

Can we avoid unnecessary polysomnographies in the diagnosis of Obstructive Sleep Apnea? A Bayesian network decision support tool

Liliana Leite, Cristina Costa-Santos

CINTESIS

Faculty of Medicine of the University of Porto
Porto, Portugal

lilianappleite@gmail.com, csantos@med.up.pt

Pedro Pereira Rodrigues

CINTESIS & LIAAD - INESC TEC

Faculty of Medicine of the University of Porto
Porto, Portugal

pprodrigues@med.up.pt

Abstract—Obstructive Sleep Apnea (OSA) affects 2-4% of the population worldwide. The standard test for OSA diagnosis is polysomnography (PSG), an expensive exam limited to urban areas. Furthermore, nearly half of all PSG tests results are negative for OSA. This work aims to reduce these unnecessary exams, by defining an auxiliary diagnostic method that could be used to assess patient's need for PSG, according to their probability of OSA diagnosis. A prospective study was conducted on adult patients with OSA suspicion who performed PSG at our sleep laboratory in Portugal. The studied clinical variables were defined after literature review and collected during consultation. Two comparable cohorts were studied for derivation (n=86) and validation (n=33) of models. Three classifiers were analyzed - a multiple logistic regression classifier (AUC=80.0%) and two Bayesian networks classifiers - Naïve Bayes (AUC=81.3%) and Tree Augmented Naïve Bayes (TAN, AUC=81.4%) - aiming at the best possible specificity (identification of unnecessary exams). Overall, sensitivity-adjusted models could detect normal patients, preventing unnecessary PSG, while keeping sensitivity high. Furthermore, the graphical representation of TAN can be explored by the physician during consultation, making it a helpful tool to assess patients' need to perform PSG.

Keywords—obstructive sleep apnea; diagnosis; clinical model; Bayesian network.

I. BACKGROUND

Obstructive Sleep Apnea (OSA) is a disease that affects approximately 4% of men and 2% of women worldwide but is still underestimated and underdiagnosed [1]. It is characterized by episodes of breathing cessation (apnea) or reduction in airflow (hypopnea) during sleep for at least 10 seconds as a result of upper airway collapse [2]. The severity of OSA is associated with the apnea-hypopnea index (AHI), documented during sleep, which can be divided into mild ($5 \leq \text{AHI} < 15$), moderate ($15 \leq \text{AHI} < 30$) and severe ($\text{AHI} \geq 30$) [3]. The standard method for assessing this index, and therefore defining the OSA diagnosis, is polysomnography (PSG). However, it is time-consuming, expensive and relatively limited to urban areas which, consequently, originates high waiting lists [4].

There is on literature a large variability about the most predictive factors that indicate a higher or lower probability of OSA. Many studies are in concordance in some factors like age, gender and body mass index (BMI) in OSA patients [1,5-9]. However, some authors referred other features such as neck circumference (NC), witnessed apneas or diurnal somnolence as important risk factors too [2,5,10]. The path for a consensus is still undetermined.

In Portugal, patients are referred by the primary care physician to a sleep consult, and then the sleep expert physicians decide the need to perform polysomnography. Although patients are screened by the physicians, based on clinical factors, the specificity of the entire process is rather low (48% of PSG performed in 2010, in our sleep laboratory, resulted negative for OSA) which, together with the limited availability of the service, yields long waiting lists both for consultation and to perform PSG. This problem is also prevalent in other sleep laboratories and several studies have been conducted construct prediction models or questionnaires, to determine the probability of having OSA, and thereby reduce and optimize the number of patients that realize PSG, assigning different priority to patients [5,11-14]. The identification of the most important risk factors is the key to build successful prediction models.

