Utility of Combination of Sleep Questionnaires in Predicting Obstructive Sleep Apnea and its Correlation with Polysomnography

Patro Mahismita1, Ojha U Chandra2, Gothi Dipti3, Jain Anshul4, Palai Shrabani5, Sah R Baboo6, Vaidya Sameer7

ABSTRACT

Introduction: A number of screening questionnaires and clinical screening models have been developed to identify patients with obstructive sleep apnea syndrome (OSAS). These questionnaires lack reliability, and their predictability varies. Hence, it is difficult to predict or rule out OSAS on one questionnaire alone. The combination of two or more questionnaires might be helpful in ruling out OSAS.

Objectives: (1) To determine the sensitivity, specificity, and predictive values of combination of two or more sleep questionnaires out of three established sleep questionnaires, i.e., Epworth sleepiness scale (ESS), perioperative sleep apnea prediction score (PSAP), STOP-Bang, in predicting OSAS and correlation with severity of OSAS. (2) To compare and correlate ESS, PSAP, and STOP-Bang individually with apnea–hypopnea index (AHI) obtained by polysomnography (PSG).

Materials and methods: It was a prospective observational study conducted in a tertiary care center from January 2018 to August 2019 involving 250 cases of suspected OSAS. All the participants were interviewed for the three questionnaires followed by diagnostic type I PSG.

Results: Comparing the individual questionnaires, ESS had higher sensitivity but low specificity, whereas PSAP had higher specificity. Perioperative sleep apnea prediction [area under curve (AUC) = 0.743 for any OSAS and 0.722 for moderate-to-severe OSAS] had a better prediction for OSAS. For predicting any OSAS, the combination of STOP-Bang + ESS + PSAP had a sensitivity of 95.76, specificity of 24.59%, and high negative predictive value (NPV) of 65.22%. For predicting moderate-to-severe OSAS, the combination of STOP-Bang + ESS + PSAP had a sensitivity of 92.59, specificity of 36.06%, and high NPV of 61.11%.

Conclusion: The combination of questionnaires especially STOP-Bang, ESS, and PSAP improves the sensitivity of detection up to 95%, and when all of them are negative, OSAS is ruled out with around 65% confidence. So, using this combination can help us to identify high-risk patients and prioritize them for PSG so that they can get early treatment.

Keywords: Obstructive sleep apnea, Polysomnography, Sleep questionnaires.

Indian Journal of Sleep Medicine (2019): 10.5005/jp-journals-10069-0046

INTRODUCTION

Obstructive sleep apnea syndrome (OSAS) is a highly prevalent disease, characterized by repetitive pattern of upper airway collapsibility and airflow obstruction resulting in recurrent arousals. Obstructive sleep apnea syndrome is defined as the occurrence of an average 5 or more episodes of obstructive respiratory events per hour of sleep with either sleep-related symptoms or comorbidities or ≥15 such episodes without any sleep-related symptoms or comorbidities. The prevalence of OSAS varies from 2 to 17% in men and 1 to 9% in women.1–4 Studies from India also have a similar prevalence of 2.4–4.96% in men and 1–2% in women.3,5–8 Untreated OSAS is recognized as a major risk factor for cardiovascular as well neurological morbidity and mortality and represents an increasing burden on healthcare resources.9

It is estimated that nearly 82% of men and 93% of women with moderate-to-severe sleep apnea are undiagnosed.10 Although the gold standard for diagnosis of OSAS is laboratory polysomnography (PSG); however, the occurrence of OSAS is far more prevalent than can be handled by the available sleep laboratories. Therefore, a screening tool is necessary to stratify patients to high risk or low risk of OSAS. This will help to identify those in urgent need of PSG to confirm the presence and severity of OSAS and/or further treatment for the same.10

