

PHÉNOTYPE DE PATIENTS AYANT UN SYNDROME D'APNÉE DU SOMMEIL

Morgane Bordelais¹, Baptiste Delahais¹ & Thomas Sauvaget²

¹ Ensai, Campus de Ker-Lann - Rue Blaise Pascal - BP 37203 - 35172 BRUZ cedex
morgane.bordelais@eleve.ensai.fr, baptiste.delahais@eleve.ensai.fr,

² Insee, DMCSI-DMS-Div.Sondages - 88 Avenue Verdier – 92120 MONTRouGE
thomas.sauvaget@insee.fr

Résumé. *[Communication dans le cadre du Prix SFDS-Ensai]* À partir des données issues de l'Observatoire Sommeil de la Fédération de Pneumologie concernant environ 45 000 patients atteint de Syndrome d'Apnée Obstructive du Sommeil (SAOS), nous construisons un modèle utilisant 24 variables basées sur des symptômes et des comorbidités. Une analyse en classes latentes révèle 8 classes qui se trouvent être nettement distinctes, homogènes, et intéressantes d'un point de vue clinique (avec notamment une classe composée uniquement de femmes, et trois autres en grande majorité d'hommes). Pour apprécier la robustesse de cette classification, un apprentissage supervisé par régression polytomique est mis en œuvre : le très faible taux de mal classés obtenu sur échantillons de validation est un résultat prometteur quant à l'intérêt de notre classification.

Mots-clés. Médecine, épidémiologie, modèles de classes latentes.

Abstract. *[Presentation given for the SFDS-Ensai student prize]* Using data from the french Sleep Observatory of the Pneumology Federation on about 45 000 patients displaying Obstructive Sleep Apnea (OSA), we build a model containing 24 variables combining various comorbidities and symptoms. A latent class analysis reveals 8 classes, which happen to be very distinct, homogeneous, and interesting from a clinical perspective (in particular, one class consists exclusively of women, and three others mostly of men). To appreciate the robustness of the classification, we perform a supervised learning by polytomic regression : the very low rate of misclassification obtained on cross-validation samples is a promising result regarding the interest of our classification.

Keywords. Medicine, epidemiology, latent class analysis.

1. Phénotype de patients ayant un syndrome d'apnée du sommeil

Obstructive Sleep Apnea (OSA) is a syndrome characterized by frequent obstruction, either partial or complete, of respiratory airways during sleep. These obstructions can induce myocardial damage. Thus OSA is associated with cardiovascular morbidity and mortality. It is therefore of considerable importance to develop efficient treatment of OSA.

Patients diagnosed with OSA can display very different profiles. Firstly, while the syndrome is more prevalent in older men with obesity, it may occur in lean subjects, women, and young people. Some patients experience a range of symptoms (among others : snoring, sleep fragmentation, morning headaches, general tiredness, and nocturia), while others do not. As for comorbidities, from coronary heart disease to hypertension and diabetes, they can be present, but not always and not necessarily concomitantly.

The aim of our study is precisely to investigate whether patients could be organized based on their characteristics into a few groups that are at the same time fairly distinct, and as homogeneous as possible. Our work builds on the 2016 paper of Sébastien Bailly [B2016], our tutor on this project, who had performed a first clustering study using Ascending Hierarchical Clustering (AHC) on data obtained from the french sleep observatory of the pneumology federation (OSFP). We had at our disposal a larger data set from OSFP.

After reviewing several methods mentioned in the literature, we settle on Latent Class Analysis (LCA), a model-based method. Then we describe our choice of variables, and optimize the Integrated Complete Likelihood (ICL) criteria, which penalizes the overlap of classes. This results in 8 clusters of OSA patients.

Our analysis reveals that these classes are very distinguishable, and possess desirable medical aspects. In particular, comorbidities such as hypertension and diabetes have very different prevalence between the clusters. One of the classes contains mostly patients with very low levels of symptoms, even snoring, and may have medical significance. We also obtain a class containing nearly exclusively women, a result which might be explained by menapausal processes and also seems especially interesting.

In order to assign new patients to classes, as would doctors need to do, we first construct a CART tree. The result is disappointing (40 % of misclassification), so we then perform a polytomic regression. Its analysis shows that new patients can be assigned to classes with high accuracy (2 % of misclassification). It is thus possible to use our classification easily and reliably.

Finally our results are compared with those of Sébastien Bailly. Some partial overlap is found with some of the 6 clusters he obtained, while some of our clusters do not appear in his results. In particular, the highly feminized class is new.

Overall, it appears that our chosen model has a number of interesting properties, and may become useful for screening and improving the care of OSA patients.

2. Bibliography

[B2016] Sébastien Bailly, *et al.* (2016). Obstructive Sleep Apnea: A Cluster Analysis at Time of Diagnosis, *PLOS ONE* 11.6 , e0157318