JCSM
Journal of Clinical Sleep Medicine

SPECIAL ARTICLES

Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline

Vishesh K. Kapur, MD, MPH[1]; Dennis H. Auckley, MD[2]; Susmita Chowdhuri, MD[3]; David C. Kuhlmann, MD[4]; Reena Mehra, MD, MS[5]; Kannan Ramar, MBBS, MD[6]; Christopher G. Harrod, MS[7]

[1]University of Washington, Seattle, WA; [2]MetroHealth Medical Center and Case Western Reserve University, Cleveland, OH; [3]John D. Dingell VA Medical Center and Wayne State University, Detroit, MI; [4]Bothwell Regional Health Center, Sedalia, MO; [5]Cleveland Clinic, Cleveland, OH; [6]Mayo Clinic, Rochester, MN; [7]American Academy of Sleep Medicine, Darien, IL

Introduction: This guideline establishes clinical practice recommendations for the diagnosis of obstructive sleep apnea (OSA) in adults and is intended for use in conjunction with other American Academy of Sleep Medicine (AASM) guidelines on the evaluation and treatment of sleep-disordered breathing in adults.

Methods: The AASM commissioned a task force of experts in sleep medicine. A systematic review was conducted to identify studies, and the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) process was used to assess the evidence. The task force developed recommendations and assigned strengths based on the quality of evidence, the balance of benefits and harms, patient values and preferences, and resource use. In addition, the task force adopted foundational recommendations from prior guidelines as "good practice statements", that establish the basis for appropriate and effective diagnosis of OSA. The AASM Board of Directors approved the final recommendations.

Recommendations: The following recommendations are intended as a guide for clinicians diagnosing OSA in adults. Under GRADE, a STRONG recommendation is one that clinicians should follow under most circumstances. A WEAK recommendation reflects a lower degree of certainty regarding the outcome and appropriateness of the patient-care strategy for all patients. The ultimate judgment regarding propriety of any specific care must be made by the clinician in light of the individual circumstances presented by the patient, available diagnostic tools, accessible treatment options, and resources.

Good Practice Statements:

Diagnostic testing for OSA should be performed in conjunction with a comprehensive sleep evaluation and adequate follow-up.

Polysomnography is the standard diagnostic test for the diagnosis of OSA in adult patients in whom there is a concern for OSA based on a comprehensive sleep evaluation.

Recommendations:

1. We recommend that clinical tools, questionnaires and prediction algorithms not be used to diagnose OSA in adults, in the absence of polysomnography or home sleep apnea testing. (STRONG)
2. We recommend that polysomnography, or home sleep apnea testing with a technically adequate device, be used for the diagnosis of OSA in uncomplicated adult patients presenting with signs and symptoms that indicate an increased risk of moderate to severe OSA. (STRONG)
3. We recommend that if a single home sleep apnea test is negative, inconclusive, or technically inadequate, polysomnography be performed for the diagnosis of OSA. (STRONG)
4. We recommend that polysomnography, rather than home sleep apnea testing, be used for the diagnosis of OSA in patients with significant cardiorespiratory disease, potential respiratory muscle weakness due to neuromuscular condition, awake hypoventilation or suspicion of sleep related hypoventilation, chronic opioid medication use, history of stroke or severe insomnia. (STRONG)
5. We suggest that, if clinically appropriate, a split-night diagnostic protocol, rather than a full-night diagnostic protocol for polysomnography be used for the diagnosis of OSA. (WEAK)
6. We suggest that when the initial polysomnogram is negative and clinical suspicion for OSA remains, a second polysomnogram be considered for the diagnosis of OSA. (WEAK)

Keywords: obstructive sleep apnea, diagnosis, polysomnography, home sleep testing

Citation: Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, Harrod CG. 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.

INTRODUCTION

The diagnosis of obstructive sleep apnea (OSA) was previously addressed in two American Academy of Sleep Medicine (AASM) guidelines, the "Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005" and "Clinical Guidelines for the Use of Unattended Portable Monitors in the Diagnosis of Obstructive Sleep Apnea in Adult Patients (2007)."[1,2] The AASM commissioned a task force (TF) of content experts to develop an updated clinical practice guideline (CPG) on this topic. The objectives of this CPG are to combine and update information from prior guideline documents regarding the diagnosis of OSA, including the optimal circumstances under which attended in-laboratory polysomnography (heretofore referred to as "polysomnography" or "PSG") or home sleep apnea testing (HSAT) should be performed.

BACKGROUND

The term sleep-disordered breathing (SDB) encompasses a range of disorders, with most falling into the categories of OSA, central sleep apnea (CSA) or sleep-related hypoventilation. This paper focuses on diagnostic issues related to the diagnosis of OSA, a breathing disorder characterized by narrowing of the upper airway that impairs normal ventilation during sleep. Recent reviews on the evaluation and management of CSA and sleep-related hypoventilation have been published separately by the AASM.[3–5]

The prevalence of OSA varies significantly based on the population being studied and how OSA is defined (e.g., testing methodology, scoring criteria used, and apnea-hypopnea index [AHI] threshold). The prevalence of OSA has been estimated to be 14% of men and 5% of women, in a population-based study utilizing an AHI cutoff of ≥ 5 events/h (hypopneas associated with 4% oxygen desaturations) combined with clinical symptoms to define OSA.[6] OSA may impact a larger proportion of the population than indicated by these numbers, as the definition of AHI used in this study was restrictive and did not consider hypopneas that disrupt sleep without oxygen desaturation. In addition, the estimate excludes individuals with an elevated AHI who do not have sleepiness but who may nevertheless be at risk for adverse consequences such as cardiovascular disease.[7–10] In some populations, the prevalence of OSA is substantially higher than this estimate, for example, in patients being evaluated for bariatric surgery (estimated range of 70% to 80%)[11] or in patients who have had a transient ischemic attack or stroke (estimated range of 60% to 70%).[12] Other disease-specific populations found to have increased rates of OSA include, but are not limited to, patients with coronary artery disease, congestive heart failure, arrhythmias, refractory hypertension, type 2 diabetes, and polycystic ovarian disease.[13,14]

The consequences of untreated OSA are wide ranging and are postulated to result from the fragmented sleep, intermittent hypoxia and hypercapnea, intrathoracic pressure swings, and increased sympathetic nervous activity that accompanies disordered breathing during sleep. Individuals with OSA often feel unrested, fatigued, and sleepy during the daytime. They may suffer from impairments in vigilance, concentration, cognitive function, social interactions and quality of life (QOL). These declines in daytime function can translate into higher rates of job-related and motor vehicle accidents.[15] Patients with untreated OSA may be at increased risk of developing cardiovascular disease, including difficult-to-control blood pressure, coronary artery disease, congestive heart failure, arrhythmias and stroke.[16] OSA is also associated with metabolic dysregulation, affecting glucose control and risk for diabetes.[17] Undiagnosed and untreated OSA is a significant burden on the healthcare system, with increased healthcare utilization seen in those with untreated OSA,[18] highlighting the importance of early and accurate diagnosis of this common disorder.

Recognizing and treating OSA is important for a number of reasons. The treatment of OSA has been shown to improve QOL, lower the rates of motor vehicle accidents, and reduce the risk of the chronic health consequences of untreated OSA mentioned above.[19] There are also data supporting a decrease in healthcare utilization and cost following the diagnosis and treatment of OSA.[20] However, there are challenges and uncertainties in making the diagnosis and a number of questions remain unanswered.

Individuals with OSA can also have other sleep disorders that may be related to or unrelated to OSA. Co-morbid insomnia has been found to be a frequent problem in patients with OSA.[21] It is also possible that undiagnosed OSA may be masquerading as another sleep disorder, such as REM Behavior Disorder.[22] Therefore, when OSA is suspected, a comprehensive sleep evaluation is important to ensure appropriate diagnostic testing is performed to address OSA, as well as other comorbid sleep complaints.

