Title: Neural Network Prediction of Obstructive Sleep Apnea From Clinical Criteria(*)
Author(s): Wayne Danter , P Eng , Kathleen A. Ferguson , Tanya Francovic , Charles F. P. George and Simon D. Kirby
Source: *Chest*. 116.2 (Aug. 1999): p409.
Document Type: Article
Full Text:

Study objectives: Clinical prediction models for the diagnosis of obstructive sleep apnea (OSA) have lacked the accuracy necessary to confidently replace polysomnography (PSG). Artificial neural networks are computer programs that can be trained to predict outcomes based on experience. This study was conducted to test the hypothesis that a generalized regression neural network (GRNN) could accurately classify patients with OSA from clinical data.

Study design: Retrospective review.

Setting: Regional sleep referral center.

Patients: Randomly selected records of patients referred for possible OSA.

Measurements: The neural network was trained using 23 clinical variables from 255 patients, and the predictive performance was evaluated using 150 other patients.

Results: The prevalence of OSA in this series of 405 patients (293 men and 112 women) was 69%. The trained GRNN had an accuracy of 91.3% (95% confidence interval [CI], 86.8 to 95.8). The sensitivity was 98.9% for having OSA (95% CI, 96.7 to 100), and the specificity was 80% (95% CI, 70 to 90). The positive predictive value that the patient would have OSA was 88.1% (95% CI, 81.8 to 94.4), whereas the negative predictive value that the patient would not have OSA (if so classified) was 98% (95% CI, 94 to 100).

Conclusions: Appropriately trained GRNN has the ability to accurately rule in OSA from clinical data, and GRNN did not misclassify patients with moderate to severe OSA. In this study, rise of the neural network could have reduced the number of PSG studies performed. Prospective validation of the neural network for the diagnosis of OSA is now required. (CHEST 1999; 116:409-415)

Key words: artificial neural networks; clinical prediction models; obstructive sleep apnea; screening

Abbreviations: AHI = apnea plus hypopnea index; ANN = artificial neural network; BMI = body mass index; CI = confidence interval; GRNN = generalized regression neural network; NC = neck circumference; NPV = negative predictive value; OSA = obstructive sleep apnea; PPV = positive predictive value; PSG = polysomnography

Obstructive sleep apnea (OSA) is a common clinical condition affecting 2 to 4% of middle-aged women and men.[1] It can lead to numerous complications[2] including decreased quality of life,[3] reduced occupational performance,[4] and increased risk of a motor vehicle accidents.[5] When untreated, significant OSA is associated with increased mortality largely due to cardiovascular causes.[6] The diagnosis of OSA or, more precisely, OSA syndrome requires the presence of symptoms and an elevated apnea plus hypopnea index (AHI). The threshold of AHI beyond which OSA is diagnosed is usually between 10 and 20 obstructive events per hour of sleep recorded during nocturnal polysomnography (PSG).

Previous efforts to develop a clinical screening or case-finding instrument for the diagnosis of OSA have been based on patient questionnaires usually combined with anthropometric and physical findings.[7-13] These studies have used a wide variety of statistical approaches and techniques to predict the presence of OSA. These approaches have been limited by poor sensitivity or less-than-optimal specificity (even when sensitivity has been high), by conflicting results between the different studies, and by the lack of prospective validation in some of the models that hinders their generalizability. It is noteworthy that the overall subjective impression of experienced sleep physicians correctly identifies about 50% of the patients with OSA.[8,9] Portable home sleep monitoring has also been proposed for the diagnosis of OSA. The potential benefits of home sleep monitoring include reduced costs and the ability to evaluate

patients in their usual environment. Potential disadvantages include the problems inherent in unattended monitoring (especially if the patient sets up the equipment) and the time and expense involved if a technician is required to instrument the patient at home. In the latter situation, there may be minimal financial savings over in-laboratory PSG. In addition, many of the home monitoring systems that are available have not been adequately validated. PSG is therefore the diagnostic standard for OSA due to the lack of well-validated clinical or other screening tools. PSG may be expensive and inconvenient for the patient, and it is a labor-intensive, specialized procedure that may have limited accessibility in many jurisdictions. Increasing awareness of the adverse health effects of OSA and the availability of effective therapies have led to increased demand for PSG and, concomitantly, concerns by health insurance providers about increased expenditures for the diagnosis and treatment of OSA.

