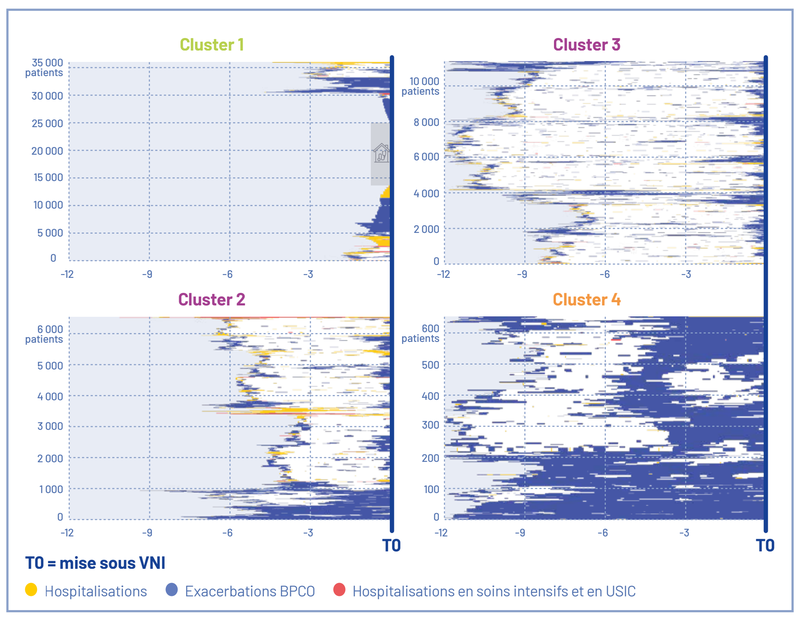
Méthode

La 2ème méthode, TAK® (Time-sequence Analysis through K-clustering)[1], était basée sur une analyse des séquences temporelles d'évènements et un regroupement des séquences similaires par un clustering HCA.

Clusters obtenus

- **Cluster 1** (66 %) : patients ayant une VNI initiée en ambulatoire ou après le premier événement aigu (exacerbation/hospitalisation), SG à 24 mois : 76 %
- **Cluster 2** (12 %) : patients ayant 2 exacerbations sévères dans les 6 mois avant mise sous VNI, SG à 24 mois : 63 %
- **Cluster 3** (21 %) : patients ayant des exacerbations fréquentes pendant l'année précédant la VNI, SG à 24 mois : 61 %
- **Cluster 4** (1 %) : patients ayant beaucoup d'hospitalisations/exacerbations dans l'année pré-VNI, SG à 24 mois : 42 %

Trajectoires de soins dans l'année précédant l'initiation de la VNI pour les 4 clusters identifiés par la méthode TAK®[1]



Guide de lecture : Une ligne représente un patient (ou un groupe de patient avec trajectoire similaire). Le rectangle gris avec un symbole de maison dans le groupe 1 indique les personnes qui ont bénéficié d'une initiation de la VNI à domicile ou lors d'une consultation en cabinet privé (cadre ambulatoire) sans hospitalisation ou exacerbation préalable (n = 11 405 ; 21 %).

Avantages

Le TAK® permet de réaliser un clustering de la séquence temporelle des patients avec une granularité temporelle très fine (au jour) et permet de visualiser lisiblement un volume d'informations très dense.

Inconvénients

Le TAK® ne permet pas de prendre en compte la singularité des événements concomitants du fait de son unique dimension temporelle.

En cas d'événements simultanés, il faut donc les prioriser ou créer un événement mixte. De plus au-delà de 10 événements le TAK perd en lisibilité.

