Characterization of obstructive sleep apnea–hypopnea syndrome (OSA) population by means of cluster analysis

DONATO LACEDONIA[1], GIOVANNA ELISIANA CARPAGNANO[1], ROBERTO SABATO[1], MARIA MADDALENA LO STORTO[1], GIUSEPPE ANTONIO PALMIOTTI[1], VITO CAPOZZI[2], MARIA PIA FOSCHINO BARBARO[1] and CRESCENZIO GALLO[2]

[1]Department of Medical and Surgical Sciences, University of Foggia, Foggia, Italy; [2]Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Italy

SUMMARY
Obstructive sleep apnea–hypopnea syndrome (OSA) is being identified increasingly as an important health issue. It is typified by repeated episodes of upper airway collapse during sleep leading to occasional hypoxaemia, sleep fragmentation and poor sleep quality. OSA is also being considered as an independent risk factor for hypertension, diabetes and cardiovascular diseases, leading to increased multi-morbidity and mortality. Cluster analysis, a powerful statistical set of techniques, may help in investigating and classifying homogeneous groups of patients with similar OSA characteristics. This study aims to investigate the (possible) different groups of patients in an OSA population, and to analyse the relationships among the main clinical variables in each group to better understand the impact of OSA on patients. Starting from a well-characterized OSA population of 198 subjects afferent to our sleep centre, we identified three different communities of OSA patients. The first has a very severe disease [apnea–hypopnea index (AHI) = 65.91 ± 22.47] and sleep disorder has a strong impact on daily life: a low level of diurnal partial pressure of oxygen (PaO_2) (77.39 ± 11.64 mmHg) and a high prevalence of hypertension (64%); the second, with less severe disease (AHI = 28.88 ± 17.13), in which sleep disorders seem to be less important for diurnal PaO_2 and have a minimum impact on comorbidity; and the last with very severe OSA (AHI = 57.26 ± 15.09) but with a low risk of nocturnal hypoxaemia (T90 = 11.58 ± 8.54) and less sleepy (Epworth Sleepiness Scale 10.00 ± 4.77).

INTRODUCTION

Obstructive sleep apnea–hypopnea syndrome (OSA) is a common disease characterized by recurrent episodes of collapse of the upper airway during sleep, with incidence estimated at 5–14% among adults aged 30–70 years (Peppard *et al.*, 2013). The presence of sleep apnea is associated with cardiovascular diseases such as hypertension, heart failure and cerebrovascular diseases (McNicholas and Bonsigore, 2007), as well as diabetes and metabolic syndrome (Nieto *et al.*, 2009).

Despite this evidence, the clinical presentation of OSA patients is characterized by a large heterogeneity. For example, there is no linear relationship between obesity, severity of OSA and symptoms (Dixon *et al.*, 2007). Moreover, it is known that there is an association between obstructive sleep apnea and the presence of multi-morbidity, but it is not clear how much this association is due directly to OSA rather than to other factors often present in the same patients (Robichaud-Hallé *et al.*, 2012).

To characterize patients with OSA more effectively, some authors have hypothesized the possible presence of different 'phenotypes' of OSA and have demonstrated that, according to polysomnographic parameters, different polysomnographic phenotypes could be identified (Joosten *et al.*, 2012).

Cluster analysis is a statistical approach for studying the relationships present among groups of patients, or variables, in a large population. In recent years, cluster analysis has been applied widely in different biological fields and is particularly useful for the analysis of gene expression (Gallo and Capozzi, 2013). By applying the same methods it is possible to identify clusters of patients who are similar among themselves but significantly different from others. By means of this analysis, two recent papers (Vavougios *et al.*, 2016; Ye *et al.*, 2014) have demonstrated the possibility of identifying different phenotypes of OSA among a population of patients suffering from this disease.

The aim of this study was to improve knowledge about characteristics of OSA patients through clustering, using not only generic patient parameters but also introducing other instrumental data such as blood gas analysis and spirometry parameters. Furthermore, we analysed the relationships among the main clinical variables in each group to understand more clearly the impact of OSA on comorbidity and gas exchange.

