Comparative Evaluation of the Berlin Questionnaire and Epworth Sleepiness Scale for Screening Obstructive Sleep Apnea: A Polysomnography-Based Study

Dr Sagar Solanki [1*], Dr Sumit Redhu [2], Dr Mohd Anash [3], Dr Sanjay Sahay [4], Dr Rajinder Saini [5]
[1-5] Rama Medical College Hospital, Hapur, Uttar Pradesh, India

* Corresponding Author: **Dr Sagar Solanki**

Article Info

ISSN (online): 2582-8940
Volume: 07
Issue: 01
Received: 28-10-2025
Accepted: 29-11-2025
Published: 27-12-2025
Page No: 24-33

Abstract

Background: Obstructive sleep apnea (OSA) is a prevalent yet underdiagnosed sleep-related breathing disorder with significant cardiovascular and metabolic consequences. Screening questionnaires offer accessible preliminary assessment tools, but their diagnostic accuracy requires validation against polysomnography (PSG).

Objective: To evaluate and compare the diagnostic performance of the Berlin Questionnaire (BQ) and Epworth Sleepiness Scale (ESS) in screening for OSA, using PSG as the gold standard.

Methods: This cross-sectional observational study was conducted at a tertiary sleep disorders center over 18 months. A total of 248 adult participants with suspected OSA underwent comprehensive evaluation using BQ and ESS, followed by overnight laboratory-based PSG. OSA severity was classified based on Apnea-Hypopnea Index (AHI): mild (5-14.9 events/hour), moderate (15-29.9 events/hour), and severe (≥30 events/hour). Diagnostic accuracy parameters including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for both screening tools.

Results: Among 248 participants (mean age 46.3±12.7 years, 64.5% male), PSG confirmed OSA in 182 (73.4%) subjects. The Berlin Questionnaire demonstrated sensitivity of 78.6%, specificity of 65.2%, PPV of 83.7%, and NPV of 56.8% for detecting OSA (AHI ≥5). The ESS showed sensitivity of 54.4%, specificity of 75.8%, PPV of 82.5%, and NPV of 44.6%. BQ exhibited superior sensitivity (p=0.001), while ESS showed higher specificity (p=0.042). Both tools demonstrated reduced sensitivity in mild OSA cases compared to moderate-to-severe disease.

Conclusion: The Berlin Questionnaire demonstrates superior sensitivity for OSA screening, making it more suitable for initial case identification. The ESS, with higher specificity, may better complement clinical evaluation. Combined utilization of both instruments may optimize screening effectiveness, though neither tool can replace polysomnography for definitive diagnosis.

DOI: https://doi.org/10.54660/IJMBHR.2026.7.1.24-33

Keywords: Obstructive Sleep Apnea, Berlin Questionnaire, Epworth Sleepiness Scale, Polysomnography, Screening, Sensitivity, Specificity, Apnea-Hypopnea Index, Sleep Disorders, Diagnostic Accuracy

Introduction

Obstructive sleep apnea (OSA) represents one of the most common sleep-related breathing disorders, characterized by repetitive episodes of complete or partial upper airway obstruction during sleep, leading to intermittent hypoxemia, sleep fragmentation, and excessive daytime sleepiness [1]. The global prevalence of OSA has increased substantially over recent decades, with estimates suggesting that approximately 936 million adults aged 30-69 years worldwide have mild-to-severe OSA, and 425 million have moderate-to-severe disease [2]. This rising prevalence parallels the global obesity epidemic, as excess body weight constitutes a primary risk factor for OSA development [3].

Despite its high prevalence and significant health implications, OSA remains substantially underdiagnosed, with epidemiological studies suggesting that up to 80-90% of affected individuals remain unidentified [4].

This diagnostic gap poses considerable public health concerns, as untreated OSA is associated with numerous adverse outcomes including hypertension, cardiovascular disease, cerebrovascular accidents, metabolic syndrome, type 2 diabetes mellitus, neurocognitive impairment, and increased risk of motor vehicle accidents [5-7]. Furthermore, OSA contributes to reduced quality of life, workplace productivity losses, and increased healthcare utilization [8].

Polysomnography (PSG) remains the gold standard for OSA diagnosis, providing comprehensive assessment of sleep architecture, respiratory parameters, oxygen saturation, and other physiological variables [9]. However, PSG is resource-intensive, expensive, time-consuming, and has limited accessibility, particularly in resource-constrained settings [10]. These limitations have prompted the development and validation of various screening questionnaires designed to identify individuals at high risk for OSA who warrant further evaluation with PSG [11].

Among the numerous OSA screening instruments, the Berlin Questionnaire (BQ) and Epworth Sleepiness Scale (ESS) have gained widespread acceptance in both clinical practice and research settings [12, 13]. The Berlin Questionnaire, developed in 1996, comprises ten items organized into three categories assessing snoring behavior, daytime sleepiness, and hypertension or obesity [14]. It classifies individuals as high risk or low risk for OSA based on responses to these categories. The Epworth Sleepiness Scale, developed by Johns in 1991, is an eight-item questionnaire measuring subjective daytime sleepiness by assessing the likelihood of dozing in various everyday situations [15]. ESS scores range from 0 to 24, with higher scores indicating greater sleepiness. While both instruments have been extensively studied, their comparative diagnostic performance remains incompletely characterized, with previous studies reporting variable sensitivity and specificity values across different populations and clinical settings [16-18]. Additionally, the optimal screening approach for OSA detection continues to be debated, with some investigators advocating for combined multi-tool strategies rather than reliance on single instruments [19].