Artificial neural networks (ANNs) are computer programs modeled after the biological nervous system, and they are capable of recognizing complex patterns in data based on experience. Their application is useful in complex problems because they can analyze a large number of linear and nonlinear variables without the operator knowing or making assumptions about the relationships between the variables. Neural networks are "trained" by presenting a set of data together with the outcomes that the trainer wishes the network to learn. The trained neural network can then be evaluated by inputting similar, but previously unseen, data. This artificial intelligence approach for outcome prediction has been used successfully in other medical applications, including the prediction of acute myocardial infarction in patients presenting to an emergency room physician,[14] the diagnosis of pulmonary embolism,[15,16] and the predicted length of ICU stay.[17] ANNs have been shown to outperform physician impression or prediction[18,19] and to equal or exceed traditional statistical modeling in the prediction of outcomes.[20,21] This study was conducted to test the hypothesis that a trained generalized regression neural network (GRNN) could accurately classify patients with OSA from clinical data.

MATERIALS AND METHODS

The clinical data were collected from a retrospective review of randomly selected patients who presented to the London Health Sciences Center Sleep Clinic for assessment of possible OSA and who went on to have PSG. Greater than 95% of patients referred for possible OSA have PSG performed in their work-up. Reasons for not going on to PSG included a clinical diagnosis other than OSA (eg, insomnia or sleep-related laryngospasm) and patient refusal. Excluded from the study were patients [is less than] 16 years of age, patients with a nondiagnostic study (eg, total sleep time [is less than] 2 h), and patients referred for split-night studies. The latter group was comprised largely of patients with morbid obesity and obesity/ hypoventilation (hypoxemia and hypercapnia), and the clinical likelihood was so great and the symptoms so severe that the diagnostic and therapeutic studies were combined in order to implement therapy as quickly as possible.

Table 1—*Patient Demographics and Clinical Features

Variables	Total n = 405	Training Set n = 255	Test Set n = 150
Demographics			
Gender†			
Male	293	194	99
Female	112	61	51
Age, yr	47.2 ± 11.2 (20 – 79)	47.5 ± 11 (20 – 72)	46.9 ± 11.6 (23 – 79)
BMI, kg/m^2†	31.5 ± 6 (16.6 – 54)	30.9 ± 5.5 (16.6 – 49)	32.6 ± 6.3 (22 – 54)
NC, cm	42.1 ± 4 (30 – 53)	42.0 ± 3.9 (30 – 53)	42.2 ± 4.1 (32 – 53)
Clinical features			
Chronic snoring	403 (99.8)	255 (100)	149 (99.3)
Observed choking†	190 (47)	108 (42.4)	82 (54.7)
Witnessed apneas	319 (79)	199 (78)	120 (80)
Hypertension‡	124 (30.6)	52 (20.4)	72 (48)
Alcohol consumption			
None	149 (37)	85 (33.3)	64 (42.7)
Mild	190 (47)	126 (49.4)	64 (42.7)
Moderate to heavy	66 (16)	44 (17.3)	22 (14.6)
Smoking, pack-yr	15 ± 16.9 (0 – 90)	14.3 ± 16.6 (0 – 90)	16.2 ± 17.5 (0 – 90)
Epworth Sleepiness Scale	11 ± 4.9 (0 – 24)	10.8 ± 4.9 (0 – 24)	11.3 ± 4.8 (0 – 22)
AHI, events/h	29.2 ± 28 (0 – 146)	27.7 ± 25 (0 – 146)	32 ± 33 (0 – 146)
Clinical score†	6.5 ± 2.1 (0 – 11)	6.2 ± 1.2 (3 – 9)	7.1 ± 3 (3 – 11)
Physical examination			
Soft palate enlargement	320 (79)	208 (81.6)	112 (74.7)
Crowded oral pharynx	315 (78)	202 (79.2)	113 (75.3)

*Date are given as mean ± SD (range) or frequency (%).
†$p \leq 0.05$ for comparisons between the training and test set.
‡$p \leq 0.0005$ for comparisons between the training and test set.

Overnight PSG

PSG was performed on 1 night in all patients. The sleep study montage included EEG (C3/A2, C4/A1, O2/A1), electro-oculogram, submental electromyogram, left and right anterior tibialis electromyogram, ECG, thoracoabdominal motion, oronasal airflow (expired [CO.sub.2]), and arterial oxygen saturation with pulse oximetry using an ear probe sensor. The studies were seared manually, and the total AHI (number per hour total sleep time) was calculated for the night. Obstructive apneas were defined as the cessation of airflow for at least 10 s accompanied by ongoing respiratory effort. Obstructive hypopneas were defined as a reduction in airflow of at least 50% for at least 10 s accompanied by a reduction in respiratory effort and by an arousal or an arterial oxygen, desaturation of at least 3%. OSA was defined as an AHI [is greater than or equal to] 10/h for the purposes of this study.