MATERIALS AND METHODS

This is a retrospective study in which we analysed data from patients with a new diagnosis of OSA who were referred to the Sleep Centre at the University of Foggia in the last 2 years (2012–14). All patients underwent cardiorespiratory monitoring, spirometry, blood gas analysis and evaluation of sleepiness using the Epworth Sleepiness Scale (ESS). Finally, all information about the comorbidities present at the time of the first visit was also collected. Only subjects with a diagnosis of 'pure' OSA [apnea–hypopnea index (AHI) ≥ 5 events h^{-1}] were included in the analysis, so patients with suspect chronic obstructive pulmonary disease (COPD), obesity hypoventilation syndrome (OHS), central apnea and neuromuscular diseases (NMD) were excluded from the study. All subjects came from our region and represented a typical Caucasian southern Italy population. According to our inclusion criteria, a final data set was made of 198 subjects, while 20 main variables were selected for analysis: sex, age, body mass index (BMI), ESS, smoke, AHI, oxygen desaturation index (ODI), T90, six most representative comorbidities (cardiovascular diseases, hypertension, endocrinal disease, diabetes, asthma), forced vital capacity (FVC%), forced expiratory volume in 1 s (FEV_1%), FEV_1/FVC, pH, partial pressure of oxygen (PaO_2), partial pressure of carbon dioxide ($PaCO_2$) and bicarbonate (HCO^{3-}). Cluster analysis was then applied in order to identify groups of patients, and after identifying them the correlation between variables in each community was studied (see Statistical analysis for details).

The study was approved by the Ethic Committee of the Ospedali Riuniti di Foggia (file no. 10/CE/2015). Written informed consent to collect and analyse data was obtained from all participants and all data were anonymized and de-identified prior to analysis.

Polygraphy

While the patient remained in the sleep laboratory, unattended cardiorespiratory overnight monitoring (Alice PDx; Philips, Amsterdam, the Netherlands) was performed which registered oronasal flow measured by a nasal cannula flow, snoring, sleep position, arterial oxygen saturation (SaO_2) by finger probe, electrocardiography (ECG), abdominal and rib-cage efforts. Sleep-disordered breathing was quantified according to standard criteria of the American Academy of Sleep Medicine manual (Iber *et al.*, 2007). The examination was considered to be good if there were at least 6 h of registration. A manual score was performed the day after registration by a doctor experienced in sleep disorder.

Spirometry

Pulmonary function tests were performed in the pulmonary function laboratory using a spirometer (SensorMedics, Yorba Linda, CA, USA). To determine if any patient was affected by COPD, patients with FEV_1/FVC lower than 70% were not included in the study.

Blood gas analysis

An arterial blood sample for the analysis of gases during room-air breathing was drawn with the patient in the sitting position, the day after polygraphy registration and within 1 h of waking up. PaO_2, $PaCO_2$ and pH were measured in a blood gas analyser (Model 1312; Instrumentation Laboratory; Milan, Italy). Patients with $PaCO_2$ higher than 45 mmHg were excluded from the study because they could be affected by OHS or NMD.

Statistical analysis

Univariate and multivariate statistical analyses were performed on the available data set using Wolfram Mathematica V9 software (Wolfram Research, Champaign, IL, USA). The single probed patient was used as a statistical unit; every patient was annotated with information regarding the selected variables; thus, a matrix including values of all 20 variables was constructed.

Data set correlations (both between patients and the studied variables) were investigated by calculating the Spearman's rank correlation coefficient (r). Student's t-test with $n - 2$ degrees of freedom (df) was used to check whether the computed r-values were significantly different from 0 ($P < 0.05$).

We performed a principal component analysis (PCA) (Tipping and Christopher, 1999) on the full data set by first testing the correlation matrix among the 20 initial variables, the variance differences of which were confirmed by the Conover test. From PCA we first obtained seven principal components (corresponding to eigenvalues >1) explaining approximately 100% of the total variance; variables were

then selected having a communality of >0.5. This is an excellent method for synthesizing data variability, and allowed us to summarize the ways in which patient profiles vary under different conditions.