A number of screening questionnaires and clinical screening models have been developed to help identify patients with OSAS. Out of those, Epworth sleepiness scale (ESS), perioperative sleep apnea prediction (PSAP), and STOP-Bang are points based, easy to administer, and can be subjected to outpatient settings. However, utility of sleep questionnaire is limited as their validation studies were conducted on different populations, and so the sensitivity and specificity of these tests varies widely. STOP-Bang has high sensitivity and thus high screening capability but lacks specificity and so has higher chances of false-positive detection. Epworth sleepiness scale is claimed to have high prediction of excessive
daytime sleepiness. The PSAP uses upper airway elements such as high modified Mallampati grade and reduced thyromental distance that are commonly assessed perioperatively and includes type II diabetes.11 The sleep questionnaires are very useful and effective tool for the evaluation and prediction of OSAS in suspected patients. There is no guideline on which questionnaire has to be used in which context. These sleep questionnaires lack reliability on some aspects, and their predictability varies widely. Hence, it is difficult to predict or rule out OSAS on one questionnaire alone owing to their variable sensitivity and specificity. The combination of two or more questionnaires might be helpful in ruling out OSAS. To the best of our knowledge, there are very few studies in literature using the combination of two or more questionnaires but only one from India.12 Earlier Ulasli et al. advocated the use of ESS and Berlin questionnaire together as they had a slightly better performance when used together than when they are taken alone in predicting OSAS.13 Saxena et al. from India have used the combination of STOP-Bang + ESS + PSAP and found a very high negative predictive value (NPV) of 90%.12 So, this study was carried out to find out if the combination of questionnaires will help rule out OSAS.

AIMS AND OBJECTIVES

- To determine the sensitivity, specificity, and predictive values of combination of two or more sleep questionnaires out of three established sleep questionnaires, i.e., ESS, STOP-Bang, and PSAP, in predicting OSAS and correlation with severity of OSAS. Following combinations will be studied.
  - ESS and PSAP
  - ESS and STOP-Bang
  - PSAP and STOP-Bang
  - ESS, STOP-Bang with PSAP
- To compare and correlate ESS, PSAP, and STOP-Bang individually with apnea–hypopnea index (AHI) obtained by PSG.

MATERIALS AND METHODS

It was a prospective observational study conducted in a tertiary care center of North India from January 2018 to August 2019 after obtaining ethics committee approval. The inclusion criteria were patients with age ≥18 years having a clinical suspicion of OSAS based on the history of snoring and/or excessive daytime sleepiness. The exclusion criteria were those who had received treatment for OSAS, those with any active psychiatric disorder, and those having exacerbation of respiratory disease or acute myocardial infarction within last 4 weeks.

Study Method

Written and informed consent was obtained from those individuals fulfilling the inclusion criteria. The study was approved by the institutional ethics committee. Patients were asked for their detailed history and subjected to thorough clinical examination. The clinical evaluation included demographics; symptoms of snoring, witnessed apnea, and excessive daytime sleepiness; patient’s vital parameters; and anthropometric measurements [body mass index (BMI), waist circumference, and neck circumference]. Body mass index was calculated using the standard formula: weight (kg)/height² (m²). The neck circumference was measured at the level of the cricothyroid membrane using a measuring tape. Waist circumference was measured at the level of midpoint between the top of the iliac crest and the lower margin of the last palpable rib in the midaxillary line. SpO₂ was measured using daytime pulse oximetry in sitting position. All study patients were interviewed for the three OSAS prediction questionnaires, i.e., ESS, STOP-Bang, and PSAP.

ESS

This is a self-administered questionnaire that asks subjects to rate their chances of falling asleep, which is scored in a scale from 0 to 3 (0 = would never doze, 1 = slight chance of dozing, 2 = moderate chance of dozing, and 3 = high chance of dozing) in eight situations that are routinely encountered in daily life. The total ESS score is the sum of eight-item scores and can range between 0 and 24. The higher the score, the higher the person’s level of daytime sleepiness. Here, we have used the Hindi language ESS. The Hindi language ESS is modified for the Indian sociocultural characteristics (the ESS-I) and is also reliable, valid, and comparable with the original scale.14,15

STOP-Bang

The STOP-Bang questionnaire is a scoring model consisting of eight easily administered questions starting with the acronym STOP-Bang and is scored based on yes/no answers (score: 1/0). Thus, the scores range from a value of 0–8.16

PSAP

The PSAP score validates six of the eight elements of the STOP-Bang model but differs in that it uses high modified Mallampati class and reduced thyromental distance and includes type II diabetes. It has a total of nine parameters: three demographic measures, three history variables, and three airway measures. Each parameter is scored one point, and the total score is calculated.17 All the subjects underwent blood investigations (hemoglobin, total and differential leukocyte count, blood urea, creatinine, sugar, liver function test, lipid profile, and thyroid profile), chest radiograph, two-dimensional echocardiograph, and spirometry to find out any other comorbidity. All the patients were subjected to diagnostic type I PSG. The full-night attended PSG was performed on Philips Alice S5® polysomnograph. It includes recording of electroencephalogram (central and occipital), electrooculogram, submental and pretibial electromyography, oronasal flow (thermistor and nasal pressure transducer), thoracoabdominal movements, and oxygen saturation. Sleep stages and respiratory events were scored according to the American Academy of Sleep Medicine (AASM) guidelines. The scoring was done manually by the sleep specialist blinded to the results of the questionnaires. The diagnosis of OSAS was based on the AASM guidelines. The severity of OSAS was classified based on the AHI values: mild OSAS, AHI ≥ 5 and <15/hour; moderate OSAS, AHI ≥ 15 and <30/hour; and severe OSAS, AHI ≥ 30/hour.