Traditionally, these models were based on questionnaires, such as the Berlin Questionnaire (BQ) or the Epworth Somnolence Scale (ESS), or on simple decision rules, the prognostic score and the classification of patients into different risk categories. This score is often based on the combination of clinical variables and was built for the general population, as well as for specific groups. However, existing studies did not fill all the requirements to enable its use in clinical practice, such as their validation or sensitivity [5,10,12,14-16]. These models need high sensitivity, avoiding false negatives, in order to prevent excluding a patient with moderate or severe OSA from performing PSG. No study was found that was fitted for 100% sensitivity and their results in clinical practice are still questionable. Vaz et al. used the BQ, one of the most recognized screening tool, to screen patients with OSA in a sleep breathing clinic in Portugal [17]. The authors achieve a sensitivity of 65.2% and specificity of 80%, what reveals a good discrimination but poor performance in OSA identification. These results are similar to other studies that

have different performances and do not reveal BQ as an alternative to screen patients [6,17].

Recently, many studies have been conducted to apply machine learning methods for medical knowledge discovery, including sleep medicine, consisting on an alternative to traditional statistic in defining diagnostic models [4,18,19]. These models can now be generated by artificial intelligence, using decision trees, neural networks, support vector machines and Bayesian networks (BN) [18,20]. To become useful, models must have certain characteristics, such as good performance, good ability to handle data entry errors or omissions, transparency of diagnostic knowledge, ability to explain decisions, and the algorithm should be able to reduce the number of tests needed to make a reliable diagnosis [15]. Overall, we did not find a valid method to screen patients with OSA in Portuguese sleep laboratories. Hence, the main objective of this work is to define an auxiliary diagnostic method, that prioritizes patients during pre-polysomnography consultation, according to their probability of OSA diagnosis. The auxiliary diagnostic method should avoid the false negatives, i.e. a patient with OSA with recommendation to not perform polysomnography, thus aiming at 100% sensitivity.

II. METHODS

A literature review was performed to define the relevant variables to be collected. Data was prospectively collected from patients in two periods of time. A derivation cohort was used to build logistic regression (LR) and BN models. Results from PSG were categorized into normal or OSA (independently of the severity) and we validate the final models, LR and BN, on a validation cohort.

A. Variables

The selection of variables was based on a literature search on Pubmed on the 27th of September of 2011, using the Mesh terms and keywords "risk factors", "obstructive sleep apnea" and "diagnosis". From the 442 articles retrieved, 127 were reviews considered for inclusion. After reading the title and abstract, we excluded 82 articles because they studied children, pregnant women, OSA as a secondary problem, only women or elderly, or were not related to risk factors. From these articles, 33 variables have been selected and collected in the study, including demographic information, clinical history, physical examination and co-morbidities information.

B. Patients

This study included patients referred to perform PSG at the Sleep Laboratory of Vila Nova de Gaia/Espinho Hospital Center, Portugal. We collected two samples with the patients that agreed to participate in the study, one for the development of the model and a second one to test the proposed model. The first sample, the derivation cohort, includes patients that realized PSG between December of 2011 to February of 2012 while the second, the validation cohort, between April and May of 2012. All adults, older than 18 years, referred by the physicians with suspected OSA were included. In case of duplicate studies from the same patient, the one with best sleep efficiency was selected. Patients with suspicion of another

disorder than OSA, patients already diagnosed (therapeutic studies), and patients with severe lung disease or neurological condition that somehow affects the respiratory function, such as neuromuscular diseases, were excluded. This study was approved by the ethics commission of the Vila Nova de Gaia/Espinho Hospital Center, and patients were not exposed to any adverse effects.

C. Data collection

Clinical information was collected prospectively during consultation, 3 months before PSG. However, all variables were checked to avoid missing information and if some variable was missing, it would be collected during PSG. For the PSG, the parameters, settings, filters, technical specifications, sleep stage, event scoring and final results were applied according to the American Academy of Sleep Medicine rules of 2007. For this study, the outcome measure was the clinical diagnosis supported on PSG results, dichotomized into normal or OSA (mild, moderate and severe). Data quality was checked and validated before analysis to avoid duplicate records and errors on data.