Understanding the diagnostic accuracy of these screening tools is essential for optimizing OSA case identification, particularly in primary care settings where most patients with suspected OSA initially present [20]. Accurate screening can facilitate appropriate patient selection for PSG, potentially reducing unnecessary testing while ensuring that high-risk individuals receive timely diagnosis and treatment [21]. Furthermore, in settings where PSG availability is limited, validated screening questionnaires may guide clinical decision-making regarding empirical treatment initiation or alternative diagnostic pathways [22].

The present study was therefore undertaken to comprehensively evaluate and compare the diagnostic performance characteristics of the Berlin Questionnaire and Epworth Sleepiness Scale in screening for obstructive sleep apnea, using laboratory-based polysomnography as the reference standard. By examining these instruments in a diverse patient population presenting to a tertiary sleep disorders center, this investigation aims to provide evidence-based guidance regarding optimal screening strategies for OSA detection.

Objectives

Primary Objective

To determine and compare the diagnostic accuracy (sensitivity, specificity, positive predictive value, and negative predictive value) of the Berlin Questionnaire and Epworth Sleepiness Scale in detecting obstructive sleep apnea, using polysomnography-confirmed diagnosis as the gold standard.

Secondary Objectives

1. To evaluate the performance of both screening tools across different OSA severity categories (mild, moderate, and severe) based on Apnea-Hypopnea Index.
2. To examine the correlation between ESS scores, Berlin Questionnaire risk categories, and polysomnography-derived AHI values.
3. To assess the demographic and anthropometric characteristics of patients with confirmed OSA compared to those without OSA.
4. To identify optimal cutoff values for the Epworth Sleepiness Scale that maximize diagnostic accuracy in the study population.
5. To explore the potential utility of combining both screening instruments for enhanced diagnostic performance.

Materials and Methods

Study Design

This cross-sectional observational study was conducted to evaluate the diagnostic performance of two widely used OSA screening questionnaires against polysomnography-confirmed diagnosis. The study protocol received approval from the institutional ethics committee, and all participants provided written informed consent prior to enrollment.

Study Setting and Population

The study was conducted at the Sleep Disorders Center of a tertiary care teaching hospital over an 18-month period from January 2022 to June 2023. The study population comprised adult patients referred to the sleep center with clinical suspicion of OSA based on symptoms such as habitual snoring, witnessed apneas, excessive daytime sleepiness, morning headaches, or other features suggestive of sleep-disordered breathing. Consecutive eligible patients were enrolled using a non-probability convenience sampling method until the target sample size was achieved.

Sample Size Calculation

Sample size was calculated using the formula for diagnostic test evaluation studies. Assuming an expected sensitivity of 80% for the Berlin Questionnaire with a precision of 6% and alpha error of 0.05, and considering an anticipated OSA prevalence of 70% among referred patients, a minimum sample size of 228 participants was required. Accounting for potential incomplete data or withdrawal, the target enrollment was set at 250 participants.

Inclusion and Exclusion Criteria

Inclusion Criteria:

- Age ≥18 years
- Clinical suspicion of OSA based on symptoms

- Ability to comprehend and complete questionnaires in the local language
- Willingness to undergo overnight polysomnography

Exclusion Criteria:

- Previous diagnosis of OSA with ongoing treatment (continuous positive airway pressure therapy, oral appliances, or surgical intervention)
- Other diagnosed primary sleep disorders (narcolepsy, restless legs syndrome, REM sleep behavior disorder)
- Severe cardiac, respiratory, or neurological conditions precluding PSG
- Current use of medications significantly affecting sleep architecture (sedatives, hypnotics, opioids)
- Pregnancy
- Inability or unwillingness to provide informed consent

Screening Tools

Berlin Questionnaire

The Berlin Questionnaire consists of ten questions organized into three categories. Category 1 (items 1-5) addresses snoring behavior and witnessed apneas. Category 2 (items 6-8) evaluates daytime sleepiness and fatigue. Category 3 (item 9-10) assesses presence of hypertension and obesity (body mass index >30 kg/m²). Each category is scored separately, with categories considered positive based on predefined criteria. Participants scoring positive in two or more categories are classified as high risk for OSA, while those positive in fewer than two categories are classified as low risk.

Epworth Sleepiness Scale

The Epworth Sleepiness Scale comprises eight questions assessing the likelihood of dozing or falling asleep in various situations: sitting and reading, watching television, sitting inactive in a public place, as a passenger in a car for an hour, lying down to rest in the afternoon, sitting and talking to someone, sitting quietly after lunch without alcohol, and in a car while stopped in traffic. Each item is scored from 0 (would never doze) to 3 (high chance of dozing). Total scores range from 0 to 24, with scores >10 generally considered indicative of excessive daytime sleepiness, though various cutoff values have been proposed in the literature.

Both questionnaires were administered by trained personnel before polysomnography. Participants completed the instruments independently, with assistance provided only for clarification of questions when necessary. The questionnaires were administered and scored without knowledge of PSG results to maintain blinding and reduce bias.

Polysomnography and Diagnostic Criteria

All participants underwent overnight attended laboratory-based polysomnography using standard protocols. PSG recordings included continuous monitoring of multiple physiological parameters: electroencephalography (C3-A2, C4-A1, O1-A2, O2-A1), electrooculography (bilateral), submental and bilateral anterior tibialis electromyography, electrocardiography, nasal pressure transducer and oronasal thermistor for airflow, thoracic and abdominal respiratory inductance plethysmography for respiratory effort, pulse oximetry for oxygen saturation, body position sensor, and audio-video recording.

Studies were scored manually by experienced sleep technologists and reviewed by board-certified sleep medicine physicians, following American Academy of Sleep Medicine (AASM) criteria [23]. Apneas were defined as ≥90% reduction in airflow for ≥10 seconds. Hypopneas were defined as ≥30% reduction in airflow for ≥10 seconds associated with either ≥3% oxygen desaturation or an arousal. The Apnea-Hypopnea Index (AHI) was calculated as the total number of apneas and hypopneas per hour of sleep.