Clinical Data

Forty-five clinical variables from nine categories were recorded in the database. These variables were chosen on the basis of previously published screening studies[7-13] and clinical experience. The categories included demographics (age, gender, marital status); nocturnal symptoms (frequent awakenings, choking, gasping); bed partner observations (snoring, witnessed apnea, observed choking, restlessness); daytime symptoms (unrefreshing sleep, morning headache, reported excessive daytime sleepiness, Epworth sleepiness scale, impaired nasal breathing); past medical history (nasal trauma, hypertension, airway surgery, allergies); medications (sedative/hypnotics, antidepressants, antihypertensives); social history (alcohol consumption, smoking history and pack-years); and anthropometrics (weight, height, body mass index [BMI] in kg/[m.sub.2], neck circumference [NC]). Alcohol consumption was categorized as none, mild (up to 1 drink per day), and moderate to heavy ([is greater than] 2 drinks per day). Data from the physical examination included systolic and diastolic BP, nasal obstruction, tonsillar enlargement, soft palate and/or uvular enlargement, crowding of the posterior oral pharynx, and the presence and grade of maxillary overjet. Data values that were not available in the chart review were obtained, whenever possible, by calling the patients by telephone or at a follow-up clinic visit. Less than 1% of the numeric data values were missing from the database. Missing numerical values (eg, NC) were recorded as the mean value from the group the subject was in (eg, women). Missing categorical values were recorded as negative.

GRNN software was used (NeuralShell Classifier Version 2; Ward Systems Group; Frederick, MD). Data preprocessing was conducted prior to neural network training to reduce the number of input variables to an optimal number of more discriminating factors. The raw data were first divided into two subsets according to AHI (ie, AHI [is less than] 10 and AHI [is greater than or equal to] 10). Differences between the subsets were then assessed using

Student's t test for each of the parametric variables and the Mann-Whitney U test for the nonparametric variables. Categorical variables were compared by [chi square] testing. All analyses were performed using appropriate software (Microsoft Excel Version 5.0 for the Macintosh; Microsoft; Redmond, WA and Systat Version 5.2 for the Macintosh; SPSS; Chicago, IL). A p value [is less than or equal to] 0.1 was arbitrarily chosen as the cutoff for the continued inclusion of a given variable in the final model.[22-25] The 45 clinical variables were reduced to 23 variables by data preprocessing. For the purpose of ANN training, the preprocessed database (n = 405) was randomly divided into a "training" set of 255 patients and a "test" set of 150 patients. During training of the network, the training set was repeatedly presented to the network until 99.6% of the patients were learned. The test set of previously unseen cases was then presented to the trained network, and predictions were compared with actual outcomes.

The aim of this study was to predict those subjects who had OSA (ie, AHI [is greater than or equal to] 10 from PSG). The accuracy, sensitivity, specificity, positive predictive values (PPVs), and negative predictive values (NPVs) were calculated. Accuracy was defined as the number of times the ANN correctly classified the patient as having or not having OSA.

Table 2—*Demographic and Clinical Features in Patients With (AHI ≥ 10) and Without (AHI < 10) OSA

Variables	OSA Negative n = 124	OSA Positive n = 281
Demographics		
Gender‡		
Male	71	222
Female	53	59
Age, yr	41.3 ± 11.8 (21 – 71)	48.5 ± 10.8 (20 – 79)
BMI, kg/m²†	28.3 ± 4.9 (16.6 – 44.9)	32.9 ± 5.9 (21.8 – 54.4)
NC, cm‡	39.7 ± 3.7 (30 – 48)	43.1 ± 3.7 (34 – 53)
Clinical features		
Chronic snoring	123 (99)	281 (100)
Observed choking‡	19 (15)	171 (61)
Witnessed apneas‡	74 (61)	245 (87)
Hypertension‡	17 (14)	107 (38)
Alcohol consumption†		
None	56 (45)	93 (33)
Mild	56 (45)	134 (48)
Moderate to heavy	12 (10)	54 (19)
Smoking, pack-yr†	11.2 ± 12 (0 – 50)	16.7 ± 18.3 (0 – 90)
Epworth Sleepiness Scale	10.4 ± 5 (0 – 24)	11.2 ± 5 (0 – 22)
AHI, events/h‡	4.6 ± 3 (0 – 9.9)	40.1 ± 27 (10 – 146)
Clinical score†	4.6 ± 1.6 (0 – 8)	7.3 ± 1.8 (3 – 11)
Physical examination		
Soft palate enlargement‡	67 (45)	253 (90)
Crowded oral pharynx‡	53 (43)	262 (93)

*Data are given as mean ± SD (range) or frequency (%).
†p = 0.05.
‡p = 0.0005.

RESULTS

The prevalence of OSA (defined as AHI [is greater than or equal to] 10) in this series was 69% (53% in women and 76% in men). The demographic and clinical features of the study population are outlined in Table 1. There were more men than women in the data set, and on average the patients were middle aged and overweight. The prevalence of OSA in the training set and the test set were 72% and 65%, respectively (p [is greater than] 0.05). There were small differences between the training set and the test set. The subjects in the test set had a larger BMI (p = 0.009) and a higher systolic BP (p = 0.002). There were proportionately more women in the test set (p = 0.03), and overall, more subjects were noted to have observed choking in the test set (p = 0.016).