D. Evaluation methodology

Receiver Operating Characteristic (ROC) curve analysis was performed to determine prediction error and area under the curve (AUC), and to achieve the thresholds for our models. Different thresholds were used to achieve two levels of sensitivity: 100% and 95%. The achieved models were evaluated with sensitivity and specificity estimates on both the derivation and validation cohorts. The two cohorts were compared using t-test (for continuous variables) and chi-square test (for categorical variables). All the analysis was performed with SPSS statistical software (SPSS, Inc, Chicago, IL, USA), unless otherwise specified. A significance level of 5% was used. Final sensitivity and specificity plots were built using RevMan 5.1.

E. Logistic regression model

Odds ratios (OR) and corresponding 95% confidence intervals were calculated using univariate LR for the 33 variables as independent variables and the presence of OSA as dependent variable. The independent variables with significant OR were used in the multiple forward conditional logistic regression analysis to construct the model.

F. Bayesian network models

Models were built using only the significant variables achieved on the univariate LR. Missing data was removed and continuous variables BMI, NC and Abdominal circumference (AC) were categorized into dichotomous variables: BMI into normal or obese (threshold: 30), NC and AC into normal or increased (thresholds NC: male 42cm, female 37cm; AC: male 94cm, female 80cm), according to the literature [5,21]. For exploration purposes, a BN was built using a hill-climbing strategy. For classification purposes, the result of PSG (normal or OSA) was defined as the class attribute to construct the models. The BN classifiers used to build the models were Naïve Bayes (NB) and Tree Augmented Naïve Bayes (TAN), given their previous good results in other clinical domains

[22-24]. We used R to learn the exploratory network, using package BNLEARN [25], Rapidminer was used to build and evaluate NB and TAN, and to check conditional probabilities given the class outcome of PSG (normal or OSA), and SamIam was used to inspect and consult the BNs.

III. RESULTS

From the 113 patients considered for inclusion in the derivation cohort, 27 were excluded (18 w/ other pathology, 3 children, 3 no information, 2 already diagnosed, 1 neuromuscular disease). We collected data from 86 patients, 69 (80%) of which were male and mean age was 56 years. Forty one patients (48%) had normal result with mean age of 54 years; of the 45 patients with OSA (52%), 17 (37%) were categorized into mild, 15 (33%) were moderate and 13 (30%) were severe, and the mean age was 57 years. To test the performance of the model in clinical practice, we collected a second cohort with 33 patients: 18 (55%) had OSA, 25 (76%) were male, with mean age of 53 years. Of the 18 patients with OSA, 6 (33%) were categorized into mild, 7 (39%) were moderate and 5 (28%) were severe.

A. Univariate regression analysis

The analysis of univariate regression (Table I) showed 6 variables with significant OR: male gender (7.259), witnessed apneas (4.725), consume of alcohol before sleep (3.307), NC (1.341), BMI (1.159), and AC (1.076). Statistical testing showed that the two cohorts (validation vs derivation) were comparable with respect to the outcome and these factors: OSA (55% vs 52%, $p=0.833$), gender male (76% vs 80%, $p=0.634$), witnessed apneas (64% vs 65%, $p=0.469$), alcohol before sleep (52% vs 44%, $p=0.466$), NC (41cm vs 41cm, $p=0.879$), augmented NC (49% vs 43%, $p=0.664$), BMI (29 vs 29, $p=0.839$), BMI obese (37% vs 36%, $p=0.956$), AC (105cm vs 105cm, 0.744) and augmented AC (94% vs 90%, 0.723).

B. Logistic regression model

A multiple forward conditional logistic regression analysis was created with NC, gender, witnessed apneas and consume of alcohol before sleep. AC and BMI variables were not considered for this model given their high co-linearity with the strongest variable, NC. After two steps, NC and witnessed apneas were the final included variables in the multivariate regression, with intercept -11.147 and coefficients 0.256 (OR=1.292) and 1.134 (OR=3.108), respectively. ROC curve analysis, on the derivation cohort, showed an AUC of 80.0%.