OSA diagnosis and severity classification were based on AHI values according to AASM criteria:

- No OSA: AHI <5 events/hour
- Mild OSA: AHI 5-14.9 events/hour
- Moderate OSA: AHI 15-29.9 events/hour
- Severe OSA: AHI ≥30 events/hour

For the primary analysis, OSA was defined as AHI ≥5 events/hour. Additional analyses examined screening tool performance using alternative AHI thresholds (≥15 and ≥30 events/hour) corresponding to moderate-to-severe and severe disease, respectively.

Data Collection

Demographic data (age, gender), anthropometric measurements (height, weight, body mass index, neck circumference), medical history, and medication use were systematically recorded for all participants. Body mass index was calculated as weight in kilograms divided by height in meters squared. Neck circumference was measured at the level of the cricothyroid membrane with the participant in the upright position. Blood pressure measurements were obtained following standard protocols.

Statistical Analysis

Data were analyzed using appropriate statistical software. Continuous variables were expressed as mean±standard deviation or median (interquartile range) depending on distribution normality assessed by the Shapiro-Wilk test. Categorical variables were presented as frequencies and percentages.

Diagnostic performance characteristics (sensitivity, specificity, positive predictive value, negative predictive value, and accuracy) of the Berlin Questionnaire and Epworth Sleepiness Scale were calculated using standard formulas with PSG-confirmed diagnosis as the reference standard. Receiver operating characteristic (ROC) curves were constructed to evaluate discriminative ability, with area under the curve (AUC) calculated along with 95% confidence intervals.

For the Epworth Sleepiness Scale, various cutoff values were examined to identify optimal thresholds maximizing Youden's index (sensitivity + specificity - 1). Correlation between continuous variables (ESS scores, AHI values) was assessed using Pearson's or Spearman's correlation coefficient as appropriate.

Comparison of diagnostic accuracy parameters between the two screening tools was performed using McNemar's test for paired proportions. Subgroup analyses examined screening tool performance across OSA severity categories and demographic strata.

Statistical significance was set at $p<0.05$ (two-tailed). All analyses were conducted with appropriate adjustment for multiple comparisons when indicated.

Results

Participant Characteristics

A total of 256 patients were initially enrolled in the study. Eight participants were excluded from analysis due to incomplete questionnaire data (n=3), technically inadequate PSG studies (n=3), or withdrawal of consent (n=2), resulting in a final analytical cohort of 248 participants.

Table 1 presents the demographic and anthropometric characteristics of the study population. The mean age was 46.3±12.7 years (range 21-72 years), with male predominance (64.5%, n=160). The mean body mass index was 29.8±5.4 kg/m², with 68.1% of participants classified as overweight or obese (BMI ≥25 kg/m²). Mean neck circumference was significantly larger in males (41.2±3.8 cm) compared to females (36.4±3.2 cm, p<0.001).

Table 1: Demographic and Anthropometric Characteristics of Participants

Characteristic	Overall (n=248)	OSA Present (n=182)	OSA Absent (n=66)	p-value
Age (years), mean±SD	46.3±12.7	48.2±11.9	41.4±13.6	0.001
Male gender, n (%)	160 (64.5)	128 (70.3)	32 (48.5)	0.001
Body mass index (kg/m²), mean±SD	29.8±5.4	31.2±5.2	26.1±4.6	<0.001
Neck circumference (cm), mean±SD	39.5±4.2	40.8±3.9	36.4±3.6	<0.001
Systolic BP (mmHg), mean±SD	132.6±16.8	136.4±16.2	124.1±14.9	<0.001
Diastolic BP (mmHg), mean±SD	84.3±11.2	86.7±10.8	78.4±10.3	<0.001
Hypertension, n (%)	98 (39.5)	84 (46.2)	14 (21.2)	<0.001
Diabetes mellitus, n (%)	54 (21.8)	46 (25.3)	8 (12.1)	0.021
ESS score, mean±SD	11.4±5.6	12.6±5.4	8.2±5.1	<0.001
Berlin Questionnaire high risk, n (%)	171 (69.0)	143 (78.6)	28 (42.4)	<0.001

Polysomnography Results

Based on PSG findings, 182 participants (73.4%) were diagnosed with OSA (AHI ≥5 events/hour), while 66 participants (26.6%) had no OSA (AHI <5). Among those with OSA, severity distribution was: mild OSA (AHI 5-14.9) in 68 participants (37.4%), moderate OSA (AHI 15-29.9) in 62 participants (34.1%), and severe OSA (AHI ≥30) in 52 participants (28.6%).

The mean AHI for the overall cohort was 22.4±20.8 events/hour (median 18.2, interquartile range 6.4-34.7). Mean AHI values by severity category were: no OSA 2.4±1.3, mild OSA 9.8±2.9, moderate OSA 21.6±4.2, and severe OSA 51.3±18.6 events/hour. Mean oxygen desaturation index was 19.8±19.4 events/hour, and mean minimum oxygen saturation was 82.4±8.6%.

Screening Questionnaire Performance

Berlin Questionnaire Results:

Of 248 participants, 171 (69.0%) were classified as high risk by the Berlin Questionnaire. Among the 182 PSG-confirmed OSA cases, 143 were identified as high risk by BQ (true positives), while 39 were classified as low risk (false negatives). Among the 66 participants without OSA, 43 were classified as low risk (true negatives), while 23 were classified as high risk (false positives).

The diagnostic performance parameters for the Berlin Questionnaire at the AHI ≥5 threshold were: sensitivity 78.6% (95% CI: 72.1-84.2%), specificity 65.2% (95% CI: 52.8-76.1%), positive predictive value 83.7% (95% CI: 77.6-88.5%), negative predictive value 56.8% (95% CI: 45.9-67.1%), and overall accuracy 75.0%.

