Research Article

Evaluating Screening Tools for Moderate to Severe Obstructive Sleep Apnea: Early Detection to Avoid Complication

Meghana M[1], Srishankar Bairy[2], Eldhos Jacob[2], Carishma S[2], Ajither PA[2]

[1]Dept. Community Medicine, Srinivasa Medical College, Mangaluru, Karnataka.
[2]Dept. of Respiratory Medicine, Father Muller Medical College Hospital, Kankanady, Mangaluru, Karnataka.

(Received: 30-10-2025 Revised: 03-04-2026 Accepted: 23-04-2026)

Corresponding Author: ***Srishankar Bairy*** Email: ***srishankarabairy2012gmail.com***

ABSTRACT

Background: Obstructive sleep apnea (OSA) is associated with a range of cardiovascular, metabolic, and neurocognitive complications. While polysomnography (PSG) remains the gold standard for diagnosing OSA, its limited accessibility necessitates the use of reliable screening tools, particularly in resource-limited settings.

Objective: This study aimed to evaluate and compare the diagnostic performance of commonly used questionnaires, Epworth Sleepiness Scale (ESS), STOP, STOP-BANG, and Berlin Questionnaire in detecting moderate to severe OSA (Apnea-Hypopnea Index >15).

Methods: In this prospective observational study, 54 adult patients with suspected OSA underwent overnight Level 2 PSG. Participants were given ESS, STOP, STOP-BANG, and Berlin questionnaires prior to the sleep study. Diagnostic metrics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC) were calculated. Agreement with PSG was assessed using Kappa statistics and McNemar's test.

Results: The prevalence of moderate to severe OSA was 68.5% (n= 37) . STOP-BANG and Berlin questionnaires demonstrated the highest sensitivity (97.3%), while ESS showed the highest specificity (64.7%) and PPV (75%). However, all tools exhibited low specificity (<12%) except ESS. Kappa values indicated poor agreement with PSG across all tools ($\kappa < 0.12$), and McNemar's test showed significant discordance due to high false-positive rates in STOP, STOP-BANG, and Berlin tools.

Conclusion: ESS offers better specificity but lower sensitivity. In resource-limited environments, these tools may be useful for initial triage, but confirmatory PSG remains essential for accurate diagnosis and management. A combined or stepwise approach may optimize diagnostic accuracy.

Keywords: Obstructive sleep apnea, Epworth Sleepiness Scale, Berlin Questionnaire, STOP BANG

1. INTRODUCTION

Multiple upper airway collapses during sleep cause sleep fragmentation and daytime drowsiness in obstructive sleep apnea (OSA). It causes several clinical symptoms [1]. It results in intermittent hypoxia, and OSA is associated with multiple systemic complications such as systemic hypertension, coronary artery disease, heart failure, arrhythmias, stroke, pulmonary hypertension, insulin resistance, type 2 diabetes mellitus, cognitive dysfunction and increased risk of motor vehicle accidents [1]. Early detection and treatment reduces morbidity and improves quality of life [2].

"Polysomnography (PSG)" "is gold standard for OSA diagnosis, however its high cost and restricted availability in health care settings have limited its use [3]. This leads to the delayed diagnosis and management of many individuals suspected of having" OSA. For overcoming, a variety of screening questionnaires have been developed to stratify the risk and refer to sleep lab at the earliest to avoid complications due to OSA.

Commonly used screening tools are "Berlin Questionnaire(BQ) , STOP-Bang Questionnaire (SBQ) , and Epworth Sleepiness Scale (ESS)" [3, 4, 5]. Excess daytime sleepiness is an important symptom in OSA. Patient first consults a doctor with these symptoms. Hence, ESS has been employed as a screening tool for OSA, even though it is not originally designed for OSA screening.

Mild OSA is generally managed conservatively, with lifestyle modifications such as weight reduction, positional therapy, and addressing nasal obstruction [6]. In contrast, moderate to severe OSA often requires active intervention, including "continuous positive airway pressure (CPAP) therapy, mandibular advancement devices, or surgical management"[6]. Moreover, individuals in these categories are at significantly higher risk for cardiovascular events, metabolic disturbances, and neurocognitive decline [7, 8]. Early detection by using screening tool and treatment can significantly reduce long-term complications and improve outcomes. Hence, in the present study, moderate to severe OSA (described as Apnea-Hypopnea Index >15) was chosen as the diagnostic cut-off. This decision was based on the clinical approach to OSA management. Therefore, by this study we are assessing whether these screening tools can effectively detect moderate to severe OSA which in turn help to determine their utility in prioritizing patients for diagnostic testing and timely initiation of appropriate therapy in resource limited health care infrastructure and also to avoid the cost of unnecessary PSGs.