C. Bayesian network model and classifiers

Figure 1 presents the built Bayesian network and the conditional probabilities of the 6 significant variables. As found on multiple LR, there is a high association between AC, BMI and NC. BMI influences NC and AC: an obese patient is certain to have AC increased, $P(AC|Obese)=1.00$, and a higher probability of having NC increased, $P(NC|Obese)=0.84$, when compared with a non-obese, $P(NC|\neg Obese)=0.19$, while AC probability is still high for non-obese patients, $P(AC|\neg Obese)=0.85$. NC influences OSA and alcohol before sleep: $P(OSA|NC)=0.75$ and $P(Alcohol|Male,NC)=0.73$, when

compared with a patient with normal NC, $P(OSA|\neg NC)=0.34$, $P(Alcohol|Male,\neg NC)=0.36$. Also, OSA influences gender: $P(Male|OSA)=0.93$, while $P(Male|\neg OSA)=0.65$. Of course, caution is advised in interpreting such associations as causations. Gender influences WA and alcohol before sleep, with the presence of these two variables more likely in men, $P(WA|Male)=0.74$ and $P(Alcohol|Male,NC)=0.73$.

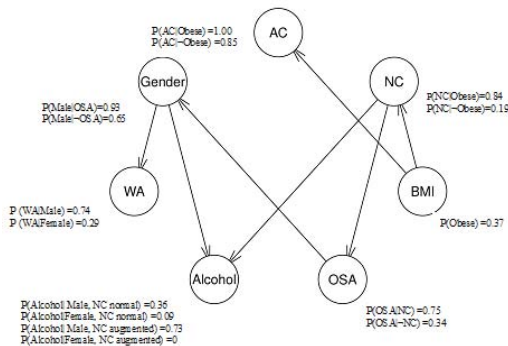


Fig. 1. Exploratory Bayesian network

The NB classifier was built with OSA as class attribute, with remaining variables being considered conditionally independent given the class. The TAN classifier resulted in BMI becoming an ancestor of both AC and NC, AC ancestor of Gender, and Gender ancestor of both alcohol before sleep and witnessed apneas. The ROC curve analysis, on the derivation cohort, demonstrated an AUC of 81.3% and 81.4%, respectively for NB and TAN.

D. Classifier validation

All three classifiers were evaluated using the simple threshold of 50%. Additionally, using ROC analysis from the derivation cohort, conservative thresholds (aiming at 100% sensitivity) were defined (LR: 10%, NB: 7%, TAN: 2%) along with less restrictive (aiming at 95% sensitivity) ones (LR: 25%, NB: 10%, TAN: 22%). Derivation cohort analysis is presented in Figure 2. Simple thresholds resulted in values for sensitivity-specificity of 71%-73% (LR), 70%-73% (NB) and 70%-73% (TAN). Conservative thresholds favored BNs with specificity of 5% (LR), 25% (NB) and 28% (TAN) while the less restrictive thresholds yielded 35% (LR), 33% (NB) and 38% (TAN) specificity.

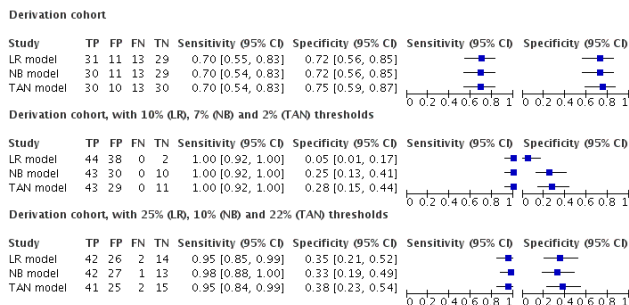


Fig. 2. Sensitivity and specificity for the three models (derivation cohort)

Figure 3 presents the application of the models to the validation cohort. Naïve threshold yielded sensitivity-specificity values of 56%-39% (LR), 56%-47% (NB) and 72%-53% (TAN), clearly favoring TAN. For both conservative and less restrictive thresholds, results were: 100%-0% and 88%-15% (LR), 94%-0% and 89%-7% (NB), 94%-7% and 89%-13% (TAN), respectively.