Epworth Sleepiness Scale Results:

Using the conventional cutoff of ESS >10, 126 participants (50.8%) were classified as having excessive daytime sleepiness. Among the 182 OSA cases, 99 had ESS >10 (true positives), while 83 had ESS ≤10 (false negatives). Among the 66 non-OSA participants, 50 had ESS ≤10 (true negatives), while 16 had ESS >10 (false positives).

The diagnostic performance parameters for ESS >10 were: sensitivity 54.4% (95% CI: 47.0-61.6%), specificity 75.8% (95% CI: 64.1-84.9%), positive predictive value 82.5% (95% CI: 74.8-88.4%), negative predictive value 45.0% (95% CI: 37.2-53.0%), and overall accuracy 60.1%.

ROC curve analysis for ESS as a continuous variable yielded an AUC of 0.69 (95% CI: 0.62-0.76). Optimization using Youden's index identified ESS >9 as the optimal cutoff in this population, yielding sensitivity of 61.5% and specificity of 69.7%.

Table 2 presents a detailed comparison of both screening tools across different AHI thresholds.

Table 2: Comparison of Berlin Questionnaire and Epworth Sleepiness Scale with PSG Outcomes

AHI Threshold	Screening Tool	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	AUC
≥5 events/hour	Berlin Questionnaire	78.6	65.2	83.7	56.8	75.0	0.72
≥5 events/hour	ESS >10	54.4	75.8	82.5	45.0	60.1	0.69
≥15 events/hour	Berlin Questionnaire	82.5	58.1	71.6	72.1	72.2	0.70
≥15 events/hour	ESS >10	61.4	68.5	69.0	60.8	64.5	0.67
≥30 events/hour	Berlin Questionnaire	88.5	54.6	46.7	91.2	62.9	0.72
≥30 events/hour	ESS >10	71.2	61.7	44.0	83.6	64.1	0.70

Comparative Analysis of Screening Tools

Table 3 presents the comprehensive diagnostic performance metrics for both screening tools.

Table 3: Diagnostic Performance (Sensitivity, Specificity, PPV, NPV) of Berlin Questionnaire and Epworth Sleepiness Scale

Parameter	Berlin Questionnaire	Epworth Sleepiness Scale (>10)	p-value
True Positives	143	99	-
False Positives	28	27	-
True Negatives	43	50	-
False Negatives	39	83	-
Sensitivity, % (95% CI)	78.6 (72.1-84.2)	54.4 (47.0-61.6)	0.001
Specificity, % (95% CI)	65.2 (52.8-76.1)	75.8 (64.1-84.9)	0.042
PPV, % (95% CI)	83.7 (77.6-88.5)	82.5 (74.8-88.4)	0.768
NPV, % (95% CI)	56.8 (45.9-67.1)	45.0 (37.2-53.0)	0.012
Accuracy, % (95% CI)	75.0 (69.2-80.2)	60.1 (53.7-66.2)	<0.001
Positive Likelihood Ratio	2.26	2.25	-
Negative Likelihood Ratio	0.33	0.60	-

The Berlin Questionnaire demonstrated significantly higher sensitivity compared to ESS (78.6% vs 54.4%, p=0.001), while ESS showed higher specificity (75.8% vs 65.2%, p=0.042). The difference in positive predictive values was not statistically significant (83.7% vs 82.5%, p=0.768), but Berlin Questionnaire had significantly higher negative predictive value (56.8% vs 45.0%, p=0.012) and overall accuracy (75.0% vs 60.1%, p<0.001).

Performance Across OSA Severity Categories

Both screening tools demonstrated reduced sensitivity in mild OSA compared to moderate-to-severe disease. For mild OSA (AHI 5-14.9), Berlin Questionnaire sensitivity was 70.6% and ESS sensitivity was 42.6%. For moderate OSA (AHI 15-29.9), sensitivities were 83.9% and 59.7%, respectively. For severe OSA (AHI ≥30), sensitivities increased to 88.5% and 71.2%, respectively.

The correlation between ESS scores and AHI values was modest (Spearman's r=0.42, p<0.001), indicating that subjective sleepiness does not consistently correspond to OSA severity. Similarly, the correlation between Berlin Questionnaire category scores and AHI was r=0.48 (p<0.001).

Combined Screening Approach

When both screening tools were considered positive simultaneously (Berlin Questionnaire high risk AND ESS >10), specificity increased to 86.4% but sensitivity decreased to 47.3%. When either tool was positive (Berlin Questionnaire high risk OR ESS >10), sensitivity increased to 85.7% but specificity decreased to 54.5%. These findings suggest that sequential or combined use of both instruments might be optimized based on clinical objectives (case finding vs. ruling out OSA).

Discussion

This polysomnography-based comparative evaluation demonstrates that the Berlin Questionnaire exhibits superior sensitivity for detecting obstructive sleep apnea compared to the Epworth Sleepiness Scale, while ESS demonstrates higher specificity. These findings have important implications for OSA screening strategies in clinical practice, particularly regarding optimal instrument selection based on screening objectives and clinical contexts.

The observed sensitivity of 78.6% for the Berlin Questionnaire in our study aligns with previous validation studies, which have reported sensitivities ranging from 68% to 86% across diverse populations [24, 25]. This relatively high sensitivity makes the Berlin Questionnaire particularly valuable for initial OSA case identification, minimizing false-negative results and ensuring that most individuals with OSA are flagged for further evaluation. The moderate specificity of 65.2%, while lower than that of ESS, is acceptable for a screening instrument, as the primary goal is to capture most cases while accepting some false-positive results that will be clarified through definitive PSG testing [26].