2. METHODOLOGY

Study design and setting

This prospective observational study was conducted over a period of 12 months from January 2024 to January 2025 at our a tertiary care teaching hospital. The study was designed to evaluate the diagnostic performance of commonly used screening questionnaires in detecting moderate to severe obstructive sleep apnea (OSA) among adults undergoing polysomnography.

Study population

Adult patients aged ≥18 years presenting to the respiratory medicine department with symptoms suggestive of obstructive sleep apnea and advised Level-2 polysomnography (PSG) were considered for inclusion. Clinical suspicion of OSA was based on history and physical examination.

OSA assessment included

- Detailed clinical history.
- Physical examination including body mass index (BMI) and neck circumference.
- Screening questionnaires.
- Overnight Level-2 polysomnography.

Inclusion criteria

Patients meeting any of the following criteria were included:

- History of habitual snoring.
- Excessive daytime sleepiness.
- Witnessed apneas during sleep.
- Obesity.
- Clinical suspicion of obstructive sleep apnea.

Exclusion criteria

Patients were excluded if they had

- Known neuropsychiatric illness.
- Incomplete screening questionnaires.
- Inadequate polysomnography data.
- Refusal to provide informed consent.

Sample size justification

The sample size was determined based on the number of eligible patients undergoing polysomnography during the study period. Due to logistical constraints and limited availability of PSG, a total of 54 patients fulfilling the inclusion criteria were enrolled.

Patient recruitment

A total of 72 patients were assessed for eligibility during the study period. Of these, 18 patients were excluded due to incomplete questionnaires (n=8), inadequate PSG data (n=5), and refusal to participate (n=5). Finally, 54 patients were included in the study analysis.

Screening tools

All participants completed the following validated screening questionnaires prior to undergoing polysomnography:

Epworth Sleepiness Scale (ESS) [5]

ESS is an 8-item self-administered questionnaire used to assess daytime sleepiness. Each item is scored from 0 to 3, with a maximum score of 24. A score greater than 11 was considered indicative of excessive daytime sleepiness.

Berlin Questionnaire [4]

The Berlin questionnaire assesses the risk of OSA based on three domains:

- Snoring severity.
- Daytime fatigue or sleepiness.
- Presence of hypertension or obesity.

Patients were categorized as high risk if they had positive scores in two or more categories.

STOP Questionnaire [3]

The STOP questionnaire consists of four yes/no questions assessing:

- Snoring.
- Tiredness during daytime.
- Observed apnea.
- High blood pressure.

A score of ≥2 was considered high risk for OSA.

STOP-BANG Questionnaire [3]

The STOP-BANG questionnaire is an 8-item screening tool assessing:

- Snoring.
- Tiredness.
- Observed apnea.
- Blood pressure.
- Body mass index.
- Age.
- Neck circumference.
- Gender.

A score of ≥3 was considered indicative of high risk for OSA.

Polysomnography

PSG was used as the gold standard for evaluating diagnostic accuracy. All participants underwent overnight Level-2 polysomnography using standard monitoring parameters including

- Electroencephalogram (EEG).
- Electrooculogram (EOG).
- Electromyogram (EMG).
- Electrocardiogram (ECG).
- Airflow monitoring.
- Respiratory effort belts.
- Pulse oximetry.

OSA severity was classified based on Apnea-Hypopnea Index (AHI)

- Mild OSA: AHI 5-15 events/hour.
- Moderate OSA: AHI 15-30 events/hour.
- Severe OSA: AHI >30 events/hour.

Moderate to severe OSA was defined as AHI >15 events/hour for the purpose of diagnostic performance analysis.

3. STATISTICAL ANALYSIS

Data were analyzed using appropriate statistical software. Demographic characteristics and baseline variables were summarized using descriptive statistics.