The behaviour and properties of systems consisting of functional parameters depend upon the ties and the relations between them, which are well depicted by a network. A non-orientated network (such as those represented in the figures) is representable by a graph $G = (V, E)$ where V is the set of nodes (or vertices) and E is the set of arcs (edges) connecting the nodes. Below the limit threshold of a node per arc, a network decomposes into aggregates of nodes (clusters) not mutually reachable, showing common characteristics.

After that, clustering was performed on data set values in order to identify groups of patients by means of a 'correlation network', which was built from the reduced data sets connecting only those patients correlating above a given threshold. In this network, edge labels show the correlation 'intensity' between nodes, while nodes represent patients. The correlation networks obtained show as 'communities', with many edges joining vertices of the same community and comparatively few edges joining vertices of different communities (Baumes *et al.*, 2005; Jeong *et al.*, 2000).

We verified the normality of the numerical variables using the Kolmogorov–Smirnov test. According to this, in order to assess statistically significant differences between communities a one-way analysis of variance (ANOVA) was applied to normal variables, while Kruskal–Wallis and chi-square were used for non-normal and categorical variables.

We also examined the obtained communities by observing correlations between variables in each community. This showed how they are manifested in principal component space reducing multi-dimensional data and determining the key variables in a multi-dimensional data set that explains the differences in the observations.

RESULTS

Clustering of patients

Our data set consists of values of nine selected variables from 198 different samples (patients), of which we analysed mutual relationships by computing the correlation coefficients between variables (or patients).

By arranging them into a visual representation (correlation graph) it is possible to discover groups of more strongly correlated elements by isolating those with a 'significant' correlation level: this can be conducted by establishing a minimum correlation threshold which allows us to highlight the most important relationships and discard the less relevant ones.

The biological networks so obtained are extremely useful for investigating important data set properties, and their structure depends upon the threshold chosen which, in turn, is established only after examining various possible values and the corresponding networks.

Two possible numbers of clusters were obtained by performing both a hierarchical and a local optimizing clustering with the squared Euclidean distance function: the first found three clusters, while the other found eight. The rationale behind our selection of the cut-off values for finding the optimal number of patient clusters was due both to the results of Ye *et al.* (2014) and our critical examination of clinical variables.

In the case of patients, we chose a minimum correlation threshold of 0.93 which allowed us to 'separate' all patients into three well-defined groups (named Communities—see Fig. 1).

Three communities (Com1, Com2 and Com3) were matched for age and for percentage of males, but many other differences arise, as summarized in Table 1. Com1, in fact, is characterized by a higher degree of obesity with a high prevalence of morbid obesity, while subjects from Com3 are in general less obese or overweight (32.96 $\pm$ 5.01). In Com1 all indices of sleep disorder are higher (AHI, ODI and T90 and also more sleepiness at ESS). Moreover, Com1 shows a lower level of PaO_2 and worse values of $FEV_1\%$ and FVC%, even if both these values are within the normal range. Prevalence of comorbidities is generally higher in Com1 than in the other communities, except for cardiovascular diseases, which were more prevalent in Com3, but no statistical difference was obtained. Thus it is possible to summarize the characteristics of the three communities as follows.

- Community 1: (the largest one) consists of younger patients with very severe OSA (mean AHI = 65.91 $\pm$ 22.47), higher risk of nocturnal hypoxaemia (T90% = 44.97 $\pm$ 21.60), morbid obesity (BMI = 37.79 $\pm$ 7.42), low levels of PaO_2 (77.39 $\pm$ 11.64 mmHg) and high prevalence of comorbidity such as hypertension (64%).
- Community 2: patients with moderate–severe OSA (AHI = 28.88 $\pm$ 17.13) and low risk of nocturnal hypoxaemia (T90% = 4.86 $\pm$ 8.35).
- Community 3: (the smaller one) older patients, overweight or mild obesity (BMI = 32.96 $\pm$ 5.01), with very severe OSA (AHI = 57.26 $\pm$ 15.09) but with a low risk of nocturnal hypoxaemia (T90 = 11.58 $\pm$ 8.54) and less sleepy (ESS = 10.00 $\pm$ 4.77).