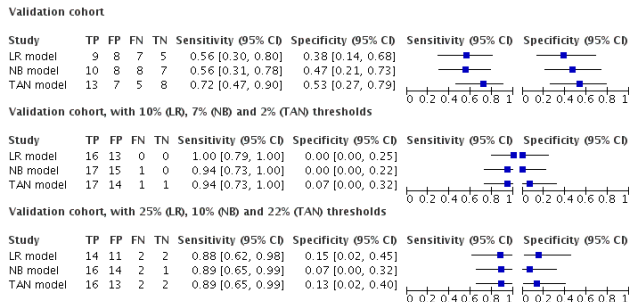


Fig. 3. Sensitivity and specificity for the three models (validation cohort)

IV. DISCUSSION

The initial step towards defining a decision support system is to assess the need for it. In the derivation cohort, the proportion of normal results was 48%, which reveals a large number of normal result exams, as we expected, hence supporting the need for a good clinical decision support system.

Although the large number of clinical variables used (33), only six had significant results in the univariate analysis: body mass index, neck circumference, abdominal circumference, gender, witnessed apneas and consume of alcohol before sleep. Even though in some studies age appears as a risk factor, in our study it did not end up having significant OR on the univariate analysis. The multivariate model showed that the most discriminative predictive factors for the presence of OSA were the neck circumference and witnessed apneas, which is in accordance with other studies, and supports the inability of some other models that were not build using these factors, in proper prediction of OSA.

The developed models were fitted for 100% sensitivity to avoid false negatives. It was our goal to reduce the number of normal result exams but we do not want to leave patients with moderate or severe OSA without performing PSG. However, other thresholds were considered to allow for multiple interpretations of results.

When we tested our models in the validation cohort, the LR model using the 10% threshold avoids false negatives but does not reduce the number of normal patients requested to perform PSG. However, this was an expected result given the reduced size of the sample (5% specificity over 15 normal patients represents less than one patient). Using a higher threshold (25%), useful for hospitals with sleep consultation but without sleep laboratory (which have to choose very well patients to refer to other laboratories), the result was good for specificity (15%) but with loss of sensitivity, 88%. Nevertheless, as further explained, we believe that the two OSA patients missed

by the model do not hinder the validity of the proposal. First, since the threshold was optimized for 95% sensitivity, we would expect one OSA patient to be missed (in this case, it was a mild OSA patient). Then, the severe OSA patient missed by the model actually presented negative values for the two included variables (witnessed apneas and NC), and could thus be considered an outlier for our modelling. Nonetheless, this case should be further considered in future work.

Given the good indications given by the LR model, the significant variables achieved in univariate analysis were used to create the BN using NB and TAN classifiers. On all validation results, TAN reveals better specificity than NB, for similar sensitivity. When we applied our BN models to the validation cohort, the results for sensitivity were the same on NB and TAN (94% and 89%) using the two types of threshold, but the results of specificity were better on TAN (7% and 13%) in the two alternatives when compared with NB (0% and 7%).

These different thresholds for the different models could be used to assign different priorities to patients, assuming for example levels of high risk, low risk and no risk of OSA diagnosis. The levels could be defined separately for a single model (e.g. TAN) or as an ensemble of models (e.g. TAN and LR). These could then be used in a clinical decision support tool applicable in sleep consultation.

A. Inference

To explain how the BN could be implemented in a clinical decision support system, Figure 4 presents the interaction between the physician and the learned TAN network. Observed variables (evidence) are marked red (100%), while unobserved variables present the posterior probability for each category of the variable.

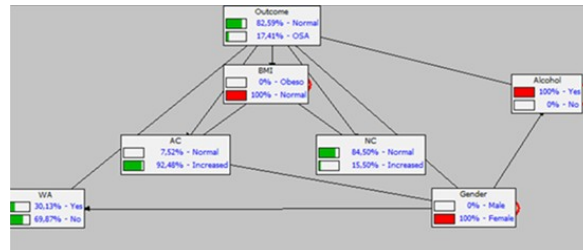


Fig. 4. Example of the inference process for a given patient using TAN

Assuming a new patient arrives at consultation, female, with normal BMI but with history of alcohol consumption before sleep, the physician can check that the probability of OSA diagnosis is only 17.41%. Even with the lack of information of the remaining three variables, the model can give a result, which rises as a great advantage compared to LR models, where all factors need to be observed.