The diagnostic performance of each screening questionnaire in detecting moderate to severe OSA (AHI >15) was evaluated by calculating:

- Sensitivity.
- Specificity.
- Positive Predictive Value (PPV).
- Negative Predictive Value (NPV).
- Overall diagnostic accuracy.
- Area under the Receiver Operating Characteristic (ROC) curve (AUC).

Cross-tabulation analysis was performed to compare questionnaire results with PSG findings. Agreement between screening tools and PSG diagnosis was assessed using Cohen's kappa statistics. McNemar's test was applied to evaluate statistical significance between paired proportions.

A p-value <0.05 was considered statistically significant.

Ethical approval

The study was approved by the Institutional Ethics Committee of Father Muller Medical College Hospital, Mangaluru. Written informed consent was obtained from all participants prior to enrolment.

4. RESULTS

This study included 54 patients. Mean age had been 52.3yrs (SD=13.4), ranging from 18 to 73. The majority had been male (59.3%). Most participants were 50–70y/o, typical population at obstructive sleep apnea risk. In this cohort, 68.5% had moderate-to-severe OSA (AHI>15).

Table 1: Diagnostic performance metrics (Sensitivity, Specificity, PPV, NPV, Accuracy, AUC) for each questionnaire.

Questionnaire	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
ESS (>11)	0.4865	0.6471	0.75	0.3667	0.537	0.567
STOP (>2)	1	0.0588	0.6981	1	0.704	0.529
STOP-BANG (>3)	0.973	0.1176	0.7059	0.6667	0.704	0.545
Berlin (Final ≥2 Category)	0.973	0.0588	0.6923	0.5	0.685	0.516

Table 2: Cross-tabulations comparing each screening tool against AHI >15 (moderate to severe OSA).

Berlin Final ≥2	AHI >15 (Yes) (moderate to severe OSA)	AHI <=15 (No) (Mild OSA)
Yes	36	16
No	1	1

Table 3: Cross-tabulations comparing each screening tool against AHI >15.

ESS >11	AHI>15 (Yes) (moderate to severe OSA)	AHI<=15 (No) (Mild OSA"
Yes	18	6
No	19	11

Confidence intervals (CI) were given at 95%. Kappa statistic evaluates agreement beyond chance between the screening tools and a reference standard. McNemar's test assesses the statistical significance of difference in paired proportions.

Table 4: Cross-tabulations comparing each screening tool against AHI >15.

STOP >2	AHI>15(Yes)(moderate to severe OSA)	AHI<=15 (No)(Mild OSA)
Yes	37	16
No	0	1

P-value < 0.05 has been considered statistically significant. AUC reflects overall ability of test in discriminating between positive and negative cases.

Table 5: Cross tabulation of accuracy, sensitivity, specificity, PPV, NPV, and AUC for each screening tool (ESS, STOP, SBQ, and BQ).

	ESS	**Stop**	**Stop Bang**	**Berlin**
Accuracy (95% CI)	0.537 (0.3961, 0.6738)	0.7037 (0.5639, 0.8202)	0.7037 (0.5639, 0.8202)	0.6852 (0.5445, 0.8048)
Kappa Statistic (95% CI)	0.1107 (-0.1221, 0.3435)	0.0789 (-0.0685, 0.2263)	0.1166 (-0.0878, 0.3210)	0.0418 (-0.1197, 0.2032)
Kappa Statistic p-value	0.359	0.136	0.177	0.566
McNemar's Test p-value	0.0164	0.0002 = <0.001	0.0012	0.0006 = <0.001
Sensitivity	0.4865	1.0000	0.9730	0.9730
Specificity	0.6471	0.0588	0.1176	0.0588
PPV	0.7500	0.6981	0.7059	0.6923
NPV	0.3667	1.0000	0.6667	0.5000
AUC (95% CI)	0.5668 (0.424, 0.7095)	0.5294 (0.4718, 0.5871)	0.5453 (0.462, 0.6286)	0.5159 (0.4525, 0.5793)

Cross-tabulations revealed that STOP, SBQ, and BQ identified nearly all moderate to severe OSA cases (AHI >15), but also misclassified substantial number of mild cases as high risk, reflecting their low specificities.