The diagnosis of OSA involves measuring breathing during sleep. The evolution of measurement techniques and definitions of abnormalities justifies updating the guidelines regarding diagnostic testing, but also complicates the evaluation and summary of evidence gathered from older research studies that have included diagnostic tests with diverse sensor types and scored respiratory events using different definitions. The third edition of the International Classification of Sleep Disorders (ICSD-3) defines OSA as a PSG-determined obstructive respiratory disturbance index (RDI) ≥ 5 events/h associated with the typical symptoms of OSA (e.g., unrefreshing sleep, daytime sleepiness, fatigue or insomnia, awakening with a gasping or choking sensation, loud snoring, or witnessed apneas), or an obstructive RDI ≥ 15 events/h (even in the absence of symptoms).[23] In addition to apneas and hypopneas that are included in the AHI, the RDI includes respiratory effort-related arousals (RERAs). The scoring of respiratory events is defined in The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.3 (AASM Scoring Manual).[24] However, it should be noted that there is variability in the definition of a hypopnea event. The AASM Scoring Manual recommended definition requires that changes in flow be associated with a 3% oxygen

desaturation or a cortical arousal, but allows an alternative definition that requires association with a 4% oxygen desaturation without consideration of cortical arousals. Depending on which definition is used, the AHI may be considerably different in a given individual.[25–27] The discrepancy between these and other hypopnea definitions used in research studies introduces complexity in the evaluation of evidence regarding the diagnosis of OSA.

Due to the high prevalence of OSA, there is significant cost associated with evaluating all patients suspected of having OSA with PSG (currently considered the gold standard diagnostic test). Further, there also may be limited access to in-laboratory testing in some areas. HSAT, which has limitations, is an alternative method to diagnose OSA in adults, and may be less costly and more efficient in some populations. This guideline addresses some of these issues using an evidence-based approach.

There are potential disadvantages to using HSAT, relative to PSG, because of the differences in the physiologic parameters being collected and the availability of personnel to adjust sensors when needed. The sensor technology used by HSAT devices varies considerably by the number and type of sensors that are utilized. Traditionally, sleep studies have been categorized as Type I, Type II, Type III or Type IV. Unattended studies fall into categories Type II through Type IV. Type II studies use the same monitoring sensors as full PSGs (Type I) but are unattended, and thus can be performed outside of the sleep laboratory. Type III studies use devices that measure limited cardiopulmonary parameters; two respiratory variables (e.g., effort to breathe, airflow), oxygen saturation, and a cardiac variable (e.g., heart rate or electrocardiogram). Type IV studies utilize devices that measure only 1 or 2 parameters, typically oxygen saturation and heart rate, or in some cases, just air flow. This classification of sleep study devices fails to consider new technologies, such as peripheral arterial tonometry (PAT), and thus an alternative classification scheme has been proposed: the SCOPER classification, which incorporates Sleep, Cardiovascular, Oximetry, Position, Effort and Respiratory parameters.[28] The SCOPER system allows for the inclusion of technologies such as PAT. However, due to the complexity of the SCOPER classification, and lack of familiarity with it amongst practicing clinicians, the TF elected to refer to HSAT devices by the traditional Type II through Type IV classification system, and to identify specific devices with technology outside of this schema when appropriate. Regardless, as can be recognized by both classifications, HSAT devices in comparison to attended studies raise risk for technical failures due to a lack of real-time monitoring, and have inherent limitations resulting from the inability of most devices to define sleep versus wake. Another potential disadvantage is that positive airway pressure (PAP) cannot be initiated during a HSAT, but can be initiated during a PSG if needed.

Measurement error is inevitable in HSAT, compared against PSG, as standard sleep staging channels are not typically monitored in HSAT (e.g., EEG, EOG and EMG monitoring are not typically performed), which results in use of the recording time rather than sleep time to define the denominator of the respiratory event index (REI; the term used to represent the frequency of apneas and hypopneas derived from HSAT). HSAT devices that use conventional sensors are unable to detect hypopneas only associated with cortical arousals, which are included in the recommended AHI scoring rule in the AASM Scoring Manual.[24] Sensor dislodgement and poor quality signal during HSAT are additional contributors to the measurement error of the REI. All these factors can result in the underestimation of the "true" AHI, and may result in the need for repeated studies due to inadequate data for diagnosis.

As a diagnostic guideline, our systematic review and recommendations incorporate evidence regarding the accuracy of HSAT for diagnosing OSA. However, diagnosis occurs in the context of management of a patient within the healthcare system, and therefore, outcomes other than diagnostic accuracy are relevant in the evaluation of management strategies. These include the impact on clinical outcomes (e.g., sleepiness, QOL, morbidity, mortality, adherence to therapy) and efficiency of care (e.g., time to test, time to treatment, costs). Therefore, these outcomes are also considered in the formulation of the current guideline.

Prior AASM guidelines[1,2] on the diagnosis of OSA included statements that the TF determined were no longer pertinent. Thus, these statements were not addressed in the current update. Moreover, prior guidelines included consensus statements that had not been specifically evaluated in clinical studies. Despite this limitation, two of these statements were adopted in the current guideline as foundational statements that underpin the provision of high quality care for the diagnosis of OSA (see good practice statements). The scope of this guideline did not include a comprehensive update of technical specification for diagnostic testing for OSA. Nevertheless, the TF considered whether currently recommended technology was used in the research studies that were evaluated. In particular, the TF determined that the use of currently AASM recommended flow (nasal pressure transducer and thermistor) and effort sensors (respiratory inductance plethysmography) during PSG and HSAT increased the value of evidence derived from validation studies.[24] As part of the data extraction process, validation studies were classified based on whether the currently recommended respiratory sensors were used for PSG or HSAT.

METHODS

Expert Task Force

The AASM commissioned a TF of board-certified sleep medicine physicians, with expertise in the diagnosis and management of adults with OSA, to develop this guideline. The TF was required to disclose all potential conflicts of interest (COI) according to the AASM's COI policy, both prior to being appointed to the TF, and throughout the research and writing of this paper. In accordance with the AASM's conflicts of interest policy, TF members with a Level 1 conflict were not allowed to participate. TF members with a Level 2 conflict were required to recuse themselves from any related discussion or writing responsibilities. All relevant conflicts of interest are listed in the Disclosures section.

Table 1—PICO questions.

1. In adult patients with suspected OSA, do clinical prediction algorithms accurately identify patients with a high pretest probability for OSA compared to history and physical exam? (*See Recommendation 1*)
2. In adult patients with suspected OSA, does HSAT accurately diagnose OSA, improve clinical outcomes and improve efficiency of care compared to PSG? (*See Recommendation 2 & 3*)
3. In adult patients undergoing HSAT for suspected OSA, is there a minimum number of hours of HSAT that accurately diagnoses OSA and improves efficiency of care? (*See Recommendation 2 & 3*)
4. In adult patients undergoing HSAT for suspected OSA, do multiple nights of HSAT accurately diagnose OSA and improve efficiency of care compared to a single night of HSAT? (*See Recommendation 2 & 3*)
5. In adult patients with comorbid conditions (poststroke, chronic heart failure, chronic obstructive pulmonary disease, opioid use, neuromuscular disease, hypoventilation, insomnia) and suspected OSA, does HSAT accurately diagnose OSA, improve clinical outcomes and efficiency of care compared to PSG? (*See Recommendation 4*)
6. In adult patients undergoing PSG for suspected OSA, does a split-night study accurately diagnose OSA and improve efficiency of care compared to a full-night study? (*See Recommendation 5*)
7. In adult patients undergoing PSG for suspected OSA, do two nights of PSG accurately diagnose OSA and improve efficiency of care compared to a single night of PSG? (*See Recommendation 6*)
8. In adult patients with diagnosed OSA, does repeat PSG or HSAT to confirm severity of OSA or efficacy of therapy improve outcomes relative to clinical follow-up without repeat testing? (*No recommendations, see Future Directions*)
9. In adult patients scheduled for upper airway surgery for snoring or OSA, does PSG or HSAT accurately identify patients with OSA and improve clinical outcomes compared to using a history and physical exam or clinical prediction algorithms? (*No recommendations, see Future Directions*)

Table 2—"Critical" outcomes by PICO.