Table 1--Patient Demographics and Clinical Features(*)

Variables	Total n = 405
Demographics	
Gender([dagger])	
Male	293

Female	112
Age, yr	47.2 [+ or -] 11.2 (20 - 79)
BMI, kg/[m.sup.2]	31.5 [+ or -] 6 (16.6 - 54)
NC, cm	42.1 [+ or -] 4 (30 - 53)
Clinical features	
Chronic snoring	403 (99.8)
Observed choking([dagger])	190 (47)
Witnessed apneas	319 (79)
Hypertension([double dagger])	124 (30.6)
Alcohol consumption	
None	149 (37)
Mild	190 (47)
Moderate to heavy	66 (16)
Smoking, pack-yr	15 [+ or -] 16.9 (0 - 90)
Epworth Sleepiness Scale	11 [+ or -] 4.9 (0 - 24)
AHI, events/h	29.2 [+ or -] 28 (0 - 146)
Clinical score([dagger])	6.5 [+ or -] 2.1 (0 - 11)
Physical examination	
Soft palate enlargement	320 (79)
Crowded oral pharynx	315 (78)

Variables	Training Set n = 255
Demographics	
Gender([dagger])	
Male	194
Female	61
Age, yr	47.5 [+ or -] 11 (20 - 72)
BMI, kg/[m.sup.2]	30.9 [+ or -] 5.5 (16.6 - 49)
NC, cm	42.0 [+ or -] 3.9 (30 - 53)
Clinical features	
Chronic snoring	255 (100)
Observed choking([dagger])	108 (42.4)
Witnessed apneas	199 (78)
Hypertension([double dagger])	52 (20.4)
Alcohol consumption	
None	85 (33.3)
Mild	126 (49.4)
Moderate to heavy	44 (17.3)
Smoking, pack-yr	14.3 [+ or -] 16.6 (0 - 90)
Epworth Sleepiness Scale	10.8 [+ or -] 4.9 (0 - 24)
AHI, events/h	27.7 [+ or -] 25 (0 - 146)
Clinical score([dagger])	6.2 [+ or -] 1.2 (3 - 9)
Physical examination	
Soft palate enlargement	208 (81.6)
Crowded oral pharynx	202 (79.2)

Variables	Test Set n = 150
Demographics	
Gender([dagger])	
Male	99
Female	51
Age, yr	46.9 [+ or -] 11.6 (23 - 79)
BMI, kg/[m.sup.2]	32.6 [+ or -] 6.3 (22 - 54)
NC, cm	42.2 [+ or -] 4.1 (32 - 53)
Clinical features	
Chronic snoring	149 (99.3)
Observed choking([dagger])	82 (54.7)
Witnessed apneas	120 (80)
Hypertension([double dagger])	72 (48)
Alcohol consumption	

None	64 (42.7)
Mild	64 (42.7)
Moderate to heavy	22 (14.6)
Smoking, pack-yr	16.2 [+ or -] 17.5 (0 - 90)
Epworth Sleepiness Scale	11.3 [+ or -] 4.8 (0 - 22)
AHI, events/h	32 [+ or -] 33 (0- 146)
Clinical score([dagger])	7.1 [+ or -] 3 (3 - 11)
Physical examination	
Soft palate enlargement	112 (74.7)
Crowded oral pharynx	113 (75.3)

(*) Date are given as mean [+ or -] SD (range) or frequency (%).

([dagger]) p [is less than or equal to] 0.05 for comparisons between the training and test set.

([double dagger]) p [is less than or equal to] 0.0005 for comparisons between the training and test set.

Patients With and Without OSA

The differences between patients with and without OSA status are summarized in Table 2. Men were more likely to have OSA than women, and the patients with OSA were older and heavier with larger NCs and BMIs than patients without OSA. A history of observed choking and witnessed apneas was associated with the presence of OSA. The patients with OSA were more likely to drink alcohol and were more likely to be smokers. The patients with OSA were more likely to have a history of hypertension, and on physical examination their systolic BP was higher (136 [+ or -] 15.9 vs 127 [+ or -] 15.8 mm Hg; p [is less than or equal to] 0.0001). The patients with OSA were also more likely to have a large soft palate and crowded posterior oral pharynx. Patients with OSA were more likely to report moderate to severe excessive daytime sleepiness (47.7%) than patients without OSA (27.4%; p = 0.001), although there was no difference in the Epworth Sleepiness Scale.

