Assessing the heterogeneity in the individuals to fit different models: an application to predict mortality in a large sample of sleep apnoea patients

G. Castellà [1], R. Boix [1], C. Colls [2], A. García-Altés [2], I. Teixidó [3], J. Mateo[3], F. Solsona [3], J. Escarrabill [4], M. Sánchez-de-la-Torre [5,6], F. Barbé [5,6], J. Valls [1].

Background

Usually, in biomedical sciences, studies and trials are designed to give an answer to particular hypothesis. Under this context or similar (called "small data"), the problem of modelling an outcome is usually solved by providing one single model, associated with an acceptable goodness of fit, using for this a set of predictors and a given modelling strategy. Thus, the same functional relationship is assumed between the predictors and the response for all the individuals in the sample, and inference is conventionally made in terms of the general expectation. However, in the "big data" context, or other intermediate scenarios where the available sample size is large, it might be reasonable to consider different clusters of homogeneous individuals where different models could be fitted.

Obstructive Sleep Apnoea (OSA) is characterised by repetitive obstructions of the upper airway during sleep. OSA is associated with a reduction of oxygen saturation, as well as cardiovascular morbidity, but it is unclear whether OSA is an independent factor associated to mortality, with heterogeneous results in the literature. In current clinical practice, the most used treatment for OSA is Continuous Positive Airway Pressure (CPAP), which consists of pumping air through the upper airway to prevent the obstruction. CPAP treatment, with an adequate compliance, has been proved to reduce excessive daytime sleepiness and the risk of hypertension, and is also associated with an increase of life quality. However, it remains unclear whether a low compliance of the CPAP treatment could be associated with and increased mortality risk. Besides, it is estimated that aproximately 70.000 out of over seven million people in Catalonia is under CPAP treatment at the moment. The *AgÃĺncia de Qualitat i AvaluaciÃş SanitÃăries de Catalunya* (AQuAS) has access to data from the catalan public health system, and particularly to all 70.000 OSA patients in Catalonia under CPAP treatment.

Aims

Here, we analysed all the patients with a diagnostic of sleep apnoea that were under CPAP treatment during 2012 and/or 2013 in Catalonia, including a total number of 75,194 patients. The goal was to model the mortality observed in these patients during this period of time, using as predictors variables related to their burden into the health system, their comorbidities and other clinical variables.

Methods

All OSA patients under CPAP treatment during 2012 and 2013 in Catalonia older than 40 years were included. Sex, age, mortality from January 2012 to May 2015, time since CPAP prescription and variables related to burden in the Health System (BHS), such as the number and length of hospitalizations, the number of visits to primary and secondary care (PC and SC) were obtained from Catalan Health Registries. First, a Principal Component Analysis was performed, using the BHS variables. Second, supervised and non-supervised clustering methods, including hierarchical clustering, k-means and classification and regression trees, were used to initially determine different partitions of the patients. Third, different modelling strategies were performed to assess differences in mortality in each of the clusters considered.

Results

75,085 OSA patients were included (75.01% males, age 63.18 ± 10.80), with a mortality of 5.50% (4,132 patients). Mean time from CPAP prescription was 3.45 ± 2.62 years, number of visits to PC and SC 12.97 ± 14.37 and 0.13 ± 0.43, respectively, and number of hospitalizations 0.49 ± 1.16, with a mean total length stay of 3.42 ± 11.30 days (Table 1).

The results showed a main cluster (59,941; 79.92%) representing the general pattern, with a mortality of 3.17% (Figure 1). Two secondary clusters (7,198 and 7,157; 9.6% and 9.54%) corresponded to patients with increased mortalities (4.89% and 19.62%), increased number of visits to SC and PC with higher presence of females and aged patients. Finally, four minor clusters (< 1%) were detected with further increased mortalities (51.29%, 57.89%, 58.33%, and 100%), some of them associated to a higher time since CPAP prescription or larger stages in hospitals.

Interestingly, when assessing differences in mortality in each cluster, age, sex, time from CPAP prescription, and BHS related variables were significant in almost all clusters. The heterogeneity analyses, also revealed that some of these effects were drastically different from cluster to cluster, particularly for sex and time from CPAP prescription. Despite the significant differences observed in the generalized linear models for these variables, the BHS related variables were way better at predicting mortality, although the patterns found were different in all clusters (Figure 2).