PICO Question	Diagnostic Accuracy*	Subjective Sleepiness	Quality of Life**	CPAP Adherence	AHI	Depression	Cardiovascular Endpoints
1	✓						
2	FN only	✓	✓	✓	✓	✓	✓
3	✓	✓	✓				✓
4	✓						
5	✓	✓	✓	✓	✓	✓	✓
6	✓	✓	✓	✓	✓	✓	✓
7	✓						
8	✓						
9		✓	✓	✓		✓	✓

* = diagnostic accuracy is determined by the number of true positive (TP), false positive (FP), true negative (TN), false negative (FN) diagnoses. ** = based on Sleep Apnea Quality of Life Index (SAQLI) and Functional Outcomes of Sleep Questionnaire (FOSQ) measures of QOL. 36-Item Short Form Survey Instrument (SF-36) measure of QOL was determined to be important but not critical for decision-making based on TF consensus.

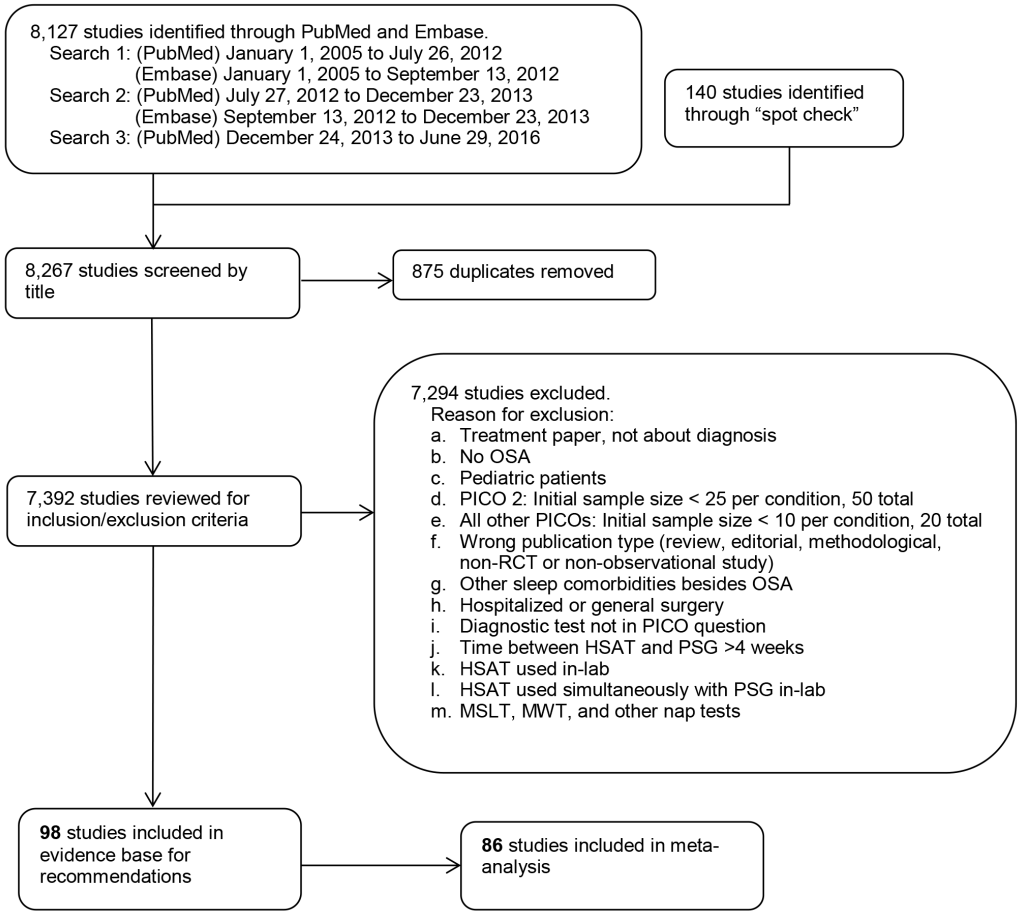
Table 3—Summary of clinical significance thresholds for clinical outcome measures.

Outcome Measure	Clinical Significance Threshold
Epworth Sleepiness Score (ESS)	2 points
Functional Outcomes of Sleep Questionnaire (FOSQ)	1 point
Sleep Apnea QOL Index (SAQLI)	1 point
CPAP Adherence (h/night)	0.5 h/night
CPAP Adherence (% nights > 4 h)	10%
SF-36 (Vitality Score)	12.5 points
SF-36 (Physical Component Summary Score)	3 points
SF-36 (Mental Component Summary Score)	3 points

PICO Questions

A PICO (Patient, Population or Problem, Intervention, Comparison, and Outcomes) question template was used to develop clinical questions to be addressed in this guideline. PICO questions were developed based on a review of the existing AASM practice parameters on indications for use of PSG and HSAT for the diagnosis of patients with OSA, and a review of systematic reviews, meta-analyses, and guidelines published since 2004. The AASM Board of Directors (BOD) approved the final list of PICO questions presented in **Table 1** before the literature search was performed. The PICO questions identify the commonly used approaches and devices for the diagnosis of OSA. Based on their expertise, the TF developed a list of patient-oriented clinically relevant outcomes that are indicative of whether a treatment should be recommended for clinical practice. A summary of the critical outcomes for each PICO is presented in **Table 2**. Lastly, clinical significance thresholds, used to determine if a change in an outcome was clinically significant, were defined for each outcome by TF clinical judgment, prior to statistical analysis. The clinical significance thresholds are presented by outcome in **Table 3**. It should be noted that there was insufficient evidence to directly address PICO question 1, as no studies were identified that compared the efficacy of clinical prediction algorithms to history and physical exam. However, the TF decided to

Figure 1—Evidence base flow diagram.



compare the efficacy of clinical prediction algorithms to PSG and HSAT.

Literature Searches, Evidence Review and Data Extraction

The TF performed a systematic review of the scientific literature to identify articles that addressed at least one of the nine PICO questions. Multiple literature searches were performed by AASM staff using the PubMed and Embase databases, throughout the guideline development process (see **Figure 1**). The search yielded articles with various study designs, however the analysis was limited to randomized controlled trials (RCTs) and observational studies. The articles that were cited in the 2007 AASM clinical practice guideline,[2] 2005 practice parameter,[1] 2003 review,[29] and 1997 review[30] were included for data analysis if they met the study inclusion criteria described below.

The literature searches in PubMed were conducted using a combination of MeSH terms and keywords as presented in the supplemental material. The PubMed database was searched from January 1, 2005 through July 26, 2012 for any relevant literature published since the last guideline. The PubMed search was expanded on September 26, 2012 to identify relevant articles published prior to January 1, 2005. Literature searches also were also performed in Embase using a combination of terms and keywords as presented in the supplemental material. The Embase database was searched from January 1, 2005 through September 13, 2012. These searches yielded a total of 3,937 articles. There were 205 duplicates identified resulting in a total of 3,732 articles from both databases.

A second round of literature searches was performed in PubMed and Embase to capture more recent literature. The PubMed database was searched from July 27, 2012 to December 23, 2013, and the Embase database was searched from September 13, 2012 to December 23, 2013. These searches yielded a total of 2,061 articles. There were 670 duplicates identified resulting in 1,391 additional papers from both databases.

A final literature search was performed in PubMed to capture the latest literature. The PubMed database was searched from December 24, 2013 to June 29, 2016 and identified 2,129 articles.

Based on their expertise and familiarity with the literature, TF members submitted additional relevant literature and screened reference lists to identify articles of potential interest. This served as a "spot check" for the literature searches to ensure that important articles were not overlooked and identified an additional 140 publications.

A total of 7,392 abstracts were assessed by two reviewers to determine whether they met inclusion criteria presented in the supplemental material. Articles were excluded per the criteria

Table 4—Summary of prevalence estimates for high risk and low risk adult sleep clinic patients with OSA by diagnostic cutoff.

Diagnostic Cutoff	High-Risk Prevalence	Low-Risk Prevalence
AHI ≥ 5	87%	55%
AHI ≥ 15	64%	25%
AHI ≥ 30	36%	10%

listed in the supplemental material and **Figure 1**. A total of 98 articles were included in evidence base for recommendations. A total of 86 studies were included in meta-analysis and/or grading.