Clustering of variables

For variables, a threshold of 0.30 was chosen for outlining relationships among variables of Com1 and Com2, while 0.25 for Com3 these values were found after trying various correlation levels, so we chose a correlation pattern which offers a better rational explanation of possible interactions among the variables.

The analysis of the relationships among the variables shows some differences in the three communities. Even if the distribution of links among variables is almost the same in both communities correlations are lower, in general, in Com2 than in Com1. Moreover, some relationships are different. In
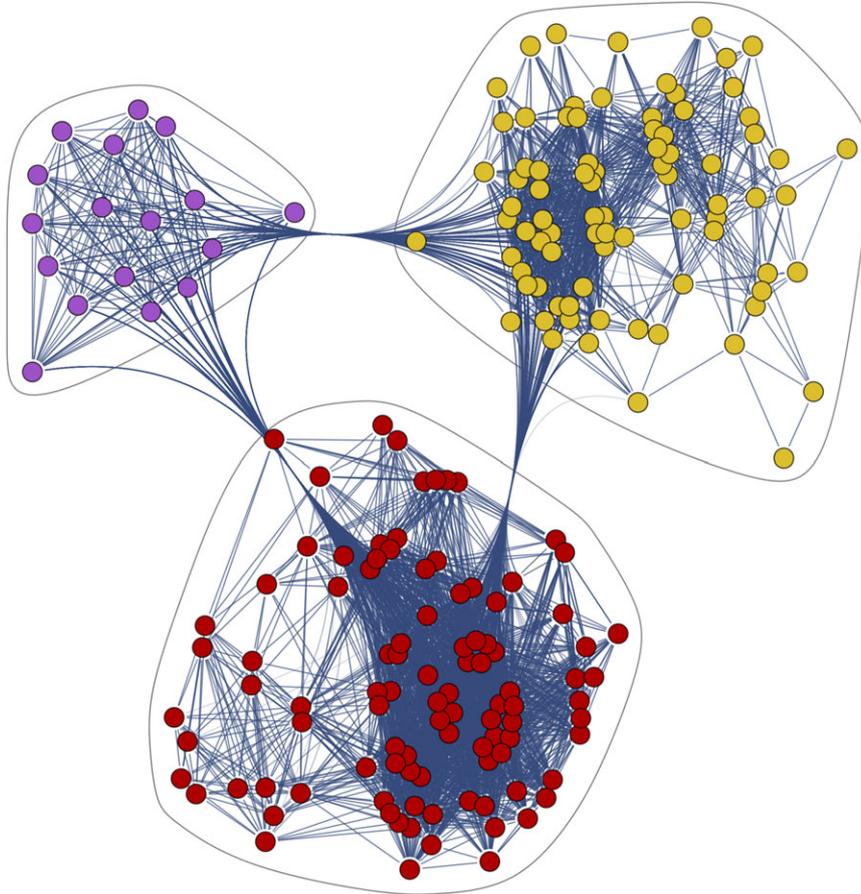
Figure 1. Distribution of patients in three communities. Distances between points are proportional to the grade of correlation.

particular, Fig. 2 shows the correlations between variables in each community. As can be seen, the values of PaO_2 in Com1 are related inversely strictly to BMI and T90, so high BMI and high T90 can induce low levels of PaO_2 during the day. In this community, age and respiratory functions do not seem to have a direct influence on oxygen levels. Conversely, in Com2—in which nocturnal desaturation was lower—PaO_2 levels are related inversely only to age. Finally, in Com3 as in Com1, BMI and T90% were correlated with diurnal PaO_2, but in this case spirometry performance also had a small role. Severity of sleep disorder is related to BMI in Com2, while weight does not seem to have any influence in Com1, in which only a slight inverse correlation was found with age, while in Com3 it was related with prevalence of cardiovascular diseases.

Sleepiness seems have no direct link with any variables in all communities.