In contrast, the ESS demonstrated lower sensitivity (54.4%) but higher specificity (75.8%) in our cohort. The modest sensitivity of ESS reflects a fundamental characteristic of this instrument: it measures subjective daytime sleepiness rather than OSA-specific symptoms [27]. Importantly, not all patients with OSA experience excessive daytime sleepiness, with studies suggesting that 30-50% of OSA patients have ESS scores ≤10 [28, 29]. This discordance between OSA severity and subjective sleepiness likely results from individual variation in arousal thresholds, sleep debt compensation mechanisms, and genetic factors influencing sleepiness perception [30]. Consequently, relying solely on ESS for OSA screening would miss a substantial proportion of affected individuals, particularly those with mild disease or without prominent sleepiness symptoms.

The higher specificity of ESS suggests that when patients do report excessive daytime sleepiness, there is a strong likelihood of underlying OSA, reflected in the positive predictive value of 82.5%. However, the low negative predictive value (45.0%) indicates that normal ESS scores do not reliably exclude OSA, limiting its utility as a standalone screening tool for ruling out the condition [31].

Our findings regarding reduced sensitivity of both instruments in mild OSA warrant emphasis. The Berlin Questionnaire sensitivity decreased from 88.5% in severe OSA to 70.6% in mild OSA, while ESS sensitivity declined from 71.2% to 42.6%. This pattern is consistent with prior research demonstrating that screening questionnaires perform better in more severe disease [32, 33]. Mild OSA cases often present with less prominent symptoms, making them difficult to capture through subjective questionnaires. This observation highlights a significant limitation of questionnaire-based screening: the patients who might benefit most from early identification and intervention may be those least likely to be detected [34].

The moderate correlation between ESS scores and AHI values (r=0.42) observed in our study corroborates previous

research demonstrating weak-to-moderate associations between subjective sleepiness and objective OSA severity [35, 36]. This dissociation underscores that ESS primarily assesses one symptom dimension of OSA rather than the full spectrum of clinical manifestations. Other symptoms such as snoring, witnessed apneas, nocturnal awakenings, and morning headaches—components assessed by the Berlin Questionnaire—may provide complementary information that enhances overall screening performance [37].

Our results regarding the Berlin Questionnaire's superior overall accuracy (75.0% vs 60.1%) support its preferential use as a primary screening tool in populations with high OSA pretest probability, such as sleep clinic referrals. However, in primary care or population screening contexts where pretest probability is lower, the higher specificity of ESS might help reduce unnecessary PSG referrals [38]. The optimal choice of screening instrument should thus be tailored to the specific clinical setting and screening objectives.

The combined screening approach analyzed in our study reveals important trade-offs. Requiring both instruments to be positive (AND strategy) increased specificity to 86.4% but reduced sensitivity to 47.3%, potentially missing many OSA cases. Conversely, considering either instrument positive (OR strategy) improved sensitivity to 85.7% but decreased specificity to 54.5%, leading to more false-positive referrals. These findings suggest that sequential screening strategies might be optimized differently based on whether the goal is to maximize case detection (use OR approach) or to prioritize PSG resource allocation (use AND approach) [39].

Several methodological strengths of our study warrant mention. The use of laboratory-based PSG with manual scoring according to AASM criteria provided high-quality reference standard diagnoses. The consecutive enrollment approach minimized selection bias, and the relatively large sample size provided adequate statistical power for comparative analyses. The diverse participant characteristics enhanced generalizability to typical sleep clinic populations.

However, certain limitations should be acknowledged. The study was conducted at a single tertiary sleep center with a referred patient population, likely enriching for higher OSA prevalence (73.4%) compared to community samples. This elevated prevalence influences positive and negative predictive values, which are prevalence-dependent metrics. Consequently, our findings may have limited generalizability to primary care or unselected populations where OSA prevalence is lower [40]. Future studies examining these screening tools in primary care settings would provide valuable complementary evidence.

The cross-sectional design precludes assessment of how screening tool performance might vary with treatment or disease progression over time. Additionally, we evaluated only the standard ESS cutoff of >10 and explored alternative thresholds through ROC analysis, but we did not examine recently proposed modified scoring approaches for ESS that might enhance performance [41].

Cultural and linguistic factors may influence questionnaire interpretation and response patterns, potentially affecting screening tool performance across different populations. While our questionnaires were administered in the local language, cross-cultural validation was beyond the scope of this study. Future research examining these instruments across diverse ethnic and cultural groups would strengthen evidence for global applicability [42].

Despite these limitations, our findings contribute to the growing body of evidence regarding optimal OSA screening strategies. The results support current recommendations that suggest using multiple assessment modalities rather than relying on single instruments [43]. Clinical evaluation incorporating detailed history, physical examination (including assessment of upper airway anatomy, body habitus, and cardiovascular comorbidities), and judicious use of validated questionnaires likely provides the most effective approach to identifying patients requiring PSG [44, 45].