This clearly shows the importance of these models, dealing without information of more than one variable, when we do not have the possibility of measuring all variables. For example, to use the LR model we need to have access to the two variables included on the regression equation: NC, which is easy to measure, and WA, which is much more difficult to assess and, in some cases, even impossible to get (e.g. if the patient does not have a bed partner). Besides the advantages described above

on dealing with missing information, the graphical representation is another capital gain, mainly using the TAN model, since it shows the interaction of more than one dependence between the variables. This graphical representation is usually well understood and adhered to by the physicians in real clinical practice.

B. Limitations

Some factors were not possible to assess due to the lack of representativeness in the sample (ethnicity, snore, decreased libido, pulmonary hypertension, congestive heart failure, metabolic syndrome, renal failure and hypothyroidism) which may have led to somewhat biased models. Also, since our study was conducted on patients referred by primary care physicians to the sleep consult, the prevalence of OSA in our sample (52%) was higher than of general populations. Hence, no OSA prevalence estimate can be inferred. Also, some authors questioned if AHI is the more accurate measure to classify OSA severity; some suggest the use of RDI in alternative to AHI. This could have also biased our results, although since most of our study is based on a dichotomous outcome (OSA or Normal) this should have had a minor influence. To recode continuous variables (NC, AC and BMI) we have used literature-based measures but there are actually no standard values to categorize values into normal or altered, so these may lead to some errors in borderline characteristics. For example, in this study we considered BMI values higher than 30 to be obese, where some authors could consider a lower threshold at 25 (pre-obese). To recode NC and AC we chose the most referred and consensual criteria on literature, but again others could exist.

V. CONCLUSIONS

In this study, the main factors found for OSA diagnosis were male gender, witnessed apneas (WA), consumption of alcohol before sleep, neck circumference (NC), body mass index (BMI), and abdominal circumference (AC). We used two different techniques to construct the diagnostic models, one based on logistic regression (LR) and another based on Bayesian networks (BN). With LR, the final model used only two variables on the regression equation: NC and WA. The great limitation on the application of this approach is the use of WA, since it is subjective and in some cases impossible to measure.

The great advantages of BN are the fact that: a) they can deal with missing information, b) the graphical representation shows the relationship among variables, and c) they allow the inspection of the values of probabilities given patient's characteristics. Also, this representation based on conditional probabilities (hence, risks) can be used as an alternative to the traditional statistical measures for this type of studies.

As we did not find a validated model, tested in Portuguese sleep laboratories, we think that our models can be incorporated, in the future, in a valid method to screen patients with suspicion of OSA, before performing PSG. Eventually, we can reduce the number of normal result exams, optimizing the available resources and making sure that no severe case waits much time for the exam and, consequently, treatment.

ACKNOWLEDGMENT

The authors acknowledge the financial support of the Master Course in Medical Informatics and CINTESIS – Faculty of Medicine of the University of Porto.

REFERENCES

- [1] Al Lawati NM, Patel SR, Ayas NT: Epidemiology, Risk Factors, and Consequences of Obstructive Sleep Apnea and Short Sleep Duration. *Progress in Cardiovascular Diseases* 2009, 51:285-293.
- [2] Redline S, Budhiraja R, Kapur V, Marcus CL, Mateika JH, Mehra R, Parthasarthy S, Somers VK, Strohl KP, Sulit LG, et al: The scoring of respiratory events in sleep: reliability and validity. *J Clin Sleep Med* 2007, 3:169-200.
- [3] Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep* 1999, 22:667-689.
- [4] Sun L, Chiu H-W, Chuang C, Liu L: A prediction model based on an artificial intelligence system for moderate to severe obstructive sleep apnea. *Sleep and Breathing* 2010:1-7.
- [5] Davies RJ, Ali NJ, Stradling JR: Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome. *Thorax* 1992, 47:101-105.
- [6] Doghranji PP: Recognition of obstructive sleep apnea and associated excessive sleepiness in primary care. *J Fam Pract* 2008, 57:S17-23.
- [7] Kapur VK: Obstructive sleep apnea: diagnosis, epidemiology, and economics. *Respir Care* 2010, 55:1155-1167.
- [8] Kohler M: Risk factors and treatment for obstructive sleep apnea amongst obese children and adults. *Curr Opin Allergy Clin Immunol* 2009, 9:4-9.
- [9] Young T, Skatrud J, Peppard PE: Risk factors for obstructive sleep apnea in adults. *JAMA* 2004, 291:2013-2016.
- [10] Young T, Shahar E, Nieto FJ, Redline S, Newman AB, Gottlieb DJ, Walsleben JA, Finn L, Enright P, Samet JM, for the Sleep Heart Health Study Research Group: Predictors of Sleep-Disordered Breathing in Community-Dwelling Adults: The Sleep Heart Health Study. *Arch Intern Med* 2002, 162:893-900.
- [11] Pouliot Z, Peters M, Neufeld H, Kryger MH: Using self-reported questionnaire data to prioritize OSA patients for polysomnography. *Sleep* 1997, 20:232-236.
- [12] Flemons WW, Whitelaw WA, Brant R, Remmers JE: Likelihood ratios for a sleep apnea clinical prediction rule. *American Journal of Respiratory and Critical Care Medicine* 1994, 150:1279-1285.
- [13] Gurubhagavatula I, Maislin G, Pack AI: An Algorithm to Stratify Sleep Apnea Risk in a Sleep Disorders Clinic Population. *Am J Respir Crit Care Med* 2001, 164:1904-1909.
- [14] Hoffstein V, Szalai JP: Predictive value of clinical features in diagnosing obstructive sleep apnea. *Sleep* 1993, 16:118-122.
- [15] Kononenko I: Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* 2001, 23:89-109.
- [16] Viner S, Szalai JP, Hoffstein V: Are history and physical examination a good screening test for sleep apnea? *Annals of Internal Medicine* 1991, 115:356-359.
- [17] Vaz AP, Drummond M, Mota PC, Severo M, Almeida J, Winck JC: Translation of Berlin Questionnaire to Portuguese language and its application in OSA identification in a sleep disordered breathing clinic. *Rev Port Pneumol* 2011, 17:59-65.
- [18] van Gerven MAJ, Taal BG, Lucas PJF: Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics* 2008, 41:515-529.
- [19] Tu JV: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996, 49:1225-1231.

[20] Lee SM, Abbott PA: Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers. *J Biomed Inform* 2003, 36:389-399.

[21] Direcção geral da saúde (Plataforma contra a obesidade)

[22] Burnside E, Rubin D, Shachter R: A Bayesian network for mammography. *Proc AMIA Symp* 2000:106-110.

[23] Montironi R, Bartels PH, Thompson D, Scarpelli M, Hamilton PW: prostatic intraepithelial neoplasia (PIN). Performance of Bayesian belief network for diagnosis and grading. *J Pathol* 1995, 177:153-162.

[24] Sakellariopoulos GC, Nikiforidis GC: Development of a Bayesian Network for the prognosis of head injuries using graphical model selection techniques. *Methods Inf Med* 1999, 38:37-42.

[25] Scutari M: Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software* 2010, 35:1-20.