Meta-Analysis

Meta-analysis was performed on both diagnostic and clinical outcomes of interest for each PICO question, when possible. Outcomes data for diagnostic approaches were categorized as follows: clinical tools, questionnaires, and prediction algorithms; history and physical exam; HSAT; attended PSG; split-night attended PSG; two-night attended PSG; single-night HSAT; multiple-night HSAT; follow-up attended PSG; and follow-up HSAT. The type of HSAT devices identified in literature search included type 2; type 3; 2–3 channel; single channel; oximetry; and PAT. A definition of these devices has been previously described.[31] Adult patients were categorized as follows: suspected OSA; suspected OSA with comorbid conditions; diagnosed OSA; and scheduled for upper airway surgery.

For diagnostic outcomes, the pretest probability for OSA (i.e., the prevalence within the study population), sensitivity and specificity of the tested diagnostic approach, and number of patients for each study was used to derive two-by-two tables (i.e., the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) diagnoses per 1,000 patients) in both high risk and low risk patients, for each OSA severity threshold (i.e., AHI ≥ 5, AHI ≥ 15, AHI ≥ 30). For analyses that included five or more studies, pooled estimates of sensitivity, specificity, and accuracy were calculated using hierarchical random effects modeling performed in STATA software (accuracy was derived by HSROC curves). When analyses included fewer than five studies, ranges of sensitivity, specificity and accuracy were used. Based on their clinical expertise and a review of available literature, the TF established estimates of OSA prevalence among "low risk" and "high risk" patients for each OSA severity threshold. The TF envisioned a sleep clinic cohort of middle-aged obese men with typical symptoms of OSA as an example of a high-risk patient population. In contrast, a sleep clinic cohort of younger non-obese women with possible OSA symptoms was used as prototype for a low risk patient population. Prevalence estimates for these populations are presented in **Table 4**.

The sensitivity and specificity of included studies were entered into Review Manager 5.3 software to generate forest plots for each analysis. The estimates of sensitivity and specificity (pooled or ranges), and OSA prevalence were entered into the GRADE (Grading of Recommendations Assessment, Development and Evaluation) Guideline Development Tool (GDT) to generate the two-by-two tables. The TF determined the downstream consequences of an accurate diagnosis versus an inaccurate diagnosis (see supplemental material, **Table S1**), and used the estimates to weigh the benefits of an accurate diagnosis versus the harms of an inaccurate diagnosis. This information was used, in part, to assess whether a given diagnostic approach could be recommended when compared against PSG.

For clinical outcomes of interest, data on change scores were entered into the Review Manager 5.3 software to derive the mean difference and standard deviation between the experimental diagnostic approach and the gold standard or comparator. For studies that did not report change scores, data from posttreatment values taken from the last treatment time-point were used for meta-analysis. All meta-analyses of clinical outcomes were performed using the random effects model with results displayed as a forest plot. There was insufficient evidence to perform meta-analyses for PICOs 3 and 9, thus no recommendations are provided.

Interpretation of clinical significance for the clinical outcomes of interest was conducted by comparing the absolute effects to the clinical significance threshold previously determined by the TF for each clinical outcome of interest (see **Table 3**).

Strength of Recommendations

The assessment of evidence quality was performed according to the GRADE process.[32] The TF assessed the following four components to determine the direction and strength of a recommendation: quality of evidence, balance of beneficial and harmful effects, patient values and preferences and resource use as described below.

1. Quality of evidence: based on an assessment of the overall risk of bias (randomization, blinding, allocation concealment, selective reporting, and author disclosures), imprecision (clinical significance thresholds), inconsistency (I^2 cutoff of 75%), indirectness (study population), and risk of publication bias (funding sources), the TF determined their overall confidence that the estimated effect found in the body of evidence was representative of the true treatment effect that patients would see. For diagnostic accuracy studies, the QUADAS-2 tool was used in addition to the quality domains for the assessment of risk of bias in intervention studies. The quality of evidence was based on the outcomes that the TF deemed critical for decision-making.
2. Benefits versus harms: based on the meta-analysis (if applicable), analysis of any harms or side effects reported within the accepted literature, and the clinical expertise of the TF, the TF determined if the beneficial outcomes of the intervention outweighed any harmful side effects.
3. Patient values and preferences: based on the clinical expertise of the TF members and any data published on the topic relevant to patient preferences, the TF determined if patients would use the intervention based on the body of evidence, and if patient values and preferences would be generally consistent.
4. Resource use: based on the clinical expertise of the TF members and a "spot check" for relevant literature

the TF determined resource use to be important for determining whether to recommend the use of HSAT versus PSG, split-night versus full-night PSG and single-night versus multiple-night HSAT diagnostic protocols, and repeat testing. Resource use was not considered in-depth for clinical tools, questionnaires and prediction algorithms, diagnosis in adults with comorbid conditions, and repeat PSG.

Taking these major factors into consideration, each recommendation statement was assigned strength ("STRONG" or "WEAK"). Additional information is provided in the form of "Remarks" immediately following the recommendation statements, when deemed necessary by the TF. Remarks are based on the evidence evaluated during the systematic review, are intended to provide context for the recommendations, and to guide clinicians in implementing the recommendations in daily practice.

Discussions accompany each recommendation to summarize the relevant evidence and explain the rationale leading to each recommendation. These sections are an integral part of the GRADE system and offer transparency to the process.

Approval and Interpretation of Recommendations

A draft of the guideline was available for public comment for a two-week period on the AASM website. The TF took into consideration all the comments received and made revisions when appropriate. The revised guideline was submitted to the AASM BOD who approved these recommendations.

The recommendations in this guideline define principles of practice that should meet the needs of most patients in most situations. This guideline should not, however, be considered inclusive of all proper methods of care or exclusive of other methods of care reasonably used to obtain the same results. A STRONG recommendation is one that clinicians should, under most circumstances, always follow (i.e., something that might qualify as a Quality Measure). A WEAK recommendation reflects a lower degree of certainty in the appropriateness of the patient-care strategy and requires that the clinician use his/her clinical knowledge and experience, and refer to the individual patient's values and preferences to determine the best course of action. The ultimate judgment regarding the suitability of any specific recommendation must be made by the clinician, in light of the individual circumstances presented by the patient, the available diagnostic tools, the accessible treatment options, and available resources.

The AASM expects this guideline to have an impact on professional behavior, patient outcomes, and possibly, health care costs. This clinical practice guideline reflects the state of knowledge at the time of the literature review and will be re-examined and updated as new information becomes available.

CLINICAL PRACTICE RECOMMENDATIONS

The following clinical practice recommendations are based on a systematic review and evaluation of evidence following the GRADE methodology. Remarks are provided to guide clinicians in the implementation of these recommendations. All figures, including meta-analyses and Summary of Findings tables are presented in the supplemental material. **Table 5** shows a summary of the recommendation statements including the strength of recommendation and quality of evidence. A decision tree for the diagnosis of patients suspected of having OSA is presented in **Figure 2**.

The following are good practice statements, the implementation of which is deemed necessary for appropriate and effective diagnosis and management of OSA.

Diagnostic testing for OSA should be performed in conjunction with a comprehensive sleep evaluation and adequate follow-up.

OSA is one of many medical conditions that may be the cause of sleep complaints and other symptoms. Therefore, diagnostic testing for OSA is best carried out after a comprehensive sleep evaluation. The clinical evaluation for OSA should include a thorough sleep history and a physical examination that includes the respiratory, cardiovascular, and neurologic systems. The examiner should pay particular attention to observations regarding snoring, witnessed apneas, nocturnal choking or gasping, restlessness, and excessive sleepiness. It is also important that other aspects of a sleep history be collected, as many patients suffer from more than one sleep disorder or present with atypical sleep apnea symptoms. In addition, medical conditions associated with increased risk for OSA, such as obesity, hypertension, stroke, and congestive heart failure should be identified. The general evaluation should serve to establish a differential diagnosis, which can then be used to select the appropriate test(s). Follow-up, under the supervision of a board-certified sleep medicine physician, ensures that study findings and recommendations are relayed appropriately; and that appropriate expertise in prescribing and administering therapy is available to the patient.