Kappa statistics indicated poor agreement with the gold standard (AHI >15) across all tools: ESS (κ = 0.11), STOP (κ=0.08), STOP-BANG (κ=0.12), and Berlin (κ = 0.04), none of which attained "statistical significance (p>0.05).

McNemar's test showed statistically significant discordance for all screening tools when compared to AHI >15, especially for STOP (p<0.001), STOP-BANG (p=0.0012), and Berlin (p<0.001), indicating" a significant number of false positives.

Receiver Operating Characteristic (ROC) curves of the ESS, STOP, STOP-BANG, and Berlin questionnaires for predicting moderate to severe OSA. Sensitivity is plotted against 1-specificity. Among the tools, STOP-BANG demonstrated better discriminative performance, while ESS showed relatively lower diagnostic accuracy.

Data analysed using R Studio version 4.4.2.

5. DISCUSSION

Questionnaires encompassing SBQ and BQ demonstrated excellent sensitivity (97.3%) in detecting moderate to severe OSA. Results align with those reported by Chung *et al.*, who observed similarly high sensitivity (93–100%) of STOP-BANG tool in preoperative patients [3].

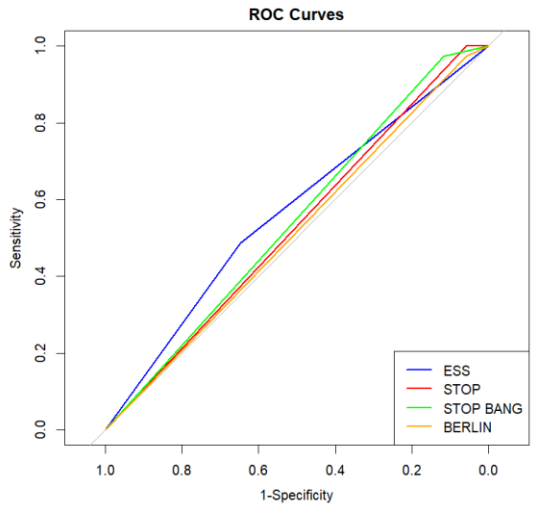

Fig. 1 Receiver Operating Characteristic (ROC) curves

However, consistent with our results, they also reported low specificity (11.8% and 5.9%, respectively), which can lead to overestimation and potential over-referral in low-risk populations [9].

The STOP questionnaire exhibited perfect sensitivity and negative predictive value (both 100%), indicating its strong utility in ruling out disease. However, it showed extremely poor specificity (5.9%), limiting its effectiveness in confirming cases. Silva et al. observed high sensitivity however low specificity for STOP questionnaire in primary care, similar to our results [10].

ESS is primarily employed for assessing excessive daytime sleepiness, demonstrated highest specificity (64.7%) and PPV (75%) among tools evaluated. However, its sensitivity was relatively low (48.7%), supporting previous findings that ESS may not reliably identify OSA severity, particularly among asymptomatic or minimally symptomatic individuals [4, 11].

Interestingly, although BQ, SBQ, and STOP showed comparable overall accuracy (approximately 68–70%), Kappa statistics revealed poor agreement with polysomnographic diagnoses. Moreover, statistically significant McNemar test results across all tools highlight considerable misclassification, largely due to low specificity and high false-positive rates.

These results raise essential considerations. While tools like STOP-BANG and Berlin are useful for identifying individuals at high risk-especially given their high sensitivity-they should not be used as standalone diagnostic tools due to their limited specificity. In resource-constrained settings, over-reliance on these tools without confirmatory testing such as polysomnography (PSG) could result in unnecessary referrals and increased patient anxiety.

However, these screening tools retain clinical value in high-risk or high-prevalence populations, where maximizing sensitivity is prioritized. In such settings, even tools with modest specificity can aid in efficient triage, thereby facilitating access to diagnostic resources like PSG. For example, Abrishami et al. highlighted practicality of STOP-BANG as perioperative screening tool due to its simplicity and strong predictive power for severe OSA [12].

In this prospective investigation, diagnostic performance of four frequently employed screening tools STOP-BANG, ESS, STOP, and Berlin in identifying moderate to severe OSA—described as an AHI>15 has been assessed and compared. Our results demonstrate that while these tools generally offer high sensitivity, they fall short in specificity and agreement with the PSG-defined gold standard. ESS demonstrated relatively better specificity compared to other tools. No single tool achieved both high sensitivity and specificity. STOP and STOP-BANG showed better overall accuracy, No single tool demonstrated ideal diagnostic accuracy.