Table 2--Demographic and Clinical Features in Patients With (AHI [is greater than or equal to] 10) and Without (AHI < 10) OSA(*)

Variables	OSA Negative n = 124
Demographics	
Gender([double dagger])	
Male	71
Female	53
Age, yr	44.3 [+ or -] 11.8 (21 - 71)
BMI, kg/[m.sup.2] ([double dagger])	28.3 [+ or -] 4.9 (16.6 - 44.9)
NC, cm([double dagger])	39.7 [+ or -] 3.7 (30 - 48)
Clinical features	
Chronic snoring	123 (99)
Observed choking([double dagger])	19 (15)
Witnessed apneas([double dagger])	74 (61)
Hypertension([double dagger])	17 (14)
Alcohol consumption(dagger])	
None	56 (45)
Mild	56 (45)
Moderate to heavy	12 (10)
Smoking, pack-yr([dagger])	11.2 [+ or -] 12 (0 - 50)
Epworth Sleepiness Scale	10.4 [+ or -] 5 (0 - 24)
AHI, events/h([double dagger])	4.6 [+ or -] 3 (0 - 9.9)
Clinical score([double dagger])	4.6 [+ or -] 1.6 (0 - 8)
Physical examination	
Soft palate enlargement([double dagger])	67 (45)

Crowded oral pharynx([double dagger])	53 (43)

Variables	OSA Positive n = 281
Demographics	
Gender([double dagger])	
Male	222
Female	59
Age, yr	48.5 [+ or -] 10.8 (20 - 79)
BMI, kg/[m.sup.2] ([double dagger])	32.9 [+ or -] 5.9 (21.8 - 54.4)
NC, cm([double dagger])	43.1 [+ or -] 3.7 (34 - 53)
Clinical features	
Chronic snoring	281 (100)
Observed choking([double dagger])	171 (61)
Witnessed apneas([double dagger])	245 (87)
Hypertension([double dagger])	107 (38)
Alcohol consumption(dagger])	
None	93 (33)
Mild	134 (48)
Moderate to heavy	54 (19)
Smoking, pack-yr([dagger])	16.7 [+ or -] 1.8.5 (0 - 90)
Epworth Sleepiness Scale	11.2 [+ or -] 5 (0 - 22)
AHI, events/h([double dagger])	40.1 [+ or -] 27 (10 - 146)
Clinical score([double dagger])	7.3 [+ or -] 1.8 (3 - 11.)
Physical examination	
Soft palate enlargement([double dagger])	253 (90)
Crowded oral pharynx([double dagger])	262 (93)

(*) Date are given as mean [+ or -] SD (range) or frequency (%).

([dagger]) $p < 0.05$.

([double dagger]) $p < 0.0005$.

Men and Women With OSA

The differences between women and men with OSA are summarized in Table 3. The women were heavier (a larger BMI), but the men had larger NC. The men with OSA were more likely to be moderate to heavy drinkers than the women with OSA. Witnessed apneas were more common in the men with OSA than in the women with OSA.

Table 3--Patient Demographics and Clinical Features of Men and Women With OSA (AHI > 10)(*)

Variables	Men n = 222
Demographics	
Age, yr	48.3 [+ or -] 10.5 (24 - 74)
Body mass index, kg/[m.sup.3] ([double dagger])	32 [+ or -] 5.2 (21.8 - 54)
NC, cm([dagger])	44 [+ or -] 3.2 (37 - 53)
Clinical features	
Chronic snoring	222 (100)
Observed choking	133 (60)

Witnessed apneas([dagger])	199 (89.6)
Hypertension	85 (38.3)
Alcohol consumption([double dagger])	
None	57 (25.7)
Mild	113 (51)
Moderate to heavy	52 (23.4)
Smoking, pack-yr	17.7 [+ or -] 18.8 (0 - 90)
AHI, events/h	40.6 [+ or -] 26.9 (10- 146)
Clinical score([double dagger])	7.1 [+ or -] 1.7 (3 - 11)
Physical examination	
Soft palate enlargement	201 (90.5)
Crowded oral pharynx	208 (93.7)

Variables	Women n = 59
Demographics	
Age, yr	49.4 [+ or -] 11.8 (20 -79)
Body mass index, kg/[m.sup.3] ([double dagger])	36.5 [+ or -] 6.5 (22.9 - 51)
NC, cm([dagger])	39.7 [+ or -] 3.3 (34 - 49)
Clinical features	
Chronic snoring	59 (100)
Observed choking	38 (64.4)
Witnessed apneas([dagger])	46 (78)
Hypertension	22 (37.3)
Alcohol consumption([double dagger])	
None	36 (61)
Mild	21 (35.6)
Moderate to heavy	2 (3.4)
Smoking, pack-yr	13.1 [+ or -] 16.7 (0 - 55)
AHI, events/h	38.4 [+ or -] 29.1 (11 - 127)
Clinical score([double dagger])	8.1 [+ or -] 1.9 (4 - 11)
Physical examination	
Soft palate enlargement	52 (88)
Crowded oral pharynx	54 (91.5)

(*) Date are given as mean [+ or -] SD (range) or frequency (%).