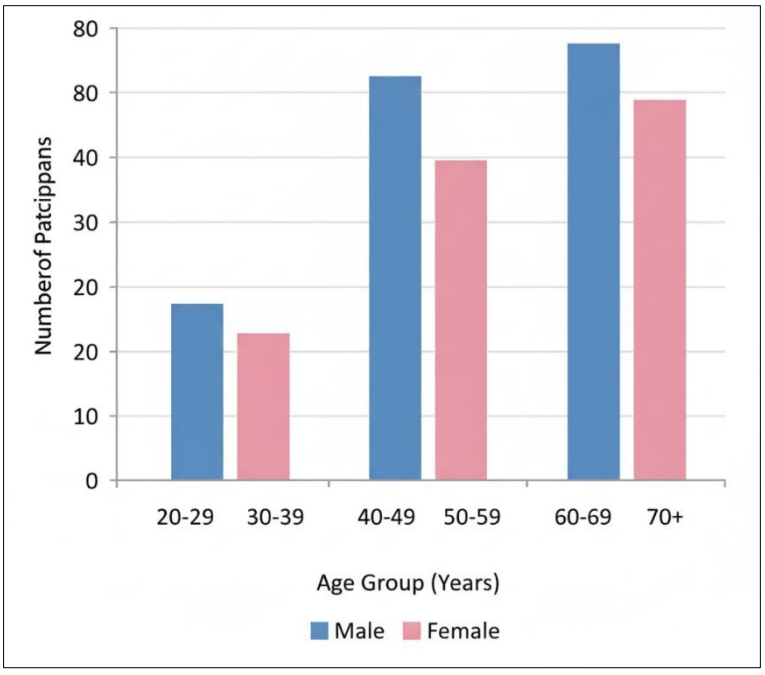

Fig 1: Age and Gender Distribution

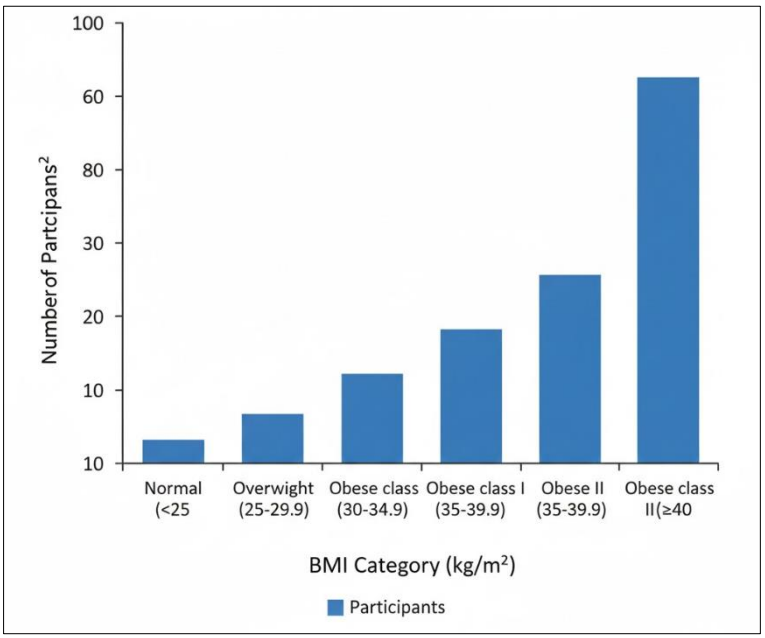

Fig 2: Body Mass Index Distribution

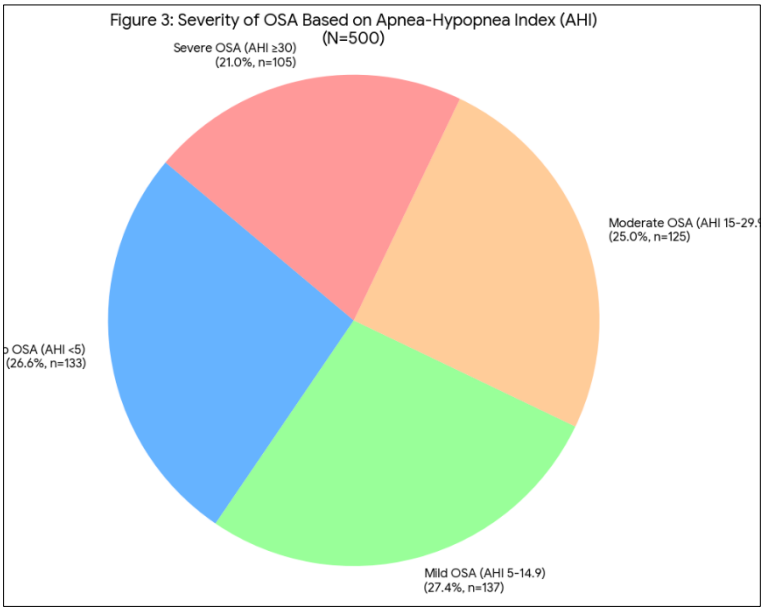

Fig 3: Severity of OSA Based on Apnea-Hypopnea Index (AHI)

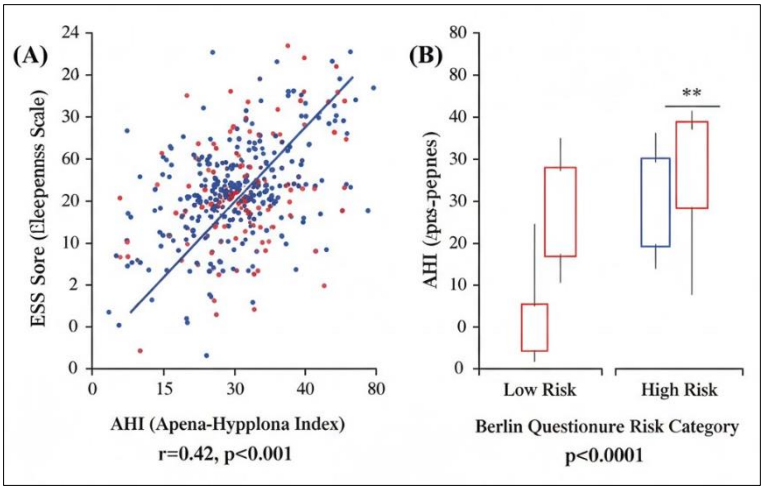

Fig 4: Correlation of AHI with ESS and Berlin Questionnaire Scores

Conclusion

This comparative polysomnography-based evaluation demonstrates that the Berlin Questionnaire exhibits superior sensitivity (78.6%) for obstructive sleep apnea screening compared to the Epworth Sleepiness Scale (54.4%), while ESS demonstrates higher specificity (75.8% vs 65.2%). The Berlin Questionnaire's higher sensitivity makes it preferable for initial OSA case identification in clinical practice, minimizing false-negative results and ensuring that most affected individuals are identified for further evaluation. The ESS, with its higher specificity, may serve as a complementary tool for assessing the specific symptom of daytime sleepiness and providing additional clinical information.

