Sleep, 20(3):232–236

Using Self-Reported Questionnaire Data to Prioritize OSA Patients for Polysomnography

Zoe Pouliot, Mel Peters, Herb Neufeld and Meir H. Kryger

Sleep Disorders Center, St. Boniface General Hospital, Section of Respiratory Diseases, University of Manitoba, Winnipeg, Manitoba, Canada

Summary: Many laboratories have large numbers of patients with suspected obstructive sleep apnea (OSA) waiting to be tested. We assessed the use of simple clinical data to detect those patients with an apnea index <20 (low AI) who could be studied less emergently.

Using questionnaires completed by patients prior to evaluation, we collected data on 354 consecutive patients (281 males, 73 females; mean age 48.6 years) referred for OSA and assessed with polysomnography (PSG). The questionnaires included the Epworth sleepiness scale (ESS), height, weight, age, and a history of observed apnea.

Analysis of receiver operating characteristics curves revealed that both body mass index (BMI) [area under curve = 0.7258, standard error (SE) = 0.03, $p < 0.01$] and ESS (area under curve = 0.5581, SE = 0.03, $p = 0.03$) were significantly better than chance alone in detecting people with AI < 20. ESS ≤12 was found in 37.9% of the subjects but 39.6% of those expected to have a low AI using ESS had an AI ≥20. A BMI ≤28 was found in 24.9% of the subjects; 14.8% of those expected to have a low AI using BMI had an AI ≥20. Combining these variables improved accuracy but resulted in smaller groups; a cut-off of ESS ≤12 and BMI ≤28 resulted in a group of 33 (9.3% of subjects), only two (6%) of whom were falsely called low AI. Adding to this the fact that apnea had not been observed resulted in a group of nine patients (2.5% of subjects), none of whom had an AI ≥20. Thus there is a tradeoff; the more variables used, the greater the accuracy but the smaller the percent of cases selected to have low AI. However, in laboratories with hundreds of patients waiting to be tested, any procedure better than chance to help prioritize patients seems worthwhile. **Key Words**: Obstructive sleep apnea—Clinical predictors—Epworth sleepiness scale.

Our laboratory and many laboratories around the world have large numbers of patients with suspected obstructive sleep apnea (OSA) waiting to be tested. Since some patients may be severely affected, while others may have a mild disorder, a simple system for prioritizing patients would be helpful. The ability to predict which patients are severely affected would allow clinicians to identify and evaluate high-risk patients on waiting lists and enable them to diagnose and treat these patients much sooner. Several groups (1–8) have examined different clinical features in an attempt to find predictive markers for OSA severity. Such data collected includes age, weight, gender, history of snoring, snorting, and witnessed apnea as well as the presence of excessive daytime sleepiness. A highly reliable set of clinical features has not yet been formulated that enables clinicians to predict the absence or presence or severity of OSA. None of the abovementioned reports focused primarily on the detection of people who did not have clinically significant apnea and who would, therefore, require less-emergent assessment.

From questionnaires completed by patients we examined the utility of using three markers, the Epworth sleepiness scale (ESS), body mass index (BMI), and whether apnea had been observed by a bedpartner. We hypothesized that if these three markers alone or in combination were "negative" then there was little likelihood that the patient had clinically significant apnea; therefore such a patient could safely be placed at the bottom of the waiting list.

METHODS

All patients seen in our sleep laboratory were referred by a primary physician. When referrals arrived in our sleep laboratory, patients were sent a questionnaire that was returned to our laboratory by mail. From the questionnaire responses we analyzed the following variables: ESS (9), height, and weight, and whether apnea had been observed in the patient. BMI was cal-

Accepted for publication January, 1997.

Address and reprint request to Dr. M. Kryger, Sleep Disorders Center, St. Boniface General Hospital, Room R2034, 351 Tache Avenue, Winnipeg, Manitoba, R2H 2A6, Canada.