6. STRENGTHS AND LIMITATIONS

The prospective design minimized recall bias and ensured systematic data collection. We directly compared multiple widely used OSA screening questionnaires against polysomnography, the gold standard diagnostic test, allowing objective evaluation of their diagnostic performance for moderate to severe OSA. Standardized PSG scoring and uniform application of screening tools further strengthened internal validity.

However, some limitations should be considered while interpreting the results. The relatively small sample size may limit statistical power and generalizability. As this was a single-centre study conducted in a tertiary care referral setting, selection bias cannot be excluded. The high prevalence of moderate to severe OSA in our

cohort may have inflated positive predictive values. Additionally, the absence of an external validation cohort limits extrapolation of our findings to other populations and clinical settings. Larger multicentric studies with external validation are needed to confirm these findings.

7. CONCLUSION

Among screening tools, STOP-BANG and Berlin showed high sensitivity while ESS showed better specificity. However no tool demonstrated balanced diagnostic performance. These tools are useful for screening and triage but cannot replace PSG. A combined approach using clinical evaluation and questionnaires may improve patient selection for PSG especially in resource-limited settings.

ACKNOWLEDGEMENT: The authors would like to acknowledge Ms. Krupa for her assistance in statistical data interpretation.

CONFLICTS OF INTEREST: All authors declare that they have no conflicts of interest.

FUNDING INFORMATION: No funding was received for this research work.

REFERENCES

1. Chen R, Zhu J, Yang Y, Liao W, Ye W, Du L, Chen M, Zhang Y, Yao W, Zheng Z. Evaluation of five questionnaires for obstructive sleep apnea screening in the elderly. *Sci Rep*. 2025 Jan 11;15(1):1689. doi:10.1038/s41598-025-86041-8. PMID: 39799222; PMCID: PMC11724924.
2. Chen L, Pivetta B, Nagappa M, Saripella A, Waseem R, Englesakis M, *et al.*, Validation of the STOP-Bang questionnaire for screening of obstructive sleep apnea in the general population and commercial drivers: a systematic review and meta-analysis. *Sleep Breath*. 2021. doi:10.1007/s11325-021-02299-y
3. 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(5):812-21. doi:10.1097/ALN.0b013e31816d83e4
4. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. *Am J Respir Crit Care Med*. 1999;159(2):494-500. doi:10.1164/ajrccm.159.2.9809054
5. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. 1991;14(6):540-5.
6. Goyal M, Johnson J. Obstructive sleep apnea diagnosis and management. *Mo Med*. 2017 Mar-Apr;114(2):120-4. PMID: 30228558; PMCID: PMC6140019.
7. Borsini E, Blanco M, Bosio M, Schrappe M, Ernst G, Nosetto D, *et al.*, Prevalence of sleep apnea and cardiovascular risk factors in patients with hypertension in a day hospital model. *Clin Exp Hypertens*. 2018;40(3):231-7. doi:10.1080/10641963.2017.1356841. Epub 2017 Sep 5. PMID: 28872361.
8. Ferini-Strambi L, Lombardi GE, Marelli S, Galbiati A. Neurological deficits in obstructive sleep apnea. *Curr Treat Options Neurol*. 2017 Apr;19(4):16. doi:10.1007/s11940-017-0451-8. PMID: 28374233.
9. Sharma SK, Vasudev C, Sinha S, Banga A, Pandey RM. Validation of the modified Berlin questionnaire to identify obstructive sleep apnea in a high-risk population. *Sleep Breath*. 2016;20(2):785-93. doi:10.1007/s11325-015-1266-2
10. Silva GE, Vana KD, Goodwin JL, 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(5):467-72. doi:10.5664/JCSM.1308
11. Chung F, Abdullah HR, Liao P. STOP-Bang Questionnaire: a practical

approach to screen for obstructive sleep apnea. *Chest*. 2016;149(3):631-8. doi:10.1378/chest.15-0903

12. Abrishami A, Khajehdehi A, Chung F. A systematic review of screening questionnaires for obstructive sleep apnea. *Can J Anaesth*. 2010;57(5):423-38. doi:10.1007/s12630-010-9280-x