Modèle de mélange

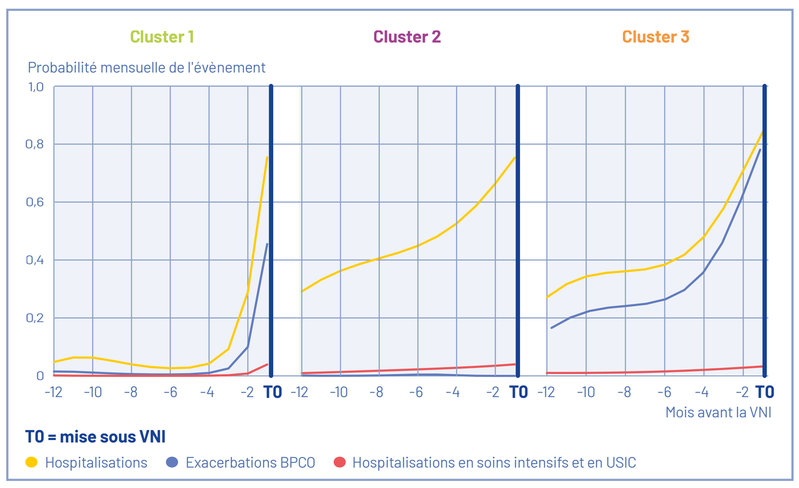
Méthode

La 3ème méthode était basée sur un modèle de mélange permettant de classifier les différentes trajectoires longitudinales. Le nombre optimal de cluster était sélectionné en comparant les critères BIC.

Clusters obtenus

- **Cluster 1** (65 %) : patients ayant une augmentation des exacerbations et des hospitalisations dans les 3 mois qui précède la mise sous VNI, SG à 24 mois : 77 %
- **Cluster 2** (9 %) : patients ayant des hospitalisations importantes hors exacerbation, SG à 24 mois : 65 %
- **Cluster 3** (26 %) : patients ayant des exacerbations fréquentes dans l'année qui précède la VNI, SG à 24 mois : 57 %

Trajectoires de soins dans l'année précédant l'initiation de la VNI pour les 3 clusters identifiés par le modèle de mélange



Guide de lecture : Chaque graphique représente la dynamique d'évolution dans les groupes, en abscisse, on a chaque mois qui précède la mise sous VNI et en ordonnée, on a la probabilité de chaque évènement d'intérêt. Pour le groupe 1, on observe une probabilité d'hospitalisations qui passe de 0.3 sur le 12ème mois avant la mise sous VNI à 0.8 le mois qui précède la mise sous VNI. Les probabilités des autres évènements restent assez faibles pour les exacerbations et les hospitalisations en cardiologie en unité intensive (<0.05).

Avantages

Cette méthode flexible peut permettre de prendre en compte plusieurs outcomes répétés dans le temps et de mettre en évidence des dynamiques d'évolution différentes entre les groupes.

Le choix du nombre de groupes peut se faire assez simplement par comparaison des critères AIC ou BIC.

Inconvénients

Le modèle de mélange fait des hypothèses sur les distributions de chaque outcome.

Dans sa version la plus simple, la prise en compte des corrélations sur les données répétées pour les patients est assez limitée et il peut y avoir des problématiques de temps de calculs long et de convergence dans sa version plus complexe.

CONCLUSION

L'ensemble des méthodes de clustering permettent de rationnaliser l'existence de plusieurs types de trajectoires de soins pré-VNI :

- 1 groupe de patients initiés très rapidement après le(s) premier(s) épisode(s) d'exacerbation(s) aigu(s).
- 1 groupe de patients pour qui la VNI a été mise en place plus tardivement, après plusieurs exacerbations / hospitalisations dans l'année précédant la mise sous VNI.

Pour les 3 méthodes de clustering, on retrouve une mortalité significativement plus importante pour les groupes pour lesquels la VNI a été mise en place plus tardivement (après plusieurs exacerbations/hospitalisations dans l'année précédant la mise sous VNI).

Les différentes méthodes de clustering permettent une visualisation complémentaire des parcours de soins. La granularité nécessaire pour décrire les parcours de soins peut impacter le choix de la méthode : un plus grand nombre de cluster peut permettre de visualiser un peu plus précisément les parcours de soins.

Références

[1] https://www.thelancet.com/journals/lanepe/article/PIIS2666-7762(23)00136-9/fulltext

Sources des données

L'étude MONTANA a été enregistrée sur le site du HDH, approuvée par le Comité Ethique et Scientifique pour les Recherches, les Études et les évaluations dans le domaine de la Santé (CESREES; ref: 3904033) et par la Commission Nationale Informatique et Liberté (CNIL, DR 2021 162 et n° 921198).

Heva
Une filiale de DOCAPOSTE

ResMed