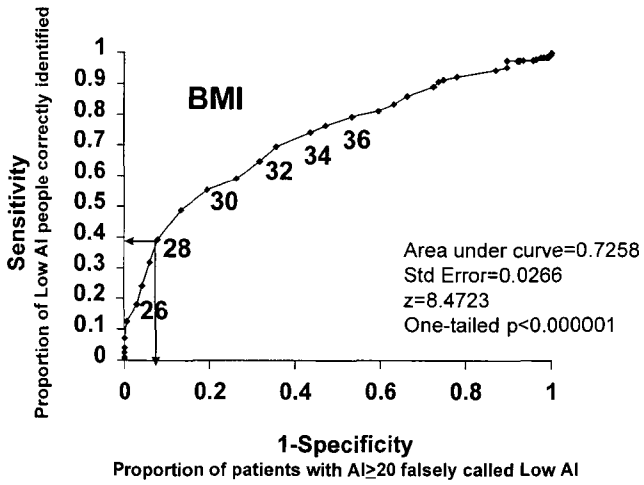

FIG. 1. Receiver operator characteristic (ROC) curve of body mass index (BMI) in detecting patients with a low apnea index (AI) (AI < 20). The numbers next to the curve are BMI values. Notice a BMI cut-off of 28 detected 39.3% of people with a low AI (horizontal arrow) while 8% of people with AI >20 were falsely labeled as "low AI" (vertical arrow). The area under the curve can vary from 0.5 (pure chance) to 1.0 (a perfect test). The z value is the area under the curve minus the expected area (about 0.5) divided by its standard error; this is an index of normal probability and values > 1.65 indicate that the probability of the result is greater than chance ($p < 0.05$).

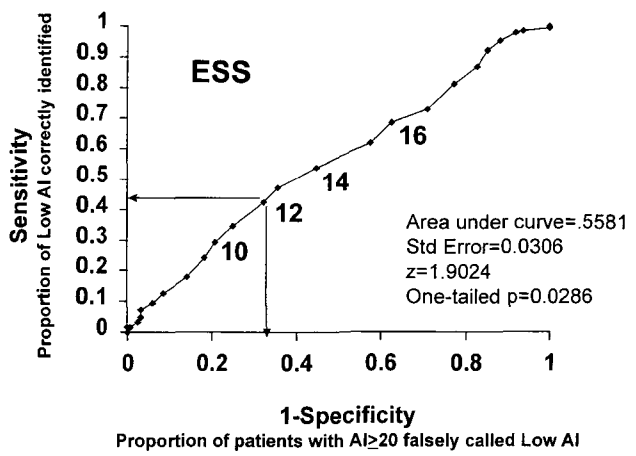

FIG. 2. ROC curve of the Epworth sleepiness scale (ESS) in detecting patients with low AI (AI < 20). The numbers next to the curve are ESS values. Notice that an ESS cut-off of 12 detected 42.4% of people with a low AI (horizontal arrow), while 32.5% of people with AI > 20 were falsely labeled as "low AI" (vertical arrow). See legend of Fig. 1 for an explanation of the statistics presented.

culated from height and weight. This questionnaire also included questions relating to other sleep disorders, medical conditions, and drug intake.

We prospectively analyzed the ability of the three abovementioned variables (ESS, BMI, and whether apnea was observed) in 354 consecutive patients (281 males, 73 females; mean age 48.6 years, range 19–89; 256 of the 354 patients snored) who were referred for suspected OSA during a year.

We defined "normal" values as follows: ESS ≤ 12, BMI ≤ 28, and no observed apneas. We hypothesized that using these cut-offs would screen for people with an apnea index < 20 (low AI). A cut-off value of 12 was chosen for ESS because the range reported in people with primary snoring (i.e. without clinical apnea) was 0–11 (see Table 2 in Ref. 9). A cut-off value for BMI of 28 was used in this study; the U.S. weight guidelines suggest that the range of desirable BMI for men and women was 21–27 kg/m^2 (10). Not all investigators agree with these guidelines, however, and believe that an upper limit of 25 may be more appropriate especially in women (11).