The TF recognizes that there may be specific contexts (e.g., preoperative evaluation of OSA) in which evaluation of OSA needs to occur in an expedited manner, when it may not be practical to perform a comprehensive sleep evaluation prior to diagnostic testing. In such situations, the TF recommends a clinical pathway be developed and administered by a board-certified sleep medicine physician or appropriately licensed medical staff member designated by the board-certified sleep medicine physician. This pathway should include the following elements: a focused evaluation of sleep apnea performed by a clinical provider, and use of tools or questionnaires that capture clinically important information that is reviewed by a board-certified sleep medicine physician prior to testing. Following testing, a comprehensive sleep evaluation and follow-up under the supervision of a board-certified sleep medicine physician should be completed.

Polysomnography is the standard diagnostic test for the diagnosis of OSA in adult patients in whom there is a concern for OSA based on a comprehensive sleep evaluation.

Misdiagnosing patients can lead to significant harm due to lost benefits of therapy in those with OSA, and the prescription of

Table 5—Summary of recommendations.

Recommendation Statement	Strength of Recommendation	Evidence Quality	Benefits versus Harms	Patient Values and Preferences
1. We recommend that clinical tools, questionnaires or prediction algorithms not be used to diagnose OSA in adults, in the absence of PSG or HSAT.	Strong	Moderate	High certainty that harms outweigh benefits	Vast majority of well-informed patients would most likely not choose clinical tools, questionnaires or prediction algorithms for diagnosis
2. We recommend that PSG, or HSAT with a technically adequate device, be used for the diagnosis of OSA in uncomplicated adult patients presenting with signs and symptoms that indicate an increased risk of moderate to severe OSA.	Strong	Moderate	High certainty that benefits outweigh harms	Vast majority of well-informed patients would want PSG or HSAT
3. We recommend that if a single HSAT is negative, inconclusive or technically inadequate, PSG be performed for the diagnosis of OSA.	Strong	Low	High certainty that benefits outweigh harms	Vast majority of well-informed patients would want PSG performed if the initial HSAT is negative, inconclusive, or technically inadequate
4. We recommend that PSG, rather than HSAT, be used for the diagnosis of OSA in patients with significant cardiorespiratory disease, potential respiratory muscle weakness due to neuromuscular condition, awake hypoventilation or suspicion of sleep related hypoventilation, chronic opioid medication use, history of stroke or severe insomnia.	Strong	Very Low	High certainty that benefits outweigh harms	Vast majority of well-informed patients would most likely choose PSG to diagnose suspected OSA
5. We suggest that, if clinically appropriate, a split-night diagnostic protocol, rather than a full-night diagnostic protocol for PSG be used for the diagnosis of OSA.	Weak	Low	Low certainty that benefits outweigh harms	Majority of well-informed patients would most likely choose a split-night diagnostic protocol to diagnose suspected OSA
6. We suggest that when the initial PSG is negative, and there is still clinical suspicion for OSA, a second PSG be considered for the diagnosis of OSA.	Weak	Very low	Low certainty that benefits outweigh harms	Majority of well-informed patients would most likely choose a second PSG to diagnose suspected OSA when the initial PSG is negative and there is still a suspicion that OSA is present

inappropriate therapy in those without OSA. As discussed in the recommendations below, sleep apnea-focused questionnaires and clinical prediction rules lack sufficient diagnostic accuracy, and therefore direct measurement of SDB is necessary to establish a diagnosis of OSA. PSG is widely accepted as the gold standard test for diagnosis of OSA. Further, this test has traditionally been used as the gold standard for comparison to other diagnostic tests, including HSAT. Besides the diagnosis of OSA, PSG can identify co-existing sleep disorders, including other forms of sleep-disordered breathing. In some cases, and within the appropriate context, the use of HSAT as the initial sleep study may be acceptable, as discussed in the recommendations below. However, PSG should be used when HSAT results do not provide satisfactory posttest probability of confirming or ruling out OSA.

Diagnosis of Obstructive Sleep Apnea in Adults Using Clinical Tools, Questionnaires and Prediction Algorithms

Recommendation 1: We recommend that clinical tools, questionnaires and prediction algorithms not be used to diagnose OSA in adults, in the absence of polysomnography or home sleep apnea testing. (STRONG)

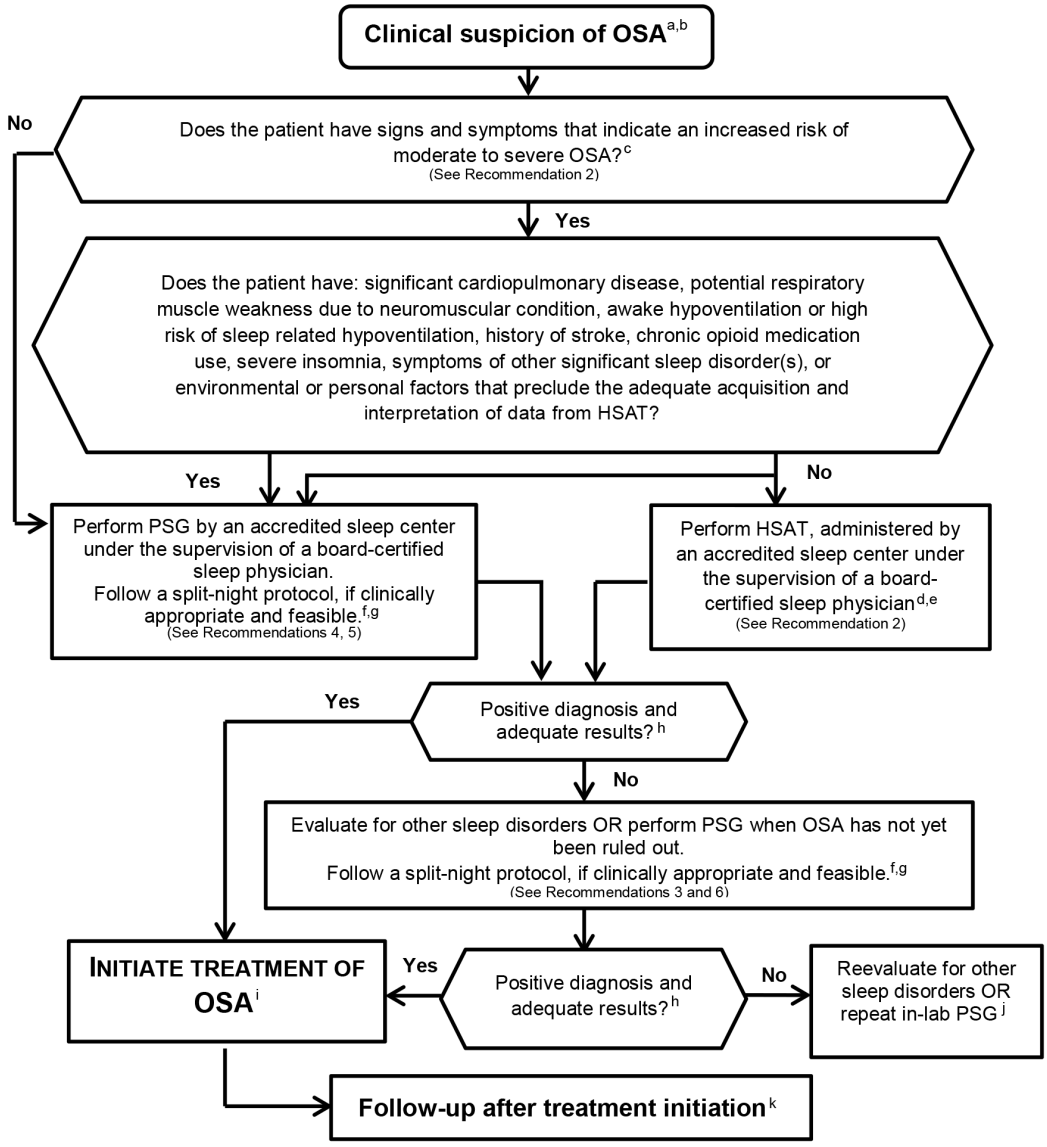
Summary

The literature search did not identify publications that directly compared the performance of clinical prediction algorithms to history and physical exam to accurately identify patients with OSA. However, our review identified forty-eight studies that compared the accuracy of clinical tools, questionnaires or prediction algorithms against PSG or HSAT. In the clinic-based setting, clinical tools, questionnaires and prediction algorithms have a low level of accuracy for the diagnosis of OSA at any threshold of AHI consideration. The overall quality of evidence was downgraded to moderate due to inconsistency and imprecision of findings.