Conclusion

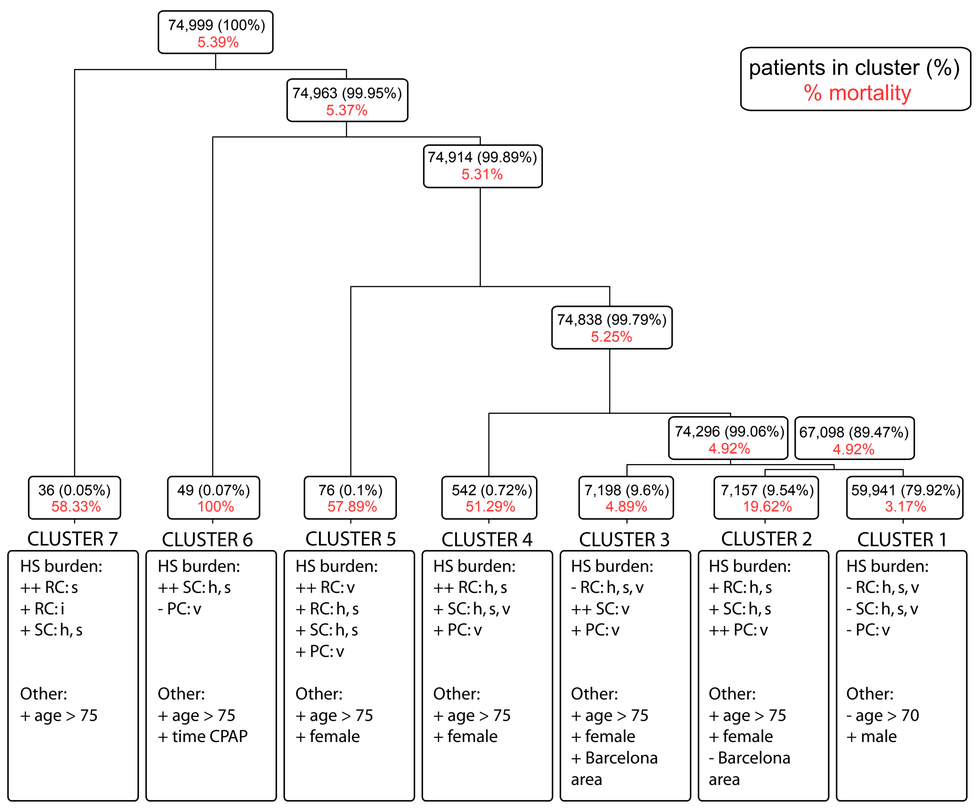
The results obtained with this approach suggest that the fit of different models for specific group of patients can be useful to reveal different patterns with relation to mortality, which would remain undetected when considering all the patients in the same model. The availability of large datasets in biomedical sciences may enable the possibility of dealing with the heterogeneity in patients and therefore allow advances in personalised medicine.

Table 1. Description of the CPAP-treated population, overall and by mortality.

	Overall (n = 75085; 100%)	Alive (n = 70953; 94.5%)	Death (n = 4132; 5.5%)
Age (years)			
40 - 49	8296 (11.66%)	8224 (12.2%)	72 (1.93%)
50 - 59	18095 (25.43%)	17754 (26.33%)	341 (9.13%)
60 - 69	24524 (34.46%)	23646 (35.07%)	878 (23.5%)
70 - 79	15463 (21.73%)	14078 (20.88%)	1385 (37.07%)
80 - 89	4653 (6.54%)	3653 (5.42%)	1000 (26.77%)
90 - 94	132 (0.19%)	72 (0.11%)	60 (1.61%)
Sex (Male)	53377 (75.01%)	50660 (75.13%)	2717 (72.72%)
Time with CPAP treatment	3.45 (2.62)	3.43 (2.61)	3.79 (2.72)
Mean Charlson Score (Enhanced)	1.33 (1.54)	1.1 (1.3)	3 (1.99)
PC visits (per year)	6.52 (7.25)	6.39 (6.88)	8.8 (11.86)
SC visits (per year)	0.07 (0.26)	0.06 (0.21)	0.13 (0.65)
SC mean days of stage (per year)	1.42 (27.17)	0.67 (2.03)	14.52 (115.87)
SC hospitalisations (per year)	0.32 (1.54)	0.2 (0.49)	2.4 (5.94)
RC visits (per year)	0 (0.02)	0 (0.01)	0.01 (0.08)
RC mean days of stage (per year)	0.63 (11.01)	0.34 (6.9)	5.65 (37.23)