Detailed polysomnography (PSG) was performed according to accepted standards (12) to establish a precise diagnosis and AI in the patients. This consisted of continuous monitoring of the electroencephalogram [EEG (C3/A2, 02/A1)], electrooculogram [EOG (ROC/LOC)], electromyogram [EMG (submental and anterior tibialis)], electrocardiogram (EKG), respiratory effort using inductance plethysmography (SARA, Ambulatory Monitoring, Redwood, CA), airflow using end-tidal CO_2 (Normocap 200, Datex Medical Instrumentation, Inc., Tewksbury, MA), and arterial-oxygen saturation [using a pulse oximeter with an ear probe (Biox 3740, Ohmeda, Louisville, CO)]. All variables were recorded simultaneously by microcomputer. Sleep breathing abnormalities are expressed as AI (number of events per hour).

Statistical methods

We examined whether the three variables (ESS, BMI, and observed apnea) alone or in combination could detect patients subsequently found to have an AI < 20. This value was chosen because it has been shown that AI ≥20 is associated with increased mortality (13). For each of the variables we used an unpaired t test to see whether AI was different in those with a normal value for the variable and those with an abnormal value. We also used standard techniques (Epistat version 5.3, Richardson, TX) to obtain receiver operator characteristics (ROC) curves.

RESULTS

The ROC curve of BMI and ESS are shown in Figs. 1 and 2. Analysis of the area under the curves revealed that both variables were highly statistically better than chance in detecting people with low AI. The area under the ROC curve for BMI revealed it was higher than the area under the ESS ROC curve ($p < 0.0001$). Notice that neither ESS nor BMI ROC curves showed a discrete, specific cut-off point.

TABLE 1. *AI in groups screening "negative" and "positive"*

Screening question	n	Apnea index[a]
ESS ≤ 12	134	22.3
ESS > 12	220	33.0
BMI ≤ 28	88	10.2
BMI > 28	266	35.1
ESS ≤ 12 and BMI ≤ 28	33	7.5
ESS > 12 and BMI > 28	321	31.3
Observed apnea = no	139	21.3
Observed apnea = yes	215	33.8

AI, apnea index; ESS, Epworth sleepiness scale; BMI, body mass index.
[a] $p < 0.005$ for all comparisons.

The screening questions expected to detect low AI (AI < 20) resulted in groups with a lower AI (see Table 1). The "best" single variable was the BMI ≤28. Patients with a BMI ≤28 had a mean AI of 10.2 vs. 35.1 for those with a BMI >28. Adding the ESS did not significantly change the results.

Table 2 shows the ability of the variables to detect patients with low AI (AI < 20). ESS ≤12 was found in 37.9% of the subjects (134 of 354); BMI ≤28 was found in 24.9% (88 of 354); BMI ≤28 and ESS ≤12 were found in 9.3% (33 of 354); BMI ≤28 and ESS ≤12 without observed apnea were found in 2.5% (9 of 354).

For a cut-off of ESS ≤12 we found that 39.6% of those expected to have a low AI had AI ≥20; for a cut-off of BMI ≤28, 14.8% of those expected to have a low AI had AI ≥20. Combining these variables improved this rate; a cut-off of ESS ≤12 and BMI ≤28 resulted in a group of 33, only two (6%) of whom were falsely called low AI. Adding to this cut-off the fact that apnea had not been observed resulted in a group of nine patients, none of whom had AI ≥20.

DISCUSSION

The purpose of this study was not to detect people with high AI or to try to correlate clinical features with apnea severity. The purpose was to use self-administered questionnaire data to reliably detect people with low AI who may not require urgent evaluation. Such information may be helpful for purposes of prioritization of PSG evaluation. Our results indicate that attempting to prioritize patients on the basis of ESS, BMI, and whether apnea is observed may be helpful. The single variable resulting in the lowest error rate was BMI ≤28. Combining variables decreased the error rate but resulted in fewer patients being detected.

In managing a waiting list one can simply test patients in order of referral. This would mean that low-AI patients would be evaluated prior to some patients with severe OSA. Another approach is to try to predict which patients have a low AI and put them lower on the waiting list. However, there is a tradeoff; the more variables one uses, the greater the accuracy but the smaller the percent of total referrals selected to have low AI. However, in laboratories with hundreds of patients waiting to be studied, any procedure better than chance to help in prioritizing patients seems worthwhile.

Our report varies from other studies (Table 3) in several ways. First, our study is the first to examine the utility of using ESS in attempting to detect low-AI patients. This is the most widely used measure of subjective sleepiness (9). Second, almost all previous studies examining such variables focused on predicting the presence of apnea or its severity, not on finding low-AI patients. Three of these reports (2,3,5) detailed how such analyses might be used to find such low-AI patients and how this might reduce the number of PSGs performed.

Previous studies have focused on methods of predicting the presence and severity of OSA and not on finding patients without OSA. The earlier studies (2,6) examining this issue involved using data from relatively small numbers of patients. Kapuniai et al. (1) analyzed data from 76 sleep-disorders patients and developed a simple apnea score based on whether apnea was observed and whether snoring was present. This

TABLE 2. *Ability to select patients with AI <20*

Variable and cut-off	n[a]	No. falsely called low AI	Sens[b]	Spec[c]	% Falsely called low AI[d]	% Apneics called low AI[e]
ESS ≤ 12	134	53	0.424	0.675	39.6	32.5
BMI ≤ 28	88	13	0.393	0.920	14.8	8.0
ESS ≤ 12, BMI ≤ 28	33	2	0.162	0.988	6.0	1.2
ESS ≤ 12, BMI ≤ 28, OA−[f]	9	0	0.047	1.00	0	0

[a] n = number of people fulfilling the screening criteria.
[b] Sens = sensitivity, the probability of a positive test for low AI in patients known to have low AI.
[c] Spec = specificity, the probability of a negative test for low AI in patients with AI ≥20.
[d] This is the percent of those who were expected to have low AI found to have AI ≥20.
[e] This is the percent of patients found to have AI ≥20 who had been screened and expected to have low AI.
[f] OA− indicates that there was no observed (witnessed) apnea.

TABLE 3. *Variables used in predicting presence or absence of apnea*

First author and date of study	BMI	Age	Gender	Observed apnea	Snoring	EDS	Other
Kapuniai, 1988 (1)				X	X		
Crocker, 1990 (2)	X	X		X			hypertension
Viner, 1991 (3)	X	X	X		X		
Hoffstein, 1993 (4)	X	X	X	X			pharyngeal examination
Rauscher, 1993 (5)				X		X[a]	weight, height
Kump, 1994 (6)				X[b]	X	X[c]	
Douglass, 1994 (7)	X	X	X	X	X	X	sleep-disorders questions[d]
Maislin, 1995 (8)	X	X	X	X	X		
Our study, 1996	X			X		X[e]	

EDS, excessive daytime sleepiness.
[a] Fallen asleep reading.
[b] Roommate observed choking.
[c] Fallen asleep driving.
[d] Also includes questions concerning narcolepsy, periodic-limb-movement disorders, and psychiatric sleep disorders.
[e] ESS.

score system was apparently 100% effective in screening when AI was >40 and 70–76% effective in screening when AI was >5. Their data suggested that using an apnea score < 2 would result in AI being ≤10 80–88% of the time.

Crocker et al. (2), using data from 100 patients and logistic regression analysis, developed a model to predict an apnea-plus-hypopnea index (AHI) >15. Observed apnea was by far the most predictive variable; the other three variables in the model were hypertension, BMI, and age. This model correctly classified 33 of 36 patients with AHI >15 (sensitivity = 92%) but only 35 of 69 patients with AHI ≤15 (specificity = 51%). Their study suggested that using such an approach might reduce the need for sleep studies by about one-third and yet identify most patients with OSA.