([dagger]) p [is less than or equal to] 0.05.

([double dagger]) $p < 0.0005$.

GRNN Prediction

The 23 variables used in the neural network are listed in Table 4. A total of 255 training cases were used with a test set of 150 patients. Overall, the trained network had a mean predictive accuracy of 91.3% (95% confidence interval [CI], 86.8 to 95.8). The sensitivity of neural network prediction for the presence of OSA was 98.9% (95% CI, 96.7 to 100), and specificity was 80% (95% CI, 70 to 90). The PPV that the patient would have OSA if the network said so was 88.1% (95% CI, 81.8 to 94.4), whereas the NPV that the patient would not have OSA (if so classified) was 98% (95% CI, 94 to 100). The ANN misclassified one subject in the test set as not having OSA who had an AHI [is greater than] 10; this subject had mild OSA with an AHI of 10.5. The ANN classified 12 subjects without OSA as having OSA. The area under the receiver operating characteristic curve was 0.94.

Table 4--Patient Demographics and Clinical Features Used in the Neural Network of Prediction of OSA

Variables	Characteristics
Demographics	Age, gender
Nighttime symptoms	Frequent awakening, experienced choking
Bed partner observations	Witnessed apneas, observed choking
Daytime symptoms	Reported excessive daytime sleepiness, Epworth sleepiness scale
Past medical history	Hypertension
Social history	Alcohol consumption, smoking in pack-yr
Anthropometrics	Height, weight, BMI, systolic BP [is greater than or equal to] 140, diastolic BP [is greater than or equal to] 90
Physical examination	Tonsillar enlargement, soft palate enlargement, crowding of the oral pharynx
Clinical score	Sum of the clinical scores for the binary categorical values (maximum score = 12)

DISCUSSION

We have shown in this retrospective study of a population of OSA patients that a trained GRNN can accurately diagnose the presence of OSA (defined as AHI [is greater than or equal to] 10/h). To our knowledge, this study is the first use of a neural network to predict the presence or absence of OSA. The prevalence of OSA in this population was 69%, with OSA more common in men than in women. In the overall group, the presence of OSA was associated with a history of witnessed apneas, observed choking, and increased smoking and alcohol intake. The patients with OSA were more obese, had a higher BP, and were more likely to have a large soft palate and a crowded posterior oral pharynx. Different from some previously published studies[7,8,11] yet compatible with others,[9] reported excessive daytime sleepiness was associated with the presence of OSA.

OSA is an ideal condition for this use of screening or ease-finding techniques. OSA is a chronic condition that is highly prevalent in the general population, and many effective therapies are available. If this neural network retains the high level of accuracy with prospective validation, it could be used as a simple clinical screening tool to reduce the number of PSG studies performed in patients without OSA. The novel approach of applying a neural network to the clinically based prediction of OSA allows the incorporation of a much larger number of variables than what are conventionally used in linear or logistic regression techniques. Perhaps more important than the variables themselves is the technique of processing that seeks to identify subtle patterns and relationships. The user is shielded from the multitude of calculations, the technique is not cumbersome, and the prediction is free from bias, fatigue, and personal opinion. In addition, a well-trained neural network can easily handle differences between groups (eg, between men and women) within the same model.

Screening tests are generally designed to favor high sensitivities so that eases are not missed. The development of clinical prediction rules emphasizing optimal sensitivity can be developed by accepting a concomitant decrease in specificity. The financial impact of varying the cutoff points is related to the costs (both financial and human) associated with false positive vs false negative results. Most previous clinical screening studies have attempted to rule in the presence of OSA with a high level of sensitivity in order not to miss significant OSA. When Hoffstein and Szalai[9] studied 594 patients referred for possible OSA, they found that several variables from the history and physical examination were predictors of the presence of OSA (age, gender, BMI, witnessed apneas, pharyngeal examination). However, physician subjective impression only correctly identified 51% of the patients with OSA and 71% of the patients without OSA (sensitivity, 60%; specificity, 65%). They concluded that clinical impression alone could not reliably identify patients with or without OSA.

Crocker and colleagues[7] used a logistic regression model developed with clinical variables to predict the presence of OSA. They found that age, witnessed apneas, BMI, and hypertension were independently associated with the presence