Clinical prediction algorithms may be used in sleep clinic patients with suspected OSA, but are not necessary to establish the need for PSG or HSAT and further are not sufficient to substitute for PSG or HSAT. In non-sleep clinic settings, these tools may be more helpful to identify patients who are at increased risk for OSA, but this was beyond the scope of this guideline.

Evaluation with a clinical tool, questionnaire or prediction algorithm may be less burdensome to patients and clinicians than HSAT or PSG; however, their low levels of accuracy make them poor diagnostic tools. Therefore, based on clinical judgment, the TF determined that the harms of using clinical

Figure 2—Clinical algorithm for implementation of clinical practice guidelines.



a = Clinical suspicion based on a comprehensive sleep evaluation. b = Clinical tools, questionnaires and prediction algorithms should not be used to diagnose OSA in adults, in the absence of PSG or HSAT. c = Increased risk of moderate to severe OSA is indicated by the presence of excessive daytime sleepiness and at least two of the following three criteria: habitual loud snoring; witnessed apnea or gasping or choking; or diagnosed hypertension. d = This recommendation is based on conducting a single HSAT recording over at least one night. e = This recommendation is based on HSAT devices that incorporate a minimum of the following sensors: nasal pressure, chest and abdominal respiratory inductance plethysmography (RIP) and oximetry; or peripheral arterial tonometry (PAT) with oximetry and actigraphy. For additional information, refer to The AASM Manual for the Scoring of Sleep and Associated Events. f = A split-night protocol should only be conducted when the following criteria are met: (1) A moderate to severe degree of OSA is observed during a minimum of 2 hours of recording time on the diagnostic PSG; AND (2) At least 3 hours are available to complete CPAP titration. If these criteria are not met, a full-night diagnostic protocol should be followed. g = Clinically appropriate is defined as the absence of conditions identified by the clinician that are likely to interfere with successful diagnosis and treatment using a split-night protocol. h = A technically adequate HSAT includes a minimum of 4 hours of technically adequate oximetry and flow data, obtained during a recording attempt that encompasses the habitual sleep period. For additional information, refer to The AASM Manual for the Scoring of Sleep and Associated Events. i = Treatment of OSA should be initiated based on technically adequate PSG or HSAT study. j = Consider repeat in-laboratory PSG if clinical suspicion of OSA remains. k = There should be early follow-up after initiation of therapy.

tools, questionnaires, and prediction algorithms to confirm a suspected diagnosis of OSA outweigh the potential benefits. The TF also determined that a vast majority of patients would not favor the use of clinical questionnaires or prediction tools alone to establish the diagnosis of OSA.

Discussion

While the literature search did not identify publications that directly compared the performance of clinical prediction algorithms to history and physical exam, it did identify forty-eight validation studies that compared the accuracy of clinical tools,

questionnaires or prediction algorithms against PSG or HSAT. Our recommendations are therefore based on these validation studies, which are described. Relevant outcomes data from these validation studies are summarized in the supplemental material, **Table S2** through **Table S36**. Due to the uncertainty regarding clinical outcomes for patients misclassified by the prediction rules, the TF was unable to establish a cutoff for number of misclassified patients that would be considered acceptable. Nevertheless, all the clinical prediction models evaluated resulted in upper ranges of predicted false negatives per 1,000 patients that exceeded 100, a number that was determined by the TF to be clearly excessive for a stand-alone diagnostic test for OSA. In summary, the diagnostic performance of clinical questionnaires, morphometric models, and clinical prediction rules that consider multiple variables including symptoms, exam findings and subject characteristics, have all been evaluated against PSG and/or HSAT. Overall, while sensitivity appears to be in the higher range it is not sufficient to adequately exclude the possibility of OSA. Specificity tends to be lower, resulting in a higher number of false positives that further limit the utility of these clinical or morphometric rules and models in the diagnosis of OSA. It should also be noted that some of these studies were conducted in focused populations (e.g., commercial drivers, elderly, bariatric surgery patients, etc.), thus limiting generalizability. The following discussion has been organized to review the data by questionnaire or clinical prediction rule.

Berlin Questionnaire: The Berlin Questionnaire consists of eleven questions divided into three categories to classify the patient as high or low risk for OSA.[33] Our review identified nineteen studies that evaluated the performance of the Berlin Questionnaire against PSG in the identification of patients with OSA.[34–52] The studies were conducted in a wide variety of geographic locations including Brazil,[38] Canada,[34,42] Greece,[37] Iran,[36] Korea,[40] Turkey,[43] and the United States.[41,44] Various patient populations were considered, including those in primary care clinics, sleep clinics, the veteran population, and patients with cardiac disease. The patients included in these studies were mostly men (50% or greater in the majority of studies) with suspected OSA; they were overweight or obese, and middle-aged. Overall, the Berlin Questionnaire produced a large number of false negative results, thereby limiting its utility as an instrument to diagnose patients with OSA. Specifically, when assessing the performance of the Berlin Questionnaire in identification of subjects with an AHI cutoff of ≥ 5, the pooled sensitivity was 0.76 (95% CI: 0.72 to 0.80), while the pooled specificity was 0.45 (95% CI: 0.34 to 0.56) (see supplemental material, **Figure S1**). Assuming a prevalence of 87% in a high-risk population, the result was an unacceptably high number of false negative results of 209 per 1,000 patients (95% CI: 174 to 244) (see supplemental material, **Table S2**). Furthermore, the questionnaire had suboptimal accuracy, ranging from 56% to 70%; accuracy became progressively more compromised with consideration of higher OSA severity thresholds (see supplemental material, **Table S2** through **Table S4** and **Figure S1** through **Figure S6**).

Five studies evaluated the performance of the Berlin Questionnaire against HSATs.[53–57] When using an AHI cutoff of ≥ 15, the pooled estimate for sensitivity was 0.76 (95% CI: 0.64 to 0.85); specificity was 0.44 (95% CI: 0.30 to 0.58) and accuracy was 67%. When using an AHI cutoff of ≥ 5 and assuming a prevalence of 87% in a high-risk population, the number of false negative results was 531 per 1,000 patients (95% CI: 357 to 679) (see supplemental material, **Figure S5** through **Figure S7**, **Table S5** and **Table S6**).

The quality of evidence for the use of the Berlin Questionnaire was low after being downgraded due to either heterogeneity, indirectness, or imprecision.

Epworth Sleepiness Scale: The Epworth Sleepiness Scale (ESS) is a self-reported questionnaire involving eight questions to assess the propensity for daytime sleepiness or dozing.[58] Our review identified seven studies that evaluated the performance of the ESS against PSG in the identification of patients with OSA. These studies were conducted in China, Brazil, Croatia, Turkey and the United States, thus reflecting a wide geographic sampling.[38,43,50,51,59–61] Participants were those suspected of OSA and included mainly male, middle-aged and overweight or obese individuals. The overall results indicate that the ESS had a large number of false negative results limiting its utility for the diagnosis of OSA. When considering an AHI of ≥ 5, the ESS revealed a range of sensitivity of 0.27–0.72 and specificity of 0.50–0.76 (see supplemental material, **Table S8**). The ESS demonstrated an accuracy ranging from 51% to 59% for the AHI ≥ 5 cutoff. Therefore, the ESS had a high number of false negative results (range of 244 to 635 per 1,000 patients; assuming a prevalence of 87%). When considering a cutoff of AHI ≥ 15 and assuming a prevalence of 64% in high-risk patients, the number of false negative results increased and ranged from 269 to 506 per 1,000 patients (see supplemental material, **Table S9**). Findings from one study, comparing the performance of the ESS against HSATs for identification of patients with OSA, showed low sensitivity of 0.36 (95% CI: 0.19 to 0.57) and higher specificity of 0.77 (95% CI: 0.66 to 0.86)[57] (see supplemental material, **Table S11**).

The quality of evidence for the use of the ESS ranged from low to high across different AHI cutoffs after being downgraded to due to heterogeneity, indirectness, or imprecision The TF determined that the overall quality of evidence across AHI cutoffs was low.