Scharf et al. (14) reported on the use of a self-administered questionnaire on a group of 40 patients with hypertension. The symptom questionnaire was found not to be useful in predicting OSA.

Other studies involved larger numbers of patients. Using data from 410 patients, Viner et al. (3) developed a prediction model using stepwise logistical regression with the independent variables being gender, age, BMI, and a history of snoring. The data were obtained by physician interview. The probability that their model could correctly identify the patient with apnea was 0.77. Using this model and considering a predicted probability of having apnea of 70% correctly identified only 28% of patients with sleep apnea and correctly excluded 95% of patients without sleep apnea. These authors suggested that this might result in a one-third reduction of PSGs. Hoffstein and Szalai (4) extended their observations to a group of 594 patients and used stepwise multiple linear regression to predict AHI. This model used BMI, age, gender, observed apnea, and pharyngeal examination. This model explained only 36% of the variance in AHI.

Douglass et al. (7) used the sleep disorders questionnaire (SDQ) in 158 patients with sleep apnea. This questionnaire has 175 items and the authors recommend that it be used as a confirmatory diagnostic tool after clinical interview in general practice. The instrument yields a score for sleep apnea, periodic-limb-movements disorders, psychiatric sleep disorders, and narcolepsy. The apnea-related questions enquire about snoring, observed apnea, history of hypertension, weight, height, effect of position, alcohol intake, and sleepiness. For males (n = 141) the calculated sensitivity was 0.85, specificity 0.76, positive predictive value 0.72, and negative predictive value 0.87. The authors have used this scale to decide whether to add a multiple sleep latency test (MSLT) to PSG based on the sleep apnea and narcolepsy scores.

Kump et al. (6) reported on a group of 465 patients made up of 38 people previously diagnosed as having apnea; the remainder were relatives or neighbors who were enrolled as part of an epidemiological study and not because of symptoms or physician referral. This group reported a final prediction model using logistical regression analysis based on intensity of snoring, roommate-observed choking, and having fallen asleep while driving. The addition of data on gender and BMI further improved the ability to predict apnea.

Rauscher et al. (5) reported on 184 snorers suspected of having OSA and referred to a sleep laboratory. Most of these patients (63%) were self-referred with the remainder referred by ear, nose, and throat, general practitioner, or other specialists. This group used a self-administered questionnaire. A stepwise logistic regression analysis was used. Only four variables contributed to the model: weight, height, whether the subject fell asleep while reading, and whether apnea was

observed. Specificity was defined by the authors as the number of true negative predictions divided by the number of patients with an AHI < 10 or < 20. This model resulted in a specificity of 0.45 in predicting an AI >10 and 0.41 in predicting an AI >20. Thus this tool was not helpful in screening for low AHI. The authors suggest that addition of oximetry improved the ability to predict low AI and could result in a one-third reduction of negative PSG in snorers.

Maislin et al. (8) used multiple logistical analysis to calculate multivariate apnea risk index (MAP) using a sample of 427 patients. The final model had BMI, age, gender, and an index (Index 1) calculated by averaging values for snoring, observed apnea, snorting, and gasping. The area under the ROC curve for MAP was 0.786, for BMI 0.73, and for Index 1 was 0.69. The difference between the areas under the curve for BMI and MAP was small but significant. The area under the ROC curve for BMI in this report (0.73) was identical to what we found.

In most laboratories ~80% of the patients have obesity as the cause of their obstructive sleep apnea. Therefore, weight itself would be expected to be an unreliable predictor of whether an individual patient has obstructive sleep apnea. Surprisingly, our study using cut-offs confirmed the impression from several studies (Table 3) using stepwise regression; almost all the models resulted in BMI being an important independent variable. Although the ESS as a predictor results in groups with statistically different apnea indices, the error rate using this variable alone was unacceptably high. This also is not surprising since in sleep apnea there is known to be a poor correlation between objective and subjective sleepiness (15).