of OSA (AHI [is greater than] 15/h) in a group of 100 patients. The model had a sensitivity of 99,% and a specificity of 51% for the diagnosis of OSA. The prevalence of OSA was only 27% in the initial group and 34% in the prospective test group of 105 patients. In the test group, 33 of 36 patients were correctly classified with OSA. The three patients in the test group who the model misclassified as not having OSA all had an elevated AHI [is greater than] 40/h. The researchers estimated that they could reduce the number of PSG studies performed by approximately one third, but would have misclassified patients with significant OSA as not having OSA. Viner and coworkers[8] developed a logistic regression model in 410 patients; age, gender, BMI, and snoring were retained as significant variables in the model. The prevalence of OSA was 46% in this population. The sensitivity was 94% and the specificity was 9,8% for the diagnosis of OSA (defined as AHI [is greater than] 10). The area under the receiver operating characteristic curve was 0.77. The physician's subjective impression of the presence of OSA had a sensitivity of 59,% and a specificity of 70% with an accuracy rate of 63%. The application of this model would have allowed them to potentially reduce the number of PSG studies by one third. This model was not prospectively validated. The merits of these two different models are difficult to compare, as the prevalence of OSA in the populations studied were different and relatively low, and this affects the PPV and NPV of the test. The neural network, however, had better sensitivities for the diagnosis of OSA than either of these models, it had better specificities, and it did not miss eases of significant OSA while potentially reducing the number of PSG studies required.

Flemons and colleagues[11] evaluated 180 patients with possible OSA, and they developed a model that included NC, hypertension, habitual snoring, and witnessed gasping or choking. The prevalence of OSA was 45% in the group (AHI [is greater than] 10/h). Likelihood ratios were determined, and the calculated clinical score from the model was used to provide a posttest probability of OSA. This model was superior to physician impression, and it was comparable or superior to previously published models.[7,8] Pradhan et al[13] conducted a prospective study of OSA prediction in 150 patients. They adjusted the cutoff values in the clinical model (containing age, gender, loud snoring, and BMI) to have a sensitivity of 100% for the diagnosis of OSA. The prevalence of OSA was 57% (AHI [is greater than] 10/ h). They also developed a model that included overnight home oximetry. They found that it was more cost effective to pursue a clinical screening strategy than a clinical strategy with oximetry. The use of oximetry would have decreased the number of PSG by 13%, compared to an 8% reduction for clinical screening alone, but this savings was eliminated by the additional cost of home monitoring. Series and colleagues[26] used a qualitative analysis of home oximetry for screening in 9,40 patients referred for assessment of OSA. They reported a sensitivity of 98% and a specificity of 48% (PPV, 61%; NPV, 97%) in a population with a prevalence of 46%. The two patients who were misclassified as not having OSA had very low AHI (14/h and 16/h, respectively). Oximetry had a high sensitivity for the diagnosis of OSA in that study and could have reduced the number of PSG studies performed by approximately 9,5%. However, it is more time consuming and expensive to organize and interpret a home study than it is to use a simple clinical prediction model such as a neural network in the office setting.

This study differs from other screening studies because of the use of the neural network. If the neural network could accurately rule in or rule out OSA, then the PSG could be eliminated from the diagnostic assessment of some non-OSA patients, thereby saving valuable resources, potentially from some OSA subjects who might proceed to a therapeutic study (continuous positive airway pressure trial) instead of a diagnostic study. Given that OSA has significant consequences, physicians would not want to use a screening toot that missed patients with significant sleep-disordered breathing. In this study, the sensitivity for the diagnosis of OSA was 98.9% (95% CI, 96.7 to 100). The most important measure of the neural network as a screening instrument is the high sensitivity coupled with the low false-negative rate. The neural network misclassified only 1 out of 150 eases as not having OSA when the AHI was [is greater than or equal to] 10/h. This subject had an AHI of 10.5 events per hour, a very low level of OSA. This suggests that the neural network does not miss serious eases of OSA when it does make a mistake. Only 19, eases without OSA were classified as having OSA and therefore underwent potentially unnecessary PSG testing. Overall, 48 patients (39.%) would not have required PSG based on the neural network prediction. Although a specific cost analysis was not performed in this study, it is apparent that the number of PSG studies could be reduced if the network could accurately rule in and rule out OSA. These patients could have been reassured that they did not have significant OSA; potentially, they could have preceded directly to treatment based on their risk factors and symptoms. Ultimately, this approach may increase the number of patients evaluated for the possibility of OSA, while reserving PSG for patients who are more likely to have significant sleep-disordered breathing. In centers with long waiting lists,

this could also reduce waiting times for evaluation.

The limitations to this study include the retrospective nature of the data review, the lack of prospective validation, and the relatively small numbers. The model performed exceedingly well, even with the small numbers; however, in general, the performance of neural networks improves with increases in the size of the data set. Future work with this prediction technique will include expansion of the training set and validation in a larger set of consecutively and prospectively collected cases. This model was developed in a population with a fairly high prevalence of OSA (referral to a sleep clinic; prevalence, 69%), and it needs to be tested in a population with a lower prevalence. This will help determine if the neural network model that was developed in this study is generalizable to other patient populations. Potentially, a neural network could be developed for children with OSA, as PSG is more difficult to perform in this specific population.

CONCLUSION

This study has shown that an optimally trained GRNN can accurately rule out OSA in a referral setting where the prevalence of OSA is high (ie, 69%). Also, the neural network did not miss important cases of OSA. This neural network was developed from readily available clinical data found in most patient charts. Approximately 32% of all PSG studies could have been avoided by using this classification tool. Training with increased numbers of subjects can be anticipated to only improve network predictive performance. Prospective validation of this new prediction methodology is warranted. The limited capital costs and expenses associated with operating such an instrument suggest that significant savings may be realized by using this novel approach.

REFERENCES

[1] Young T, Palta M, Dempsey J, et al. The occurrence of sleep-disordered breathing among middle-aged adults. New Eugl J Med 1993; 328:1230-1235

[2] Ferguson KA, Fleetham JA. Sleep-related breathing disorders: 4. Consequences of sleep disordered breathing. Thorax 1995; 50:998-1004

[3] Flemons WW, Tsai W. Quality of life consequences of sleep-disordered breathing. J Allergy Clin Immunol 1997; 99:S750-S756

[4] Ulfberg J, Carter N, Talback M, et al. Excessive daytime sleepiness at work and subjective work performance in the general population and among heavy snorers and patients with obstructive sleep apnea. Chest 1996; 110:659-663

[5] George CF, Nickerson PW, Hanly PJ, et al. Sleep apnoea patients have more automobile accidents [letter]. Lancet 1987; 2:447

[6] He J, Kryger MH, Zorick FJ, et al. Mortality and apnea index in obstructive sleep apnea. Chest 1988; 94:9-14

[7] Crocker B, Olson LG, Saunders NA, et al. Estimation of the probability of disturbed breathing during sleep before a sleep study. Am Rev Respir Dis 1990; 142:14-18

[8] Viner S, Szalai JP, Hoffstein V. Are history and physical examination a good screening test for sleep apnea? Ann Intern Med 1991; 115:356-359

[9] Hoffstein V, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. Sleep 1993; 16:118-122

[10] Dealberto MJ, Ferber C, Garma L, et al. Factors related to sleep apnea syndrome in sleep clinic patients. Chest 1994; 105:1753-1758

[11] Flemons WW, Whitelaw WA, Brant R, et al. Likelihood ratios for a sleep apnea clinical prediction rule. Am J Respir Crit Care Med 1994; 150:1279-1285

[12] Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. Sleep 1995; 18:158-166

[13] Pradhan PS, Gliklich RE, Winkelman J. Screening for obstructive sleep apnea in patients presenting for snoring surgery. Laryngoscope 1996; 106:1393-1397

[14] Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991; 115:843-848

[15] Scott JA, Palmer EL. Neural network analysis of ventilation-perfusion lung scans. Radiology 1993; 186:661-664

[16] Patti S, Henry JW, Rubenfire M, et al. Neural network in the clinical diagnosis of acute pulmonary embolism. Chest 1993; 104:168.5-1689

[17] Danter WR, Wood T, Morrison NJ, et al. Artificial neural network prediction of survival or death following admission to ICU. Clin Invest Med 1996; 19:S14

[18] Tourassi GD, Floyd CE, Sostman HD, et al. Acute pulmonary embolism: artificial neural network approach for diagnosis. Radiology 1993; 189:555-558

[19] Bottaci L, Drew PJ, Hartley JE, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. Lancet 1997; 350:469-472

[20] Pilon S, Tandberg D. Neural network and linear regression models in residency selection. Am J Emerg Med 1997; 15:361-364

[21] Kennedy RL, Harrison RF, Burton AM, et al. An artificial neural network system for diagnosis of acute myocardial infarction (AMI) in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. Comput Methods Programs Biomed 1997; 52:93-103

[22] Lawrence J. Introduction to neural networks: design, theory and applications. Nevada: California Scientific Software 1994: 207-210

[23] Crooks T. Care and feeding of neural networks. A1 Expert 1992; 7:36-41.

[24] Stein R. Selecting data for neural networks. AI Expert 1993; 8:42-47

[25] Lawrence J. Data preparation for a neural network. AI Expert 1991; 6:34-41.

[26] Series F, Marc I, Cormier Y, et al. Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome. Ann Intern Med 1993; 119:449-453

(*) From the Departments of Otolaryngology (Drs. Kirby and George) and Medicine (Drs. Danter, Ferguson, and Francovic), The University of Western Ontario, London, Ontario, Canada. Presented in part at the Meeting of the Eastern Section of the American Laryngological, Rhinological and Otological Society Inc., New York, NY, January, 1998.

Manuscript received May 7, 1998; revision accepted March 2, 1999.

Correspondence to: Kathleen A. Ferguson BSc, MD, FCCP, University Campus, London Health Sciences Centre, 339 Windermere Rd, London, Ontario, Canada N6A 5A5; e-mail: kafergus@ julian.uwo.ca3