Both screening instruments demonstrate reduced sensitivity in mild OSA compared to moderate-to-severe disease, highlighting an inherent limitation of questionnaire-based screening approaches. The modest correlation between subjective symptoms and objective OSA severity underscores that neither tool can replace polysomnography for definitive diagnosis.

Combined or sequential use of both instruments may optimize screening effectiveness depending on clinical objectives, with the choice between maximizing sensitivity (OR approach) or specificity (AND approach) tailored to specific healthcare contexts and resource availability. However, neither screening tool alone nor in combination can substitute for comprehensive clinical evaluation and polysomnographic confirmation when OSA is suspected based on clinical presentation.

Recommendations

Based on the findings of this study, the following recommendations are proposed:

1. **Primary Screening Tool Selection:** The Berlin Questionnaire should be prioritized as the primary screening instrument in clinical settings where maximizing OSA case identification is the objective, given its superior sensitivity and overall diagnostic accuracy.
2. **Complementary Assessment:** The Epworth Sleepiness Scale should be used as a complementary tool to assess subjective daytime sleepiness, which has important implications for treatment decisions, particularly regarding CPAP therapy initiation and monitoring treatment response.
3. **Comprehensive Clinical Evaluation:** Screening questionnaires should be integrated into comprehensive clinical assessment that includes detailed history, physical examination, and consideration of comorbid conditions, rather than being used as standalone diagnostic tools.
4. **Context-Specific Strategies:** Healthcare systems should adapt screening approaches based on local resources and objectives. In settings with limited PSG availability, higher specificity thresholds (ESS or combined instruments) might be appropriate to optimize resource utilization, while in primary care settings focused on case finding, higher sensitivity approaches should be prioritized.
5. **Patient Education:** Clinicians should educate patients that normal screening questionnaire scores do not definitively exclude OSA, particularly in individuals with high pretest probability based on risk factors such as obesity, male gender, or witnessed apneas.
6. **Sequential Screening:** In resource-constrained settings, sequential screening strategies might be implemented where initial positive Berlin Questionnaire results trigger ESS administration, with PSG referral reserved for patients positive on both instruments.
7. **Periodic Reassessment:** Patients with initially negative screening results who have persistent symptoms or develop new OSA risk factors (such as significant weight gain) should undergo repeat screening and consideration for PSG.
8. **Quality Improvement:** Sleep centers and primary care practices should implement systematic OSA screening programs using validated questionnaires, with periodic audit of screening performance and refinement of local protocols based on outcomes data.
9. **Research Priorities:** Future research should focus on: (a) validating these screening tools in diverse populations and primary care settings, (b) developing and validating objective screening technologies that may complement or enhance questionnaire performance, (c) examining cost-effectiveness of various screening strategies, and (d) evaluating whether enhanced screening and earlier OSA detection improve long-term cardiovascular and metabolic outcomes.
10. **Integration with Technology:** Healthcare systems should explore integration of electronic health record-based automatic screening tools that incorporate questionnaire data, anthropometric measures, and comorbidity information to generate OSA risk scores that trigger appropriate clinical pathways.

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(9):1006-14.
2. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, *et al*. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med. 2019;7(8):687-98.
3. Young T, Skatrud J, Peppard PE. Risk factors for obstructive sleep apnea in adults. JAMA. 2004;291(16):2013-6.
4. Kapur V, Strohl KP, Redline S, Iber C, O'Connor G, Nieto J. Underdiagnosis of sleep apnea syndrome in U.S. communities. Sleep Breath. 2002;6(2):49-54.
5. Marin JM, Carrizo SJ, Vicente E, Agusti AG. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. Lancet. 2005;365(9464):1046-53.
6. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med. 2005;353(19):2034-41.
7. Punjabi NM, Caffo BS, Goodwin JL, Gottlieb DJ, Newman AB, O'Connor GT, *et al*. Sleep-disordered breathing and mortality: a prospective cohort study. PLoS Med. 2009;6(8):e1000132.
8. Streatfeild J, Smith J, Mansfield D, Pezzullo L, Hillman D. The social and economic cost of sleep disorders. Sleep. 2021;44(11):zsab132.

9. Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr, *et al*. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep. 2005;28(4):499-521.
10. Flemons WW, Douglas NJ, Kuna ST, Rodenstein DO, Wheatley J. Access to diagnosis and treatment of patients with suspected sleep apnea. Am J Respir Crit Care Med. 2004;169(6):668-72.
11. Abrishami A, Khajehdehi A, Chung F. A systematic review of screening questionnaires for obstructive sleep apnea. Can J Anaesth. 2010;57(5):423-38.
12. Ramachandran SK, Josephs LA. A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology. 2009;110(4):928-39.
13. Chiu HY, Chen PY, Chuang LP, Chen NH, Tu YK, Hsieh YJ, *et al*. Diagnostic accuracy of the Berlin questionnaire, STOP-BANG, STOP, and Epworth sleepiness scale in detecting obstructive sleep apnea: a bivariate meta-analysis. Sleep Med Rev. 2017;36:57-70.
14. 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(7):485-91.
15. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991;14(6):540-5.
16. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, *et al*. Validation of the Berlin questionnaire and American Society of Anesthesiologists checklist as screening tools for obstructive sleep apnea in surgical patients. Anesthesiology. 2008;108(5):822-30.
17. Sforza E, Chouchou F, Pichot V, Herrmann F, Barthélémy JC, Roche F. Is the Berlin Questionnaire a useful tool to diagnose obstructive sleep apnea in the elderly? Sleep Med. 2011;12(2):142-6.
18. Tan A, Yin JD, Tan LW, van Dam RM, Cheung YY, Lee CH. Using the Berlin Questionnaire to predict obstructive sleep apnea in the general population. J Clin Sleep Med. 2017;13(3):427-32.
19. Ahmadi N, Chung SA, Gibbs A, Shapiro CM. The Berlin questionnaire for sleep apnea in a sleep clinic population: relationship to polysomnographic measurement of respiratory disturbance. Sleep Breath. 2008;12(1):39-45.
20. Chai-Coetzer CL, Antic NA, Rowland LS, Reed RL, Esterman A, Catcheside PG, *et al*. Primary care vs specialist sleep center management of obstructive sleep apnea and daytime sleepiness and quality of life: a randomized trial. JAMA. 2013;309(10):997-1004.
21. Pang KP, Terris DJ. Screening for obstructive sleep apnea: an evidence-based analysis. Am J Otolaryngol. 2006;27(2):112-8.
22. Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ, *et al*. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. J Clin Sleep Med. 2007;3(7):737-47.
23. Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, *et al*. Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. J Clin Sleep Med. 2012;8(5):597-619.
24. Kunisaki KM, Brown KE, Fabbrini AE, Wetherbee EE, Rector TS. Validation of the Berlin Questionnaire in a sleep clinic population. Sleep Breath. 2011;15(4):825-37.
25. Sharma SK, Vasudev C, Sinha S, Banga A, Pandey RM, Handa KK. Validation of the modified Berlin questionnaire to identify patients at risk for the obstructive sleep apnoea syndrome. Indian J Med Res. 2006;124(3):281-90.
26. Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea?: The Rational Clinical Examination systematic review. JAMA. 2013;310(7):731-41.
27. Gottlieb DJ, Whitney CW, Bonekat WH, Iber C, James GD, Lebowitz M, *et al*. Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. Am J Respir Crit Care Med. 1999;159(2):502-7.
28. Baldwin CM, Griffith KA, Nieto FJ, O'Connor GT, Walsleben JA, Redline S. The association of sleep-disordered breathing and sleep symptoms with quality of life in the Sleep Heart Health Study. Sleep. 2001;24(1):96-105.
29. Chervin RD. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. Chest. 2000;118(2):372-9.
30. Gottlieb DJ, Hek K, Chen TH, Watson NF, Eiriksdottir G, Byrne EM, *et al*. Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. Mol Psychiatry. 2015;20(10):1232-9.
31. Benbadis SR, Mascha E, Perry MC, Wolgamuth BR, Smolley LA, Dinner DS. Association between the Epworth sleepiness scale and the multiple sleep latency test in a clinical population. Ann Intern Med. 1999;130(4 Pt 1):289-92.
32. 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(5 Pt 1):1279-85.
33. Rowley JA, Aboussouan LS, Badr MS. The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. Sleep. 2000;23(7):929-38.
34. Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, *et al*. A survey screen for prediction of apnea. Sleep. 1995;18(3):158-66.
35. Vana KD, Silva GE, Goldberg R. Predictive abilities of the STOP-Bang and Epworth Sleepiness Scale in identifying sleep clinic patients at high risk for obstructive sleep apnea. Res Nurs Health. 2013;36(1):84-94.
36. Bixler EO, Vgontzas AN, Lin HM, Calhoun SL, Vela-Bueno A, Kales A. Excessive daytime sleepiness in a general population sample: the role of sleep apnea, age, obesity, diabetes, and depression. J Clin Endocrinol Metab. 2005;90(8):4510-5.
37. Kribbs NB, Pack AI, Kline LR, Smith PL, Schwartz AR, Schubert NM, *et al*. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. Am Rev Respir Dis. 1993;147(4):887-95.
38. Chai-Coetzer CL, Antic NA, Hamilton GS, McArdle N, Wong K, Yee BJ, *et al*. Physician decision making and clinical outcomes with laboratory polysomnography or limited-channel sleep studies for obstructive sleep apnea: a randomized trial. Ann Intern Med. 2017;166(5):332-40.

39. Senaratna CV, Perret JL, Lodge CJ, Lowe AJ, Campbell BE, Matheson MC, *et al*. Prevalence of obstructive sleep apnea in the general population: a systematic review. Sleep Med Rev. 2017;34:70-81.
40. Masa JF, Corral J, Pereira R, Duran-Cantolla J, Cabello M, Hernández-Blasco L, *et al*. Effectiveness of sequential automatic-manual home respiratory polygraphy scoring. Eur Respir J. 2013;41(4):879-87.
41. Kendzerska T, Gershon AS, Hawker G, Leung RS, Tomlinson G. Obstructive sleep apnea and risk of cardiovascular events and all-cause mortality: a decade-long historical cohort study. PLoS Med. 2014;11(2):e1001599.
42. Rosenthal LD, Dolan DC. The Epworth sleepiness scale in the identification of obstructive sleep apnea. J Nerv Ment Dis. 2008;196(5):429-31.
43. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, *et al*. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med. 2017;13(3):479-504.
44. Patil SP, Ayappa IA, Caples SM, Kimoff RJ, Patel SR, Harrod CG. Treatment of adult obstructive sleep apnea with positive airway pressure: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. J Clin Sleep Med. 2019;15(2):301-34.
45. Lim DC, Pack AI. Obstructive sleep apnea: update and future. Annu Rev Med. 2017;68:99-112.

How to Cite This Article
Solanki S, Redhu S, Anash M, Sahay S, Saini R. Comparative evaluation of the Berlin Questionnaire and Epworth Sleepiness Scale for screening obstructive sleep apnea: a polysomnography-based study. Int J Med All Body Health Res. 2026;7(1):24–33. doi:10.54660/IJMBHR.2026.7.1.24-33.