Statistical Analysis

A statistical analysis was performed with the help of a statistician using SPSS software version 17. The sample size was calculated based on the following formula: 

\[ n = \frac{Z_{\alpha/2}^2 \times \sigma^2}{d^2}, \]

where 

- \( n \) = no. of patients required,
- \( Z \) = confidence interval of 95%,
- \( \sigma \) = anticipated population proportion, and
- \( d \) = precision required on either side of proportion.18 Considering a precision of 8%, a minimum sample size of 150 patients was required. All the data were presented as mean ± standard deviation. Data were compared among patients with OSAS and those
without OSAS using Student’s t test, χ², or Mann–Whitney U test wherever applicable. A correlation analysis was done to find out the correlation between the questionnaires with AHI. The cutoff value of each questionnaire was calculated from the receptor operating characteristics (ROC) analyses. Using those cutoffs, various predictive parameters, i.e., sensitivity, specificity, positive predictive value (PPV), and NPV of each questionnaire and all possible combinations were calculated. The polysomnographic findings were taken as the gold standard for the diagnosis of OSAS. Results of all three questionnaires were compared individually and taken together. The p value of <0.05 was considered to be statistically significant.

**Results**

A total of 250 cases were included in the study. The mean age was 50.42 ± 10.42 years, and mean BMI was 31.42 ± 6.34 kg/m² with 70.8% being men. The prevalence of OSAS was 75.6% (189/250). Table 1 shows the demographic characteristics, anthropometric parameters, and sleep-related parameters of OSAS patients. The mean age of OSAS patients was 51.19 ± 10.38 years, the average BMI was 32.27 ± 6.15 kg/m², and average AHI was 33.24 ± 23.68 events/hour. Most of the patients, i.e., 46.03%, had severe OSAS followed by moderate OSAS in 26.98% patients and mild OSAS in 26.98% patients.

Three questionnaires, STOP-Bang, ESS, and PSAP, were used for determining the probability of OSAS. On the correlation analysis, we found a significant positive correlation of all the three questionnaires with AHI (Table 2). Receptor operating characteristics curves were utilized to determine the cutoffs for predicting any OSAS (AHI ≥ 5) and moderate-to-severe OSAS (AHI ≥ 15) separately (Fig. 1). Table 3 shows the area under curve (AUC) and cutoffs derived for each of the questionnaires for predicting OSAS. Perioperative sleep apnea prediction score (AUC = 0.743 for any OSAS and 0.722 for moderate-to-severe OSAS) had a better prediction for OSAS than STOP-Bang and ESS according to ROC.

Using those cutoffs (rounded off to the nearest figure), the sensitivity, specificity, PPV, and NPV of each questionnaire and various combinations of them were obtained to predict any OSAS and moderate-to-severe OSAS separately (Table 4). Figure 2 shows the line diagram of the predictive parameters of the questionnaires for any OSAS (Fig. 2A) and moderate-to-severe OSAS (Fig. 2B). The cutoff obtained for STOP-Bang questionnaire for predicting any OSAS and moderate-to-severe OSAS was same at 4.5. For predicting any OSAS (AHI ≥ 5), the score was taken positive if >4 for STOP-Bang, >9 for ESS, and >4 for PSAP. For predicting moderate-to-severe OSAS (AHI ≥ 15), the score was taken positive if >4 for STOP-Bang, >10 for ESS, and >5 for PSAP.

Comparing the individual questionnaires, ESS had higher sensitivity but low specificity when compared with STOP-Bang and PSAP (sensitivity of ESS is 81.48% for any OSAS and 77.25% for moderate-to-severe OSAS), whereas PSAP had higher specificity. On combining the questionnaires, the sensitivity increased at the cost of reduced specificity.

### Table 1: Characteristics of obstructive sleep apnea syndrome patients

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Observed value in OSAS patients (n = 189)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic characteristics</td>
<td></td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>51.19 ± 10.38</td>
</tr>
<tr>
<td>Sex (male/female)</td>
<td>140:49</td>
</tr>
<tr>
<td>Anthropometric parameters</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>32.27 ± 6.15</td>
</tr>
<tr>
<td>Neck circumference (cm)</td>
<td>40.44 ± 7.32</td>
</tr>
<tr>
<td>Mallampati grade</td>
<td>2.92 ± 0.98</td>
</tr>
<tr>
<td>Thyromental distance (cm)</td>
<td>7.14 ± 1.56</td>
</tr>
<tr>
<td>Sleep questionnaires scores</td>
<td></td>
</tr>
<tr>
<td>STOP-Bang</td>
<td>5.13 ± 1.51</td>
</tr>
<tr>
<td>ESS</td>
<td>14.79 ± 6.22</td>
</tr>
<tr>
<td>PSAP</td>
<td>5.6 ± 1.77</td>
</tr>
<tr>
<td>AHI (events/hour)</td>
<td>33.24 ± 23.68</td>
</tr>
</tbody>
</table>

### Table 2: Correlation coefficient of the questionnaires with apnea–hypopnea index

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Correlation coefficient (r)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>STOP-Bang</td>
<td>0.34</td>
<td>0.004</td>
</tr>
<tr>
<td>ESS</td>
<td>0.39</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PSAP</td>
<td>0.33</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Figs 1A and B: (A) Receptor-operating characteristics curves for predicting any obstructive sleep apnea syndrome; (B) Moderate-to-severe obstructive sleep apnea syndrome.
Sleep Questionnaires in Predicting Obstructive Sleep Apnea

Among the combination of two questionnaires, for predicting any OSAS, the sensitivity of ESS + PSAP was very high (93.65%), whereas specificity of STOP-Bang + PSAP was higher (50.81%). The combination of ESS + PSAP also had a higher NPV of 60%. All the three questionnaires taken together (STOP-Bang + ESS + PSAP) had a sensitivity of 95.76%, specificity of 24.59%, and high NPV of 65.22%. For predicting moderate-to-severe OSAS, sensitivity of STOP-Bang + ESS was high, whereas specificity of STOP-Bang + PSAP was higher. A combination of STOP-Bang + ESS + PSAP had a sensitivity of 92.59, specificity of 36.06%, and high NPV of 61.11%.

**DISCUSSION**

This study was aimed at comparing various established pretest probability questionnaires regarding their predictive probabilities for OSAS. We studied three questionnaires (STOP-Bang, ESS, and PSAP) and their various combinations. These questionnaires were evaluated in patients of suspected OSAS and were compared against the PSG-based AHI serving as the “gold standard” diagnosis for OSAS.

Among the OSAS prediction tools used in our study, a correlation analysis, all three (STOP-Bang, ESS, and PSAP) correlated positively with AHI. This was consistent with earlier studies. Most studies consider STOP-Bang score of ≥3 as high risk for OSAS. On increasing the cutoff, the specificity increases at the cost of reduced sensitivity. In our study, the cutoff obtained for predicting OSAS and moderate-to-severe OSAS was same at ≥5 and higher when compared with previous studies. Using this cutoff, the sensitivity and specificity of STOP-Bang were 65.60 and 65.57%, respectively, for predicting both OSAS and moderate-to-severe OSAS. Our study reported a lower sensitivity but higher specificity for prediction of OSAS when compared with previous studies. The range of sensitivities reported by previous studies varies from 83.9% to 97% for detection of OSAS and from 92.9 to 95% for moderate-to-severe OSAS. However, the specificities of these tests were low, ranging from 18% to 40%. Some studies have used a BMI cutoff of >35 kg/m² when compared with the standard of BMI >35 kg/m² in the STOP-Bang. This could be a factor for the difference in cutoffs obtained and resulting difference in predictive power.

For ESS, the cutoffs obtained in our study for predicting any OSAS (AHI ≥5) and moderate-to-severe OSAS (AHI ≥15) were 9 and 11, respectively. This is more or less similar to prior studies. The standard cutoff used in most studies for high risk of ESS and hence OSAS is ≥10 as was obtained in our study for predicting AHI ≥15. In our study, the sensitivity and specificity of ESS for predicting AHI ≥5 were 81.48 and 44.26%, respectively, and for AHI ≥15, 77.25 and 45.90%, respectively. Previous studies have reported a sensitivity range of 27–72% and a specificity range of 50–77% for predicting AHI ≥5 at cutoff of >10. Some studies have had higher sensitivity and lower specificity when compared with previous studies. This may be due to the use of Hindi language ESS in our study. Hindi language ESS is also reliable, validated, and comparable with the original scale, but it uses certain modifications to avoid false low scores in Indian patients who often do not drive four wheelers.

For PSAP, the sensitivity and specificity in our study for predicting AHI ≥5 were 73.54 and 65.57%, respectively, and for AHI ≥15, 57.14 and 78.69%, respectively. In the original study where PSAP was developed and validated, the score of ≥4 had specificity of 77.3% and sensitivity of 66.7%. With the increase in threshold level to ≥6, the specificity increased to 91% at the expense of sensitivity (23.9%). This is comparable with our study. Only limited studies have subsequently been performed on PSAP and mostly in surgical patients only.

On comparing the three questionnaires, PSAP was found to be the best predictor for OSAS in view of higher AUC on ROC, better sensitivity, and specificity. Among the other two, ESS had higher sensitivity, whereas STOP-Bang had better specificity. Earlier studies have been done comparing different questionnaires but only one using PSAP. Cowan et al. in their study concluded that STOP-Bang had superior predictive performance to STOP and Berlin questionnaire. Another study comparing four sleep questionnaires, i.e., Berlin, STOP, STOP-Bang questionnaires, and ESS, on 234 screened patients found that the STOP-Bang, Berlin, and STOP questionnaires had the highest sensitivity to predict OSAS (95.75, 95.07, and 91.67%, respectively) but with a very low specificity (26.32, 25, and 25%, respectively), while the ESS had the highest specificity of 75% to predict OSAS but with the lowest sensitivity of 72.55%. Silva et al. evaluated four questionnaires: four-variable screening tool, STOP, STOP-Bang, and ESS. They found that the STOP-Bang questionnaire had higher sensitivity to predict...
moderate-to-severe and severe sleep-disordered breathing (87 and 70.4%, respectively), while the four-variable screening tool had higher specificity to predict moderate-to-severe and severe sleep-disordered breathing (93.2% for both). Saxena et al. have used the same three questionnaires as our study and found PSAP a better predictor similar to our results. 

With the hope to achieve improved performance, we studied all possible combinations using the three questionnaires and assessed their predictive parameters. Among the combination of two questionnaires, STOP-Bang + ESS appears to be the most effective dual combination with higher specificity without much compromise in sensitivity. The combination of ESS + PSAP had a higher NPV of 60%. When all the three questionnaires were combined in parallel (i.e., if any of them is positive, it is considered positive, whereas a negative result in all the three is considered negative), the sensitivity increased up to 95.76% for predicting any OSAS at the cost of very low specificity. The NPV for predicting any OSAS was high at 65.22% which means that if all three questionnaires are negative, there is 65.22% possibility of OSAS being ruled out. To our knowledge, there are very few studies in literature using the combination of questionnaires with only one from India. Earlier Ulasli et al. advocated the use of ESS and Berlin questionnaires together as they had a slightly better performance (specificity of ESS and Berlin questionnaires together was 72%). Pataka et al. attempted to find the best combination out of five different questionnaires—STOP, STOP-Bang, Berlin questionnaires, ESS, and four-variable screening tool—but no combination improved their predictive values. Saxena et al. have used the same combination as our study (STOP-Bang + ESS + PSAP) in 69 patients from India and have found a very high NPV of 90% when compared with only 65% in our study. The difference might be due to the difference in sample size as well as difference in the cutoffs obtained for the questionnaires in both the studies. Our study having a greater sample size of 250 would be more validated for the results. As per our results, the combination of STOP-Bang + ESS + PSAP is a very good screening test for OSAS though it cannot be used for ruling out OSAS with much confidence as reported earlier.

**Conclusion**

Our study supports the current recommendation by AASM guidelines that prediction questionnaires should not be used for the diagnosis of OSAS, but they can be used for screening of OSAS for Indian population too. The combination of questionnaires, especially STOP-Bang, ESS, and PSAP, improves the sensitivity of detection up to 95%. And when all of them are negative, OSAS is ruled out with around 65% confidence. So, using this combination can help us to identify high-risk patients and prioritize them for PSG
so that they can get early treatment. To reduce the current huge burden on the sleep laboratories in the resource-limited settings, the combination of STOP-Bang, ESS, and PSAP questionnaires is a suitable option.

REFERENCES