DISCUSSION

In line with other recent studies, this work contributes to investigating the possible identification of subgroups of OSA patients by application of cluster analysis, which is a method of statistical analysis useful in clinical research. Our results confirm that, in a population of patients affected by OSA, it is possible to identify different clusters of patients (three in our case) having some common characteristics. However, there are main differences from the other two previous studies (Vavougios *et al.*, 2016; Ye *et al.*, 2014) as well as for the number of clusters found. First, we included arterial blood gases (ABGs) and spirometry in the analysis, which gave us more information about the clinical characteristics of patients. Secondly, we used PCA to improve selection criteria of variables and provide a clearer identification of the main parameters involved in the population clustering. Finally, we also analysed the relationships among variables in the different communities, so our work improves the knowledge about the correlation between OSA, comorbidities and other variables. Furthermore, exploratory correlation analysis demonstrates that in only a subset of OSA patients (i.e. Com1) is the presence of sleep breathing disorder the main factor in determining the PaO_2 diurnal level.

To our knowledge, the first study which used the cluster analysis approach on OSA was performed by Ye *et al.* (2014), who studied a cohort of the entire population of Iceland (822 patients) with mild to moderate OSA (AHI ≥ 15). In this work they identified three clusters of patients matched by age, BMI and OSA severity but divided according to their symptoms into 'disturbed sleep group' (cluster 1), 'minimally symptomatic group' (cluster 2) and 'excessive daytime

Table 1 Characteristics of the three communities of patients obtained after cluster analysis.

	Com1		Com2		Com3			
	Mean	SD	Mean	SD	Mean	SD	P	
n	99		82		17			
Sex (%)	77.00		85.00		82.00		0.34	
Age	53.33	13.23	54.16	11.29	56.24	12.20	0.66	
BMI	37.79	7.42	33.67	7.06	32.96	5.01	0.00	1 > (2 = 3)
ESS	12.69	5.02	10.94	6.08	10.00	4.77	0.04	1 > (2 = 3)
Smoke (%)	70.00		65.00		71.00		0.74	
AHI	65.91	22.47	28.88	17.13	57.26	15.09	0.00	(1 = 3) > 2
ODI	63.52	22.52	21.09	19.43	36.28	13.73	0.00	1 > 3 > 2
T90 (%)	44.97	21.60	4.86	8.35	11.58	8.54	0.00	1 > (2 = 3)
Cardiovascular diseases (%)	32.00		26.00		47.00		0.20	
Hypertension (%)	64.00		60.00		59.00		0.84	
Endocrinal disease (%)	15.00		13.00		6.00		0.59	
Diabetes (%)	20.00		20.00		18.00		0.97	
Asthma (%)	0.00		1.00		12.00		0.00	3 > (2 = 1)
FVC%	92.51	20.00	100.76	18.05	105.88	14.32	0.00	1 < (2 = 3)
FEV$_1$%	92.71	19.05	100.24	16.42	105.53	16.39	0.00	1 < (2 = 3)
Tiff	80.71	6.38	80.67	5.95	82.18	9.15	0.67	
pH	7.41	0.03	7.41	0.02	7.41	0.02	0.60	
PaO$_2$	77.39	11.64	82.57	11.72	82.22	13.80	0.01	3 = 1 < 2
PaCO$_2$	40.78	3.68	39.84	3.30	39.29	3.14	0.10	