In addition, our study highlights the difference between being able to differentiate groups and the ability to exclude individuals. The variables by themselves, although able to separate the patients into groups of different apnea indices, were inadequate in assuring that a patient with an AI < 20 would not be missed. In combination, the results were improved. About 9.3% (33) of our 354 patients screened by the combination of BMI and ESS were expected to have AI < 20; only two of them ended up having AI ≥20. Putting such patients in a low-priority group would have put two people at risk. Although adding observed apnea to ESS and BMI resulted in no patients being missed, only nine of the 354 patients were detected by this combination.

What other inexpensive evaluation may help in prioritizing patients for further investigation? We believe that overnight oximetry may be helpful in this setting. We have previously shown that overnight oximetry was able to exclude patients with severe apnea (16), but, of course, other sleep disorders (narcolepsy, periodic movements, upper airway resistance syndrome, etc.) would likely be missed. Rauscher et al. (5) showed that adding oximetry results to a regression model for predicting OSA resulted in an improved ability to detect patients with AI < 10.

We conclude that attempting to prioritize patients with suspected sleep apnea on the basis of ESS, BMI, and whether apnea is observed may be helpful when these variables are used in combination. This does not mean that such patients screened to have low AI should not be studied since they can still have significant morbidity and sleepiness. It means that such patients may be studied less emergently.

Acknowledgements: The authors would like to thank David Anyadike, Ken Delaive, and Kori Foran for their assistance in collecting and analyzing questionnaire data.

REFERENCES

1. Kapuniai LE, Andrew DJ, Crowell DH, Pearce JW. Identifying sleep apnea from self-reports. *Sleep* 1988;11(5):430–6.
2. Crocker BD, Olson LG, Saunders NA, et al. Estimation of the probability of disturbed breathing during sleep before a sleep study. *Am Rev Respir Dis* 1990;142:14–8.
3. Viner S, Szalai MP, Hoffstein V. Are history and physcial examination a good screening test for sleep apnea? *Ann Intern Med* 1991;115:356–9.
4. Hoffstein V, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. *Sleep* 1993;16(2):118–22.
5. Rauscher H, Popp W, Zwick H. Model for investigating snorers with suspected sleep apnoea. *Thorax* 1993;48:275–9.
6. Kump K, Whalen C, Tishler PV, et al. Assessment of the validity and utility of a sleep-symptom questionnaire. *Am J Respir Crit Care Med* 1994;150:735–41.
7. Douglass AB, Bornstein R, Nino-Murcia G, et al. The sleep disorders questionnaire I: creation and multivariate structure of SDQ. *Sleep* 1994;17(2):160–7.
8. Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. *Sleep* 1995;18(3):158–66.
9. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14(6):540–5.
10. U.S. Department of Agriculture, U.S. Department of Health and Human Services. *Nutrition and your health: dietary guidelines for Americans.* 3rd edition. Washington, DC: U.S. Government Printing Office, 1990.
11. Willett WC, Manson JE, Stampfer MJ, et al. Weight, weight change, and coronary heart disease in women: risk within the 'normal' weight range. *JAMA* 1995;273:461–5.
12. Martin RJ, Block AJ, Cohn MA, et al. Indications and standards for cardiopulmonary sleep studies. *Sleep* 1985;8(4):371–9.
13. He J, Kryger MH, Zorick FJ, Conway W, Roth T. Mortality and apnea index in obstructive sleep apnea: experience in 385 male patients. *Chest* 1988;94:9–14.
14. Scharf SM, Garshick E, Brown R, et al. Screening for subclinical sleep-disordered breathing. *Sleep* 1990;13(4):344–53.
15. George CF, Kryger MH. Sleep and sleepiness and the pulmonologist. In: Simmons DH, ed. *Current pulmonology.* Chicago: Year Book Medical Publishers, 1990.
16. Yamashiro Y, Kryger MH. Nocturnal oximetry: is it a screening tool for sleep disorders? *Sleep* 1995;18(3):167–71.