TABLE I. DESCRIPTION AND ODDS RATIOS FOR THE 33 STUDIED VARIABLES

| | Normal n=41 | OSA n=45 | simple OR | 95%CI |
|---|----------------|-------------|--------------|-----------------------|
| <i>Gender, n (%)</i> | | | | |
| Female | 14 (34) | 3 (6) | Ref. | |
| Male | 27 (66) | 42 (93) | 7.259 | [1.096;27.651] |
| <i>Ethnicity, n (%)</i> | | | | |
| European | 40 (98) | 44 (98) | - | - |
| African | 1 (2) | 1 (2) | - | - |
| <i>Age, mean (sd)</i> | 54 (14) | 57 (13) | 1.020 | [0.988;1.052] |
| <i>Snore, n (%)</i> | 41 (100) | 45 (100) | - | - |
| <i>Witnessed apneas, n (%)</i> | 20 (49) | 36 (82) | 4.725 | [1.772;12.599] |
| <i>Gasping/Shocking, n (%)</i> | 7 (17) | 14 (31) | 2.194 | [0.783;6.142] |
| <i>Motor Vehicle Crashes, n (%)</i> | 3 (8) | 3 (7) | 0.872 | [0.165;4.608] |
| <i>Refreshing Sleep, n (%)</i> | 17 (41) | 22 (49) | 1.350 | [0.575;3.169] |
| <i>Humor alterations, n (%)</i> | 2 (5) | 2 (4) | 0.907 | [0.122;6.751] |
| <i>Nocturia, n (%)</i> | 16 (39) | 16 (36) | 0.862 | [0.359;2.069] |
| <i>Restless Sleep, n (%)</i> | 4 (10) | 9 (20) | 2.312 | [0.653;8.185] |
| <i>Decreased libido, n (%)</i> | 0 (0) | 1 (2) | - | - |
| <i>Morning headaches, n (%)</i> | 10 (24) | 8 (18) | 0.670 | [0.236;1.906] |
| <i>Alcohol before sleep, n (%)</i> | 12 (29) | 26 (58) | 3.307 | [1.350;8.100] |
| <i>Smoker, n (%)</i> | | | | |
| No | 22 (54) | 25 (56) | Ref. | |
| Yes | 10 (24) | 7 (16) | 0.616 | [0.200;1.894] |
| Ex-smoker | 9 (22) | 13 (29) | 1.271 | [0.456;3.543] |
| <i>Sedative use, n (%)</i> | 8 (18) | 9 (20) | 1.031 | [0.356;2.986] |
| <i>ESS, median (range)</i> | 8 (19) | 8 (24) | 0.980 | [0.908;1.050] |
| <i>Concentration decrease, n (%)</i> | 8 (19) | 3 (7) | 0.295 | [0.072;1.198] |
| <i>BMI, mean (sd)</i> | 28 (4) | 30 (5) | 1.159 | [1.030;1.303] |
| <i>NC, mean (sd)</i> | 39 (3.4) | 43 (3.8) | 1.341 | [1.159;1.550] |
| <i>AC, mean (sd)</i> | 100 (10) | 108 (12) | 1.076 | [1.025;1.129] |
| <i>Craniofacial and upper airway abnormalities, n (%)</i> | 17 (41) | 28 (62) | 2.325 | [0.979;5.526] |
| <i>Atrial fibrillation, n (%)</i> | 1 (2) | 1 (2) | 0.909 | [0.055;15.020] |
| <i>Stroke, n (%)</i> | 1 (2) | 2 (4) | 1.860 | [0.162;21.319] |
| <i>Myocardial infarction, n (%)</i> | 4 (9) | 2 (4) | 0.430 | [0.075;2.484] |
| <i>Pulmonary hypertension, n (%)</i> | 0 (0) | 0 (0) | - | - |
| <i>Congestive heart failure, n (%)</i> | 1 (2) | 0 (0) | - | - |
| <i>Diabetes, n (%)</i> | 8 (19) | 9 (20) | 1.031 | [0.356;2.986] |
| <i>Metabolic Syndrome, n (%)</i> | 0 (0) | 0 (0) | - | - |
| <i>Renal failure, n (%)</i> | 0 | 0 | - | - |
| <i>Hypothyroidism, n (%)</i> | 3 (7) | 0 | - | - |
| <i>Gastroesophageal reflux disease, n (%)</i> | 2 (5) | 3 (7) | 1.393 | [0.221;8.783] |
| <i>Hypertension, n (%)</i> | 21 (51) | 22 (49) | 0.911 | [0.391;2.124] |

BMI: Body Mass Index; NC: Neck circumference; AC: Abdominal circumference; ESS: Epworth Somnolence Scale; OR: Odds Ratio; CI: Confidence Interval