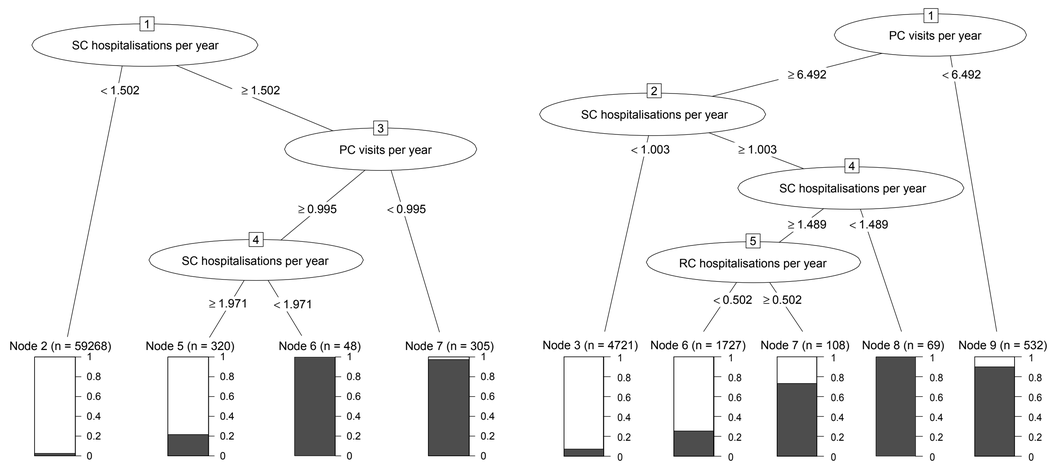
Data are presented as mean (Standard Deviation) or frequency (percentage) for quantitative and qualitative variables respectively.

Figure 1. Clusters of CPAP-treated OSA patients.



Total number of patients (and percentage) and mortality percentage (red) are shown for each cluster. Height of the branches is proportional to the distance among clusters. Characteristics defining each cluster are shown, with relation to the burden of the patients in the HS or other variables (++, +, -, - - indicate an increase in the mean or presence of the variables in the cluster with respect to the global pattern). PC, SC and RC stand for primary, secondary and respite care units; h, s, v for hospitalization, stage and visits.

Figure 2. Classification trees for predicting mortality in clusters 1 (left) and 2 (right).



Binary splits are displayed with the variable inside a circle and two branches with the values for the cut. Splits were performed based on the Gini criterion. Frequency and proportion of deaths (using bars) are also shown in each leaf node.

Institutions:
[1] Unit of Biostatistics and Epidemiology, Biomedical Research Institute of Lleida, Lleida, Spain
[2] Public Health Department, Government of Catalonia, Barcelona, Spain
[3] Computer Science & INSPIRES, University of Lleida, Lleida, Spain
[4] Respiratory Medicine Department, Government of Catalonia, Barcelona, Spain
[5] Respiratory Medicine Research Group. IRBLLeida. Lleida.
[6] CIBERes, Madrid, Spain.
* Corresponding authors: jvalls@irblleida.cat, gcastella.91@gmail.com

Supported by: ISCIII, ResMed Ltd. (Australia), Fondo de Investigación Sanitaria (PI10/02763 and PI10/02745), the Spanish Respiratory Society, the Catalonian Cardiology Society, Esteve-Teijin (Spain), Oxigen Salud, and ALLER.

Translational Research in Respiratory Medicine

BIOstatnet

Agència de Qualitat i Avaluació Sanitàries de Catalunya

Universitat de Lleida