Data are presented as mean $\pm$ standard deviation for continuous variables and as percentage of subjects for categorical variables. FVC, forced vital capacity; FEV$_1$, forced expiratory volume in 1 s; AHI, apnea–hypopnea index; BMI, body mass index; ESS, Epworth Sleepiness Scale; SD, standard deviation; ODI, oxygen desaturation index; PaO$_2$, partial pressure of oxygen; PaCO$_2$, partial pressure of carbon dioxide.
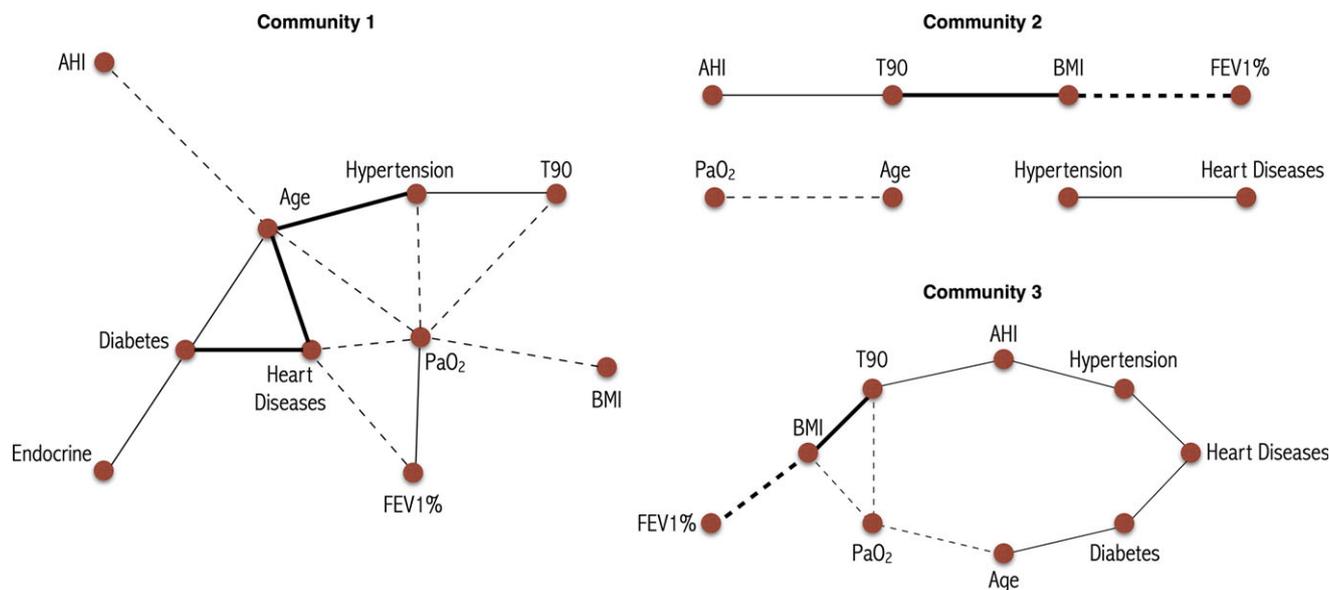

Figure 2. Correlation analysis of variables into communities. Only significant correlations are shown (P = 0.05). Positive correlations are continuous lines, while negative correlations are dashed. The thickness of each line is proportional to the level of correlation; thin line: $r < 0.40$, thick line: $r \geq 0.40$.

sleepiness' (cluster 3). The groups also differed regarding the presence of cardiovascular comorbidities, which were higher in cluster 2 and lower in cluster 3. More recently, a new study was performed at a Greek University (Vavougios *et al.*, 2016) on a larger population (1472) afferent to a sleep laboratory. In this case, six clusters of patients were found, and each was

different from the others for age, BMI, ESS and number of comorbidities evaluated by the Charlson Comorbidity Index.

Our results are more in line with this last study because, although we found only three clusters (due perhaps to the lower number of patients) in our work, the main differences between groups were also a consequence of severity of

OSA, BMI and ESS. Moreover, we also included some variables that were not considered in previous studies, such as spirometry performance and diurnal value of blood gas analysis. The results of these studies are extremely interesting, because they can help in focusing our attention on a detailed description of OSA patients. We usually classify OSA severity according to AHI (Iber *et al.*, 2007), and many studies have demonstrated that higher AHI is related to higher mortality and morbidity (Lavie and Lavie, 2008; Young *et al.*, 2008). Thus, this approach remains the simpler and more effective method for classifying OSA severity. However, the results of clustering seem to underline that adding more parameters can contribute to defining more clearly subjects with more severe disease than others.

Even if it is substantially intuitive that patients with higher AHI are clinically different from those with lower AHI, in some cases a moderate disease associated with higher sleepiness could be more deleterious than having a severe disease without any other risk factor, such as obesity or sleepiness (Drager *et al.*, 2013). The results of clustering could be useful to note this difference. According to these results we can hypothesize that severity of OSA also depends on other parameters, such as nocturnal hypoxaemia, which also have a strong impact in diurnal gas exchange and the presence of comorbidity. Knowledge of the main parameters involved in developing different grades of diseases could be useful to stratify the mortality and morbidity risk of OSA population, and therefore improve the approach to clinical management of these patients.