STOP-BANG Questionnaire: The STOP-BANG questionnaire is an OSA screening tool consisting of four yes/no questions and four clinical attributes.[62] We identified ten studies, involving primarily middle-aged, obese males suspected of OSA that evaluated the performance of STOP-BANG questionnaire against PSG in the identification patients with OSA.[42,49–52,63–67] The overall findings reveal that the STOP-BANG questionnaire had high sensitivity, but low specificity for the detection of OSA. These findings became more pronounced when higher levels of OSA category cutoffs were considered. The number of potential false negative diagnostic results limits use of the STOP-BANG as an instrument to diagnose individual patients with OSA. Specifically, when considering an AHI ≥ 5 and assuming a prevalence of 87% in high-risk patients, the sensitivity in the studies was 0.93

(95% CI: 0.90 to 0.95), but specificity was 0.36 (95% CI: 0.29 to 0.44) with a range of accuracy of 52 to 53%. The number of false negatives when compared against PSG was 61 per 1,000 patients (95% CI: 43 to 87), assuming a prevalence of 87% (see supplemental material, **Figure S8** and **Table S12**). The sensitivity further improved and specificity was further compromised when progressively higher level of AHI cutoffs were considered (see supplemental material, **Figure S10** through **Figure S12**, **Table S13** and **Table S14**). The sensitivity and specificity of the STOP-BANG was similar when compared against HSAT[55,68] (see supplemental material, **Table S15** through **Table S17**), or against PSG or HSAT[62,68] (see supplemental material, **Tables S18** through **Table S20**).

The quality of evidence for the use of the STOP-BANG questionnaire ranged from low to high across different AHI cutoffs was after being downgraded due to either indirectness, inconsistency, or imprecision. The TF determined that the overall quality of evidence across AHI cutoffs was moderate.

STOP Questionnaire: Our review identified five studies that evaluated the diagnostic performance of the STOP questionnaire against PSG.[49–51,67,69] The STOP questionnaire showed moderate to high sensitivity, low specificity, and moderate accuracy (see supplemental material, **Figure S14** and **Table S21** through **Table S23**). When considering an AHI ≥ 5, the sensitivity was 0.88 (95% CI: 0.77 to 0.94), the specificity was 0.33 (95% CI: 0.18 to 0.52), and the accuracy in a high-risk population ranged from 74% to 86%. Assuming a prevalence of 87%, the number of false negatives was 104 per 1,000 patients (95% CI: 52 to 200) (see supplemental material, **Table S21**). When considering an AHI cutoff of ≥ 15, the sensitivity ranged from 0.62–0.98, the specificity ranged from 0.10–0.63, and the accuracy in a high-risk population ranged from 60% to 79%. Assuming a prevalence of 64% in a high-risk population, the number of false negatives ranged from 13 to 243 per 1,000 patients (see supplemental material, **Table S22**). When considering an AHI cutoff of ≥ 30, the sensitivity ranged from 0.91–0.97, the specificity ranged from 0.11–0.36, and the accuracy in a high-risk population ranged from 48% to 49%. Assuming a prevalence of 36% in a high-risk population, the number of false negatives ranged from 11 to 32 per 1,000 patients (see supplemental material, **Table S23**).

The quality of evidence for the use of the STOP questionnaire ranged from low to moderate across different diagnostic cutoffs and risk groups after being downgraded due to heterogeneity or imprecision. The TF determined that the overall quality of evidence across AHI cutoffs was low.

Morphometric Models: Our review identified two studies that used morphometric models to predict OSA that was confirmed using sleep study data.[70,71] In a group of hypertensive patients, a multivariable apnea prediction score that combined symptoms, body mass index, age and sex was used to assess OSA risk.[70] In another study involving primarily middle-aged males, those with OSA were compared to those without OSA by using a morphometric clinical prediction formula incorporating measures of craniofacial anatomy (e.g., palatal height, maxillary and mandibular intermolar distances).[71] While these studies demonstrate relatively high sensitivity (range of 0.88–0.98) to predict AHI ≥ 5, the specificity was quite low (range of 0.11–0.31) (see supplemental material, **Table S24**). When considering adjusted neck circumference in both the hypertensive and chronic kidney disease populations, there are similar findings of relatively high sensitivity, but poor specificity, with improvements in specificity using higher AHI cutoffs[55,70] (see supplemental material, **Table S25** and **Table S26**).

The quality of evidence for the use of morphometric models and adjusted neck circumference was moderate after being downgraded due to imprecision.

Multivariable Apnea Prediction Questionnaire: The performance of the Multivariable Apnea Prediction (MAP) questionnaire has been evaluated against PSG in those with suspected OSA,[35,72–74] a sample of hypertensive patients,[70] and also a sample of older adults[75] with findings of lower levels of specificity and high numbers of false positive results (see supplemental material, **Table S27** and **Table S28**).

The quality of evidence for the use of the MAP questionnaire was judged moderate; it was downgraded due to imprecision.

Clinical Prediction Models: Four studies evaluated the performance of clinical prediction models against PSG,[61,76–78] and three studies[75,79,80] evaluated these models against HSAT. Two of the studies compared respiratory parameters against PSG: a study involving a Chinese cohort that evaluated snoring while sitting[76] and another single study assessing respiratory conductance and oximetry.[78] Results demonstrated a sensitivity ranging from 0.33–0.90, and a specificity ranging from 0.50–1.00 using an AHI cutoff of ≥ 5. Other studies compared clinical prediction rules including age, waist circumference, ESS score and minimum oxygen saturation, and another evaluated gender, nocturnal choking, snoring and body mass index against PSG; these reported reasonably high sensitivity (range of 0.72–0.94) and specificity (range of 0.75–0.91) considering different AHI thresholds.[61,77] Clinical prediction rules have been evaluated against HSAT in select populations, i.e., the elderly,[75] bariatric surgery candidates,[79] and commercial drivers.[80] These studies reported sensitivities ranging from 0.76–0.97 and specificities ranging from 0.19–0.75 using an AHI cutoff of ≥ 30[75,79,80] (see supplemental material, **Table S29** through **Table S31**).

The quality of evidence for the use of clinical prediction models ranged from moderate to high across the different AHI cutoffs after being downgraded due to imprecision. The TF determined that the overall quality of evidence across AHI cutoffs was moderate.

Other OSA Prediction Tools: Our literature review identified other OSA prediction tools, including the OSA50, the clinical decision support system, the OSAS score, and the Kushida Index. The OSA50 questionnaire involves four components including age ≥ 50, snoring, witnessed apneas and waist circumference.[81] A study involving Turkish bus drivers[82] and a validation study for the OSA50 in the primary care setting[81] showed a sensitivity ranging from 0.49–0.98 and a specificity of 0.82 in both studies (see supplemental material, **Table S32** and **Table S33**).

The performance of a hand-held clinical decision support system (assessing sleep behavior, breathing during sleep and daytime functioning) against PSG was evaluated in a study of veterans with ischemic heart disease. The system showed a high sensitivity of 0.98 (95% CI: 0.92 to 1.00) and a high specificity of 0.87 (95% CI: 0.66 to 0.97)[41] (see supplemental material, **Table S34**).

The OSAS score involves assessment of the Friedman tongue position, tonsil size, and body mass index. In a sample of individuals suspected to have OSA, the sensitivity of the OSAS score was 0.86 (95% CI: 0.80 to 0.91) against PSG at an AHI > 5 cut = off, however; specificity was lower at 0.47 (95% CI: 0.34 to 0.56) with a high number of false positives in the low-risk group[39] (see supplemental material, **Table S35**).

One study evaluating the performance of the Kushida Index against PSG showed a high sensitivity of 0.98 (95% CI: 0.95 to 0.99) and high specificity of 1.00 (95% CI: 0.92 to 1.00) to detect AHI ≥ 5 (see supplemental material, **Table S36**).

The quality of the evidence for other prediction tools ranged from low to high across different tools, diagnostic cutoffs, and risk groups after being downgraded due to imprecision and indirectness.

Overall Quality of Evidence: The quality of evidence for specific clinical tools, questionnaires and prediction algorithms ranged from very low to high after being downgraded due to imprecision, indirectness, and heterogeneity. However, due to the heterogeneity of the tools, questionnaires and prediction algorithms described above combined with the low likelihood that future research would result in a change of the accuracy of these tools, the TF determined that the overall quality of evidence for the recommendation against using clinical tools, questionnaires or predictive tools was moderate.

Benefits versus Harms: These clinical tools, questionnaires and prediction algorithms carry the risk of not capturing the diagnosis of OSA when indeed OSA is present. Given the downstream effects of false negative diagnostic results, this would translate into high levels of OSA-related decrements in QOL, morbidity, and mortality due to undiagnosed and untreated OSA. On the other hand, false positive results would result in unnecessary testing and treatment for sleep apnea. Therefore, the TF determined that the potential harms outweigh the potential benefits of using clinical tools, questionnaires and prediction algorithms alone to diagnose OSA.

Patients' Values and Preferences: Evaluation with clinical tools, questionnaires or prediction algorithms may be less burdensome to the patient and physician, when compared to HSAT or PSG. However, this must be weighed against their low levels of accuracy and the likelihood of misdiagnosis. In contrast, PSG and HSAT require more resources and create more burden for the patient and physician; however, they provide greater value in terms of higher diagnostic accuracy and therefore a higher likelihood that patients will receive appropriate treatment. Based on its clinical judgment, the TF determined that the vast majority of patients would not favor the use of clinical questionnaires or prediction tools alone for the diagnosis of OSA.

Home Sleep Apnea Testing for the Diagnosis of Obstructive Sleep Apnea in Adults

Recommendation 2: We recommend that polysomnography, or home sleep apnea testing with a technically adequate device, be used for the diagnosis of OSA in uncomplicated adult patients presenting with signs and symptoms that indicate an increased risk of moderate to severe OSA. (STRONG)

Recommendation 3: We recommend that if a single home sleep apnea test is negative, inconclusive or technically inadequate, polysomnography be performed for the diagnosis of OSA. (STRONG)

Remarks: The following remarks are based on specifications used by studies that support these recommendation statements:

An uncomplicated patient is defined by the absence of:

1. *Conditions that place the patient at increased risk of non-obstructive sleep-disordered breathing (e.g., central sleep apnea, hypoventilation and sleep related hypoxemia). Examples of these conditions include significant cardiopulmonary disease, potential respiratory muscle weakness due to neuromuscular conditions, history of stroke and chronic opiate medication use.*
2. *Concern for significant non-respiratory sleep disorder(s) that require evaluation (e.g., disorders of central hypersomnolence, parasomnias, sleep related movement disorders) or interfere with accuracy of HSAT (e.g., severe insomnia).*
3. *Environmental or personal factors that preclude the adequate acquisition and interpretation of data from HSAT.*

An increased risk of moderate to severe OSA is indicated by the presence of excessive daytime sleepiness and at least two of the following three criteria: habitual loud snoring, witnessed apnea or gasping or choking, or diagnosed hypertension.

HSAT is to be administered by an accredited sleep center under the supervision of a board-certified sleep medicine physician, or a board-eligible sleep medicine provider.

A single HSAT recording is conducted over at least one night.

A technically adequate HSAT device incorporates a minimum of the following sensors: nasal pressure, chest and abdominal respiratory inductance plethysmography, and oximetry; or else PAT with oximetry and actigraphy. For additional information regarding HSAT sensor requirements, refer to The AASM Manual for the Scoring of Sleep and Associated Events.[24]

A technically adequate diagnostic test includes a minimum of 4 hours of technically adequate oximetry and flow data,

obtained during a recording attempt that encompasses the habitual sleep period.

Summary

Twenty-six validation studies suggested potential for clinically significant diagnostic misclassification using HSAT when compared against PSG. However, seven RCTs failed to find, after PAP initiation, that patient-reported sleepiness, QOL, and continuous positive airway pressure (CPAP) adherence were significantly different when HSAT was used. The RCTs used HSAT in the context of a management pathway that required PSG confirmation for patients in whom HSAT did not establish an OSA diagnosis, under the conditions specified in the above remarks. The overall quality of evidence was moderate due to imprecision, inconsistency, or indirectness. Therefore, in this context, either PSG or HSAT is recommended for the diagnosis of OSA. However, two other considerations are also key. First, a clinician's choice of study type for a particular patient should be guided by clinical judgment and incorporate consideration of patient preferences. Second, it is essential to note that the need for diagnosis of OSA is not limited to uncomplicated patients who are at increased risk for moderate to severe OSA. In patients who do not meet these criteria, but in whom there is a concern for OSA based on a comprehensive sleep evaluation, PSG is recommended.

HSAT is less sensitive than PSG in detection of OSA and a false negative test could result in harm to the patient due to denial of a beneficial therapy. For this reason, the majority of RCTs that were judged most generalizable to clinical practice required that PSG eventually be performed when HSAT did not confirm a diagnosis of OSA.[83–85] Performing a repeat HSAT is not recommended when an initial test is negative, inconclusive or technically inadequate, due to the higher likelihood that a second test will also be negative, inconclusive or technically inadequate. There is also an increased risk that the patient will not complete the diagnostic process prior to a definitive diagnosis. Therefore, after a single negative, inconclusive or technically inadequate HSAT result, performance of a PSG is strongly recommended.

Finally, use of HSAT to diagnose OSA has been shown to provide adequate clinical outcomes and efficiency of care when performed with adequate clinical and technical expertise, using specific types of HSAT devices, in an appropriate patient population, and within an appropriate management pathway. Use of HSAT in other contexts may not provide similar benefit, and therefore the recommendations for the use of HSAT are limited. On the other hand, unstudied or understudied contexts could exist in which HSAT may provide benefit to a patient.

The TF determined that the benefits of HSAT compared to PSG balanced the potential harms, when used in the patient populations and under the conditions specified in the above remarks and recommendations. Based on clinical judgment, the TF determined that many patients would value the convenience and potential cost savings of HSAT, while other patients would prefer the superior accuracy of PSG, the increased probability that only one diagnostic test will be needed, and the potential ability to titrate positive airway pressure if indicated.

Discussion

The formulation of these recommendation statements was guided by evidence from twenty-six validation studies that evaluated the diagnostic accuracy of HSAT against PSG,[35,53,62,67,81,86–106] as well as seven RCTs that compared clinical outcomes from management pathways.[83–85,107–110] Four of these RCTs were determined to be most relevant to clinical practice, as they did not require oximetry testing as a criterion for inclusion and used conventional methods for determination of PAP pressures (i.e., APAP or attended titration).[83–85,110] This subset of studies will be referred to as "RCTs most generalizable to clinical practice" for the remainder of this discussion section.

ACCURACY: The following paragraphs are organized by type of HSAT device and components or combinations of components, as described in the literature.

A total of twenty-six validation studies were identified that reported accuracy outcomes. The data from these validation studies are summarized in the supplemental material, **Table S37** through **Table S58**. In two studies that evaluated the performance of Type 2 HSAT devices against PSG,[67,86] when using an AHI ≥ 5 cutoff, accuracy in a high-risk population (assuming a prevalence of 87%) ranged from 84% to 91%. Using a cutoff of AHI ≥ 15, the accuracy of these devices was 88% in a high-risk group (see supplemental material, **Table S37** and **Table S38**).[67,86]

Seven studies evaluated the performance of Type 3 HSAT devices against PSG, but the AHI cutoffs employed varied across studies, resulting in sub-grouping by AHI cutoffs for our analyses.[87–93] When using an AHI ≥ 5 cutoff, accuracy in a high-risk population (assuming a prevalence of 87%) ranged from 84% to 91%, whereas in a low-risk population (assuming a prevalence of 55%) accuracy ranged from 70% to 78% based on the seven studies (see supplemental material, **Table S39**). Using a cutoff of AHI ≥ 15, the accuracy of these devices in a high-risk population ranged from 65% to 91%, based on six studies[87,89–92,94] (see supplemental material, **Table S40**). Using a cutoff of AHI ≥ 30, the accuracy of the devices in the high-risk population was 88% (95% CI: 81% to 94%), based on five studies (see supplemental material, **Table S41**).

Five studies evaluated the performance of 2–3 channel HSAT devices against PSG. In a high-risk population using cutoffs of AHI ≥ 5,[95–97] AHI ≥ 15,[95–99] and AHI ≥ 30,[96,97] accuracy ranged from 81% