Are Sleep Questionnaires Valid in All Adult Age Groups as Screening Tools for Obstructive Sleep Apnea?

Ho Min Lee, MD, Jae Ki Kim, MD, Jung Gwon Nam, MD, PhD, and Tae-Hoon Lee, MD, PhD

Department of Otolaryngology-Head and Neck Surgery, Ulsan University Hospital, University of Ulsan College of Medicine, Ulsan, Korea

Background and Objectives: Evaluation of Epworth Sleepiness Scale (ESS), Berlin, STOP, and STOP-Bang questionnaire validities for obstructive sleep apnea (OSA) screening among various adult age groups.
Materials and Method: Results for each of those questionnaires were compared with diagnostic overnight polysomnography (PSG) data obtained for 396 patients suffering either insomnia, sleep apnea, excessive daytime sleepiness, or chronic snoring who had been divided into three age groups (20−39, 40−59, or ≥60 years). For each questionnaire, the sensitivity, specificity, accuracy, and area under the curve (AUC) were calculated.
Results: Among the OSA group [apnea hypopnea index (AHI) cutoff >5], Berlin and STOP questionnaire sensitivity and specificity were significantly different among the age groups. Among the moderate-to-severe OSA sub-group (AHI cutoff >15), the specificity of Berlin, STOP, and STOP-Bang questionnaire was significantly different among age groups.
Conclusion: The Berlin and STOP questionnaires differed with patient age in OSA screening. The ESS questionnaire, by contrast, did not show any age-related differences of sensitivity and specificity in OSA screening or moderate-to-severe OSA screening.

KEY WORDS: Obstructive sleep apnea · OSA · Screening · Questionnaire · Age.

INTRODUCTION

Obstructive sleep apnea (OSA), occurring in 26% of people aged 30−70 years, manifests as repetitive apnea and hypopnea episodes related to sleep-time upper-airway obstruction.[1] OSA causes sleep fragmentation, is associated with excessive daytime sleepiness and decreased concentration, and has been linked to metabolic disorders, cardiovascular and neurovascular diseases, and neurocognitive impairment.[2-4]

The Epworth Sleepiness Scale (ESS), Berlin, STOP, and STOP-Bang questionnaires have been developed as OSA-screening-tool alternatives to the expensive and time-consuming polysomnography (PSG) test. There have already been several reports published on the sensitivity and specificity of these screening tests.[5-8]

For most sleep questionnaires, respondent age can be problematic. Hypertension, which longitudinal and cross-sectional studies have definitively shown to be strongly dose-response-related to OSA severity, is included in most of the current questionnaires.[9,10] The problem that arises is that hypertension is significantly correlated not only with OSA severity but also with older age. Therefore many relatively young OSA patients, even those with severe OSA, do not have hypertension, which can lead to false negatives (i.e., OSA sufferers deemed "no OSA"). Also, age usually is related to both social activity and physical strength, both of which can affect the answers to questions on sleep-de-

privation-caused daytime fatigue. However, age is not adjusted for in the STOP or Berlin questionnaires or indeed in several others. On this basis, we hypothesized that the reliability of sleep questionnaires could differ according to patient age. In our study, we evaluated the ESS, Berlin, STOP, and STOP-Bang questionnaire validities as OSA-screening tests for various adult age groups.

MATERIALS AND METHODS

We retrospectively reviewed the medical records of patients who referred to a sleep clinic with symptoms such as insomnia, snoring, sleep apnea or excessive daytime sleepiness between April 2012 and May 2016. A total of 396 patients aged 20 years or over were enrolled in this study. After categorizing the patients into three groups according to age (20−39, 40−59, or ≥60 years), patient data, including age, sex, height, weight, history of hypertension and overall medical history, were collected. The ESS, Berlin, STOP, and STOP-Bang questionnaires, which had been translated into Korean and certified, and the diagnostic overnight PSG results were also collected. Those for whom the total sleep time was under 180 min in the PSG were excluded. The relevant Institutional Review Board approved this study.

The epworth sleepiness scale

The Epworth Sleepiness Scale (ESS) is a self-reported questionnaire developed by Dr. Murray Johns for assessment of excessive daytime sleepiness as well as daytime sleep propensity.[5] It consists of eight items, being daily routines for which respondents assess their propensity to fall asleep on a scale from 0 to 3. Those scoring 11 or over are deemed to be at high risk of hypersomnia.

The berlin questionnaire

The Berlin questionnaire is a widely used OSA-assessment tool that includes 11 questions broken down into three categories (snoring, daytime sleepiness, hypertension). For category 1 (snoring), "high risk" is persistent symptoms (>3−4 times/week) in at least two questions. For category 2 (daytime sleepiness), "high risk" is persistent symptoms (>3−4 times/week) concerning either daytime sleepiness generally, drowsiness during driving specifically, or both. For category 3 (hypertension), "high risk" is marked by either a body mass index (BMI) of more than 30 kg/m^2 or a history of high blood pressure. A patient whose scores indi-

cate positivity for high risk in at least two categories is classified as high risk for OSA.[6]

The STOP questionnaire

The OSA-screening STOP questionnaire was designed to reduce some of the risk associated with general anesthesia. It defines high risk as positivity for two or more of the following: high blood pressure, tiredness during the day, snoring, and observed apnea.[7]

The STOP-bang questionnaire

The STOP-Bang questionnaire employs the STOP questionnaire to evaluate subjects for the following risk factors: age >50 years, BMI >35 kg/m^2, neck circumference >40 cm, and male sex. High risk status is indicated by affirmative answers to three or more of the eight questions (items) included in the questionnaire.[7]

Statistical analysis

The sensitivity, specificity, accuracy, and area under the curve (AUC) values for all of the questionnaires for each age group were analyzed according to receiver operating characteristic (ROC) curves. The significances of the inter-age-group differences were evaluated by chi-square test. A p-value <0.05 was considered statistically significant.

RESULTS

The patients were each assigned to one of three groups based on age: 20−39 (young adult), 40−59 (middle-aged), and 60 and over (elderly). Table 1 summarizes the patients' characteristics. The 396-patient cohort comprised 332 males and 64 females of 47 years mean age. Gender significantly differed among the three age groups (p=0.002). Hypertension, identified in 100 patients (25.3%) also showed significant differences by age group (p=0.000). The OSA group (AHI >5) included 313 patients (79.0%), and the normal (control) group, 83 patients (21.0%). There were significant OSA-distribution differences among the age groups (p=0.022). Within the OSA group, the moderate-to-severe OSA sub-group (AHI >15) included 137 patients (34.6%), and the severe OSA sub-group (AHI >30), 64 patients (16.2%). There were also significant differences in these two groups' respective distributions among the age groups (p=0.000) (Table 1).

For the OSA group (AHI >5), sensitivity and accuracy

were highest for the STOP-Bang questionnaire, specificity for the ESS, and AUC for the Berlin questionnaire. The specificity and specificity of the Berlin and the STOP questionnaire significantly differed by age. As for the Berlin questionnaire's AUC, the confidence interval did not include 0.5 for the young-adult or middle-aged group. However, it did not include 0.5 only for the young-adult in the STOP and STOP-Bang questionnaire's AUC (Table 2).

Meanwhile, about the moderate-to-severe OSA sub-group (AHI >15), the STOP-Bang questionnaire showed the highest sensitivity, the ESS the highest specificity and accuracy, and the Berlin questionnaire the highest AUC. By age group, the Berlin, STOP, and STOP-Bang questionnaire's speci-

ficity significantly differed. As for the Berlin and STOP-Bang questionnaires' AUC, the confidence interval did not include 0.5 for the young-adult or middle-aged group (Table 3).

DISCUSSION

A comprehensive approach including physical examination, thorough history-taking, and overnight PSG is required for accurate diagnosis of OSA. Complaints of daytime sleepiness, fatigue and snoring should be considered suspicious of OSA. Additionally, the presence of any underlying disease, not to mention apnea-related information from a family member of a person living with the patient, also should

Table 1. Demographic characteristics and PSG findings for all age groups

	Total	20−39	40−59	60−	p
Number	396	119	192	85	
Gender (male/female)	332/64 (83.8/16.2)	110/9 (92.4/7.6)	159/33 (82.8/17.2)	63/22 (74.1/25.9)	0.002*
Hypertension	100 (25.3)	18 (15.1)	37 (19.3)	45 (52.9)	0.000*
OSA					
All (AHI>5)	313 (79.0)	84 (70.6)	157 (81.8)	72 (84.7)	0.022*
Mod/Sev (AHI>15)	137 (34.6)	31 (26.1)	62 (32.3)	44 (51.8)	0.000*
Severe (AHI>30)	64 (16.2)	7 (5.9)	28 (14.6)	29 (34.1)	0.000*

Differences in overall scores were evaluated using the chi-squared test. *: p<0.05. PSG: polysomnography, OSA: obstructive sleep apnea, AHI: apnea-hypopnea index, Mod/Sev: moderate/severe

Table 2. Predictive parameters of sleep questionnaires by age group (AHI>5 as cutoff). The age intervals are in years

	Total	20−39	40−59	60−	p
ESS					
Sensitivity %	43.8	44.0	48.4	33.3	0.102
Specificity %	66.3	65.7	62.9	76.9	0.777
Accuracy %	48.5	50.4	51.0	40.0	0.216
AUC (95% CI)	0.550 (0.492−0.608)	0.549 (0.453−0.645)	0.556 (0.466−0.647)	0.551 (0.420−0.683)	0.096
Berlin					
Sensitivity %	74.8	76.2	68.8	86.1	0.016*
Specificity %	48.2	54.3	54.3	15.4	0.039*
Accuracy %	69.2	69.7	66.1	75.3	0.321
AUC (95% CI)	0.615 (0.556−0.674)	0.652 (0.557−0.748)	0.615 (0.524−0.707)	0.508 (0.398−0.617)	
STOP					
Sensitivity %	82.4	78.6	80.3	91.7	0.048*
Specificity %	32.5	51.4	20.0	15.4	0.010*
Accuracy %	72.0	70.6	69.3	80.0	0.168
AUC (95% CI)	0.575 (0.520−0.630)	0.650 (0.555−0.745)	0.501 (0.427−0.575)	0.535 (0.428−0.642)	
STOP-Bang					
Sensitivity %	91.1	88.1	93.0	90.3	0.401
Specificity %	24.1	34.3	20.0	7.7	0.128
Accuracy %	77.0	72.3	79.7	77.6	0.320
AUC (95% CI)	0.576 (0.527−0.625)	0.612 (0.525−0.699)	0.565 (0.495−0.635)	0.49 (0.407−0.573)	

Differences in overall scores were evaluated using the chi-squared test. *: p<0.05. ESS: Epworth Sleepiness Scale, SBQ: STOP-Bang questionnaire, AUC: area under the curve

Table 3. Predictive parameters of sleep questionnaires by age group (AHI >15 as cutoff). The age intervals are in years

	Total	20−39	40−59	60−	p
ESS					
Sensitivity %	46.7	41.9	54.8	38.6	0.211
Specificity %	61.0	59.1	57.7	75.6	0.108
Accuracy %	56.1	54.6	56.8	56.5	0.929
AUC (95% CI)	0.539 (0.487−0.590)	0.505 (0.403−0.607)	0.563 (0.487−0.638)	0.571 (0.473−0.571)	
Berlin					
Sensitivity %	81.0	83.9	75.8	86.4	0.358
Specificity %	35.9	38.6	40.8	14.6	0.005*
Accuracy %	51.5	50.4	52.1	51.8	0.964
AUC (95% CI)	0.585 (0.541−0.629)	0.613 (0.529−0.696)	0.583 (0.514−0.651)	0.505 (0.430−0.580)	
STOP					
Sensitivity %	88.3	83.9	87.1	93.2	0.403
Specificity %	25.5	35.2	23.1	12.2	0.014*
Accuracy %	47.2	47.9	43.8	54.1	0.273
AUC (95% CI)	0.569 (0.531−0.607)	0.596 (0.513−0.678)	0.551 (0.495−0.607)	0.527 (0.464−0.590)	
STOP-Bang					
Sensitivity %	94.2	96.8	96.8	88.6	0.225
Specificity %	15.4	23.9	12.3	7.3	0.025*
Accuracy %	42.7	42.9	39.6	49.4	0.312
AUC (95% CI)	0.548 (0.519−0.578)	0.603 (0.548−0.658)	0.545 (0.509−0.581)	0.480 (0.418−0.542)	

Differences in overall scores were evaluated using the chi-squared test. *: p<0.05. ESS: Epworth Sleepiness Scale, SBQ: STOP-Bang questionnaire, AUC: area under the curve

be considered.[11] However, history-taking and physical examination are insufficient for distinguishing primary snoring from sleep apnea. Thus, supplementary overnight PSG, the current standard, is considered necessary for definitive OSA.[12]

Because nearly 80% of men and approximately 93% of women with moderate-to-severe sleep apnea go undiagnosed, necessary screening tests must be further developed and improved to increase diagnosis rates.[13] Indeed, several studies focusing on the various questionnaires and the clinical OSA-associated factors have done just this.[5-8]

The Berlin questionnaire has been widely employed since its introduction at the 1996 Conference on Sleep in Primary Care in Berlin, Germany. It is used to investigate excessive daytime sleepiness, the presence of fatigue and hypertension, snoring, and sleep apnea. One OSA-diagnostics study reported sensitivities ranging between 54 and 86%, specificities between 43 and 87%, and a PPV of 89%.[6] The OSA-screening STOP questionnaire was designed to reduce some of the risk associated with general anesthesia. Significantly, this tool's sensitivity has been found to improve with higher AHI: with an AHI $\geq$5, the sensitivity was 65.6%, with an AHI of 15, 74.3%, and with an AHI of 30, 79.5%.[7] The STOP-Bang questionnaire has been demonstrated to offer higher sensitivity and NPV, especially for moderate-to-severe OSA patients.[7] In light of these three (Berlin, STOP, STOP-Bang) questionnaires' relative commonness, less complicated criteria, and relative ease of application in the clinical setting, we evaluated them in the present study. We additionally evaluated the ESS that was designed originally for assessment of excessive daytime sleepiness, as it has been widely utilized by primary care providers to identify potential sleep disorders.

Two studies that compared STOP and STOP-Bang with Berlin and/or ESS, determining STOP-Bang to have the highest sensitivity and AUC, recommended it as the better OSA-screening tool.[14)15] Another, a meta-analysis of OSA-screening tests found Berlin to be the most accurate questionnaire for prediction of OSA diagnosis. The least precise questionnaire was determined to be ESS, due possibly to the fact that excessive daytime sleepiness is a common syndrome among obese persons without OSA, as driven by other-than-nighttime-sleep-deprivation mechanisms. The simplest questionnaire, STOP, was deemed a relatively poor predictor of OSA, while STOP-Bang was judged to be average.[16]

As for the effect of age on OSA, increasing age has been correlated with both increasing prevalence and decreasing

severity.[17] Male dominance for OSA prevalence and severity disappears above 55 years of age.[18][19] The present data showed similar results: the female ratio was found to have been significantly increased in the elderly group. Overall OSA prevalence increased significantly with age.

Also in the current study, hypertension prevalence increased significantly with age; in fact, it was markedly higher in the elderly group than in either of the other age groups. OSA prevalence, meanwhile, was much higher in both the middle-aged and elderly groups than in the young-adult group. These results strongly suggest the possibility of age-related prevalence differences between hypertension and OSA. Notably, this might have influenced the Berlin and STOP questionnaire's sensitivity and specificity differences among the different age groups because hypertension plays a relatively significant role in the Berlin and STOP questionnaires.[9][10]

The ideal diagnostic test for an otherwise healthy population would offer the following: relatively high sensitivity, sufficient specificity, minimal intrusiveness, relative cheapness, and a utility for identification of patients early in the disease course.[20] We found that the ESS showed the lowest sensitivity; the STOP-Bang questionnaire, while affording the highest sensitivity, as found in previous studies, offered only inadequate specificity; the Berlin questionnaire's AUC results showed it to be recommendable for evaluation of patients with suspected OSA in young-adult and middle-aged groups, notwithstanding its relatively lengthy question list and complicated scoring system.[14][15]

There are some research limitations. For example, both the OSA patients and the normal (control) group showed sleep-related symptoms. Also, the number of patients in the elderly group was small. Furthermore, we were unable to examine patient occupation and lifestyle, both of which factors might have affected the results.

In conclusion, the Berlin and the STOP questionnaire showed significant sensitivity and specificity differences among the age groups. Also, for the moderate-to-severe OSA patients, there were the Berlin, STOP, and STOP-Bang questionnaire differences in specificity among the age groups. The ESS questionnaire, however, did not show any age-related differences.

REFERENCES

1) Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. Am J Epidemiol 2013;177:1006-14.

2) Redline S, Strohl KP. Recognition and consequences of obstructive sleep apnea hypopnea syndrome. Clin Chest Med 1998;19:1-19.

3) Mo JH. Obstructive Sleep Apnea and Systemic Diseases. J Rhinol 2013;20:8-13.

4) McNicholas WT, Bonsignore MR; Management Committee of EU COST ACTION B26. Sleep apnoea as an independent risk factor for cardiovascular disease: current evidence, basic mechanisms and research priorities. Eur Respir J 2007;29:156-78.

5) Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540-5.

6) Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131:485-91.

7) Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology 2008;108:812-21.

8) Flemons WW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. Am J Respir Crit Care Med 1994;150:1279-85.

9) Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. N Engl J Med 2000;342:1378-84.

10) Nieto FJ, Young TB, Lind BK, Shaher E, Samet JM, Redline S, et al. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. JAMA 2000;283:1829-36.

11) Haponik EF, Frye AW, Richards B, Wymer A, Hinds A, Pearce K, et al. Sleep history is neglected diagnostic information. Challenges for primary care physicians. J Gen Intern Med 1996;11:759-61.

12) Gondim LM, Matumoto LM, MeloJúnior MA, Bittencourt S, Ribeiro UJ. Comparative study between clinical history and polysomnogram in the obstructive sleep apnea/ hypopnea syndrome. Braz J Otorhinolaryngol 2007;73:733-7.

13) Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 1997;20:705-6.

14) Silva GE, Vana KD, Goodwin JL, Sherrill DL, Quan SF. Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth sleepiness scales. J Clin Sleep Med 2011;7:467-72.

15) Luo J, Huang R, Zhong X, Xiao Y, Zhou J. STOP-Bang questionnaire is superior to Epworth sleepiness scales, Berlin questionnaire, and STOP questionnaire in screening obstructive sleep apnea hypopnea syndrome patients. Chin Med J 2014;127:3065-70.

16) Ramachandran SK, Josephs LA. A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology. 2009;110:928-39.

17) Bixler EO, Vgontzas AN, Ten Have T, Tyson K, Kales A. Effects of age on sleep apnea in men: I. Prevalence and severity. Am J Respir Crit Care Med 1998;157:144-8.

18) Resta O, Caratozzolo G, Pannacciulli N, Stefàno A, Giliberti T, Carpagano GE, et al. Gender, age and menopause effects on the prevalence and the characteristics of obstructive sleep apnea in obesity. Eur J Clin Invest 2003;33:1084-9.

19) Chung YS, Jang YJ, Lee BJ, Lee SA, Choi SJ, Kang WS, et al. Gender Differences of Polysomnographic Findings in Snoring Patients. J Rhinol 2004;11:48-51.

20) Ross SD, Sheinhait IA, Harrison KJ, Kvasz M, Connelly JE, Shea SA, et al. Systemic review and meta-analysis of the literature regarding the diagnosis of sleep apnea. Sleep 2000;23:519-32.