The correlation analysis between variables demonstrated that in Com1, AHI has no direct link with hypertension and others diseases, while T90 seems to contribute, directly or indirectly, to the development of hypertension and heart diseases. In Com2, severity of OSA is not related to any comorbidities, which are related substantially to age or to other factors, while in Com3 it is possible to find a positive correlation between AHI and hypertension, and anyone with other diseases. This only partially seems in contrast with respect to other studies. Different prospective studies reported that there is an independent association between OSA and cardiovascular events, suggesting that OSA is a risk factor for developing cardiometabolic dysregulation (Marin *et al.*, 2005). In the Wisconsin study and in the Sleep Heart Study the association between OSA and cardiovascular diseases was particularly true only in subjects with an AHI greater than 30 (Gottlieb *et al.*, 2010; Young *et al.*, 2009). Although our results cannot be compared with these studies, because the methods of analysis and the observed populations are completely different, they confirm substantially that subjects with more severe OSA, associated with other risk factors, have a higher prevalence of cardiovascular diseases. Perhaps the entity of nocturnal desaturation could have a stronger impact upon the development of comorbidities than the simply the number of apneas, and this is particularly true in younger patients (as in Com1). However, it is also important to underline that the vast majority of previous studies are prospective or observational, and have demonstrated that OSA is a risk factor for the development of cardiovascular diseases. Our study, instead, examined the association between OSA and prevalence of other comorbidities. Another interesting result of our study is the role that sleep disorder breathing has on diurnal gas exchange. In a previous study, our group demonstrated that in OSA patients the level of PaO_2 in daytime is influenced directly by age, BMI, FEV_1 and T90 (Lacedonia *et al.*, 2013). This observation is confirmed partially in this study, but is clarified more clearly. In fact, after identifying the communities of patients it seems clear that when the OSA is less severe (and in particular T90 was lower, as in Com2), age is the main factor for defining the level of daytime PaO_2. In contrast, in the case of more severe diseases (see Com1 and Com3), T90, BMI and FEV_1 contributes in different ways to PaO_2. This confirms, once again, that in the absence of lung diseases the severity of OSA directly influences the diurnal level of PaO_2, and in particular that nocturnal hypoxia is decisive for decreasing daytime PaO_2, and not vice versa (Fanfulla *et al.*, 2008).

The main limitation of cluster analysis is that the results are connected closely to the number and quality of variables selected. This means that the number of final clusters, also in the same population, can be different if the number and type of chosen variables are changed. Another limitation is that the majority of our population consisted of patients with moderate to very severe disease. This may have induced some bias into the analysis, because patients with mild OSA were not greatly represented. Thus, our results suggest only another possible way to phenotyping OSA patients—not necessarily the best one.

As in previous studies, another limitation of this work is that it is a retrospective study developed in a single centre. Larger and multi-centre analysis could provide more information about the phenotyping of the OSA population and increase the knowledge of the relationship between OSA and comorbidity.

The exact number of clusters in the OSA population is still unknown, as well as whether clustering can really be useful for changing our clinical approach to the management of OSA. However, our study demonstrated that by means of cluster analysis it is possible to gain more information concerning the characteristics of OSA patients, and to have a detailed image of the relationships among a large number of variables involved in the clinical presentation of OSA. The application of these new methods to the study of this complex disease could also be used in the future to develop some useful computing tools for disease phenotyping. It would also help the physician in the screening and follow-up of OSA patients in different ways according to their phenotype.

AUTHOR CONTRIBUTIONS

statistical analysis, MPFB contributed to critical review and final approval of the manuscript and CG performed the statistical analysis, supervised this project and participated in its coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

No conflicts of interest declared.

REFERENCES

Baumes, J., Goldberg, M. K., Krishnamoorthy, M. S., Magdon-Ismail, M. and Preston, N. Finding communities by clustering a graph into overlapping subgraphs. *IADIS AC*, 2005, 5: 97–104.

Dixon, J. B., Dixon, M. E., Anderson, M. L., Schachter, L. and O'Brien, P. E. Daytime sleepiness in the obese: Not as simple as obstructive sleep apnea. *Obesity*, 2007, 15: 2504–2511.

Drager, L. F., Togeiro, S. M., Polotsky, V. Y. and Lorenzi-Filho, G. Obstructive sleep apnea: a cardiometabolic risk in obesity and the metabolic syndrome. *J. Am. Coll. Cardiol.*, 2013, 62: 569–576.

Fanfulla, F., Grassi, M., Taurino, A. E., D'Artavilla-Lupo, N. and Trentin, R. The relationship of daytime hypoxemia and nocturnal hypoxia in obstructive sleep apnea syndrome. *Sleep*, 2008, 31: 249–255.

Gallo, C. and Capozzi, V. Clustering techniques for revealing gene expression patterns. In: M. Khosrow-Pour (Ed.) *Encyclopedia of Information Science and Technology*, 3rd edn. IGI Global, Hershey, PA, 2013: 438–447.

Gottlieb, D., Yenokyan, G., Newman, A. *et al*. Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure: the sleep heart health study. *Circulation*, 2010, 122: 352–360.

Iber, C., Ancoli-Israel, S., Chesson, A. and Quan, S. F.; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*, 1st edn. AASM, Westchester, 2007.

Jeong, H., Tombor, B., Albert, R., Oltval, Z. N. and Barabasi, A. L. The large-scale organization of metabolic networks. *Nature*, 2000, 407: 651–654.

Joosten, S. A., Hamza, K., Sands, S., Turton, A., Berger, P. and Hamilton, G. Phenotypes of patients with mild to moderate obstructive sleep apnea as confirmed by cluster analysis. *Respirology*, 2012, 17: 99–107.

Lacedonia, D., Carpagnano, G., Aliani, M. *et al*. Daytime pao_2 in OSAS, COPD and the combination of the two (overlap syndrome). *Respir. Med.*, 2013, 107: 310–316.

Lavie, P. and Lavie, L. Cardiovascular morbidity and mortality in obstructive sleep apnea. *Curr. Pharm. Des.*, 2008, 14: 3466–3473.

Marin, J. M., Carrizo, S. J., Vicente, E. and Augusti, A. G. Long-term cardiovascular outcomes in men with obstructive sleep apnea–hypopnea with or without treatment with continuous positive airway pressure: an observational study. *Lancet*, 2005, 365: 1046–1053.

McNicholas, W. and Bonsigore, M. Sleep apnea as an independent risk factor for cardiovascular disease: current evidence, basic mechanisms and research priorities. *Eur. Respir. J.*, 2007, 29: 156–178.

Nieto, F., Peppard, P. and Young, T. Sleep disordered breathing and metabolic syndrome. *WMJ*, 2009, 108: 263–265.

Peppard, P. E., Young, T. B., Barnet, J. H., Palta, M., Hagen, E. W. and Hla, K. M. Increased prevalence of sleep-disordered breathing in adults. *Am. J. Epidemiol.*, 2013, 177: 1006–1014.

Robichaud-Hallé, L., Beaudry, M. and Fortin, M. Obstructive sleep apnea and multimorbidity. *BMC Pulm Med.*, 2012, 12: 60.

Tipping, M. E. and Christopher, M. B. Probabilistic principal component analysis. *J. R. Stat. Soc. Series B Stat. Methodol.*, 1999, 61: 611–622.

Vavougios, G. D., Natsios, G., Pastaka, C., Zarogiannis, S. G. and Gourgoulianis, K. I. Phenotypes of comorbidity in OSAS patients: combining categorical principal component analysis with cluster analysis. *J. Sleep Res.*, 2016, 25: 31–38.

Ye, L., Pien, G., Ratcliffe, S. *et al*. The different clinical faces of obstructive sleep apnea: a cluster analysis. *Eur. Respir. J.*, 2014, 44: 1600–1617.

Young, T. B., Finn, L., Peppard, P. E. *et al*. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep*, 2008, 31: 1071–1078.

Young, T. B., Palta, M., Dempsey, J., Peppard, P. E., Nieto, F. J. and Hla, K. M. Burden of sleep apnea: rationale, design and major findings of the Wisconsin sleep cohort study. *WMJ*, 2009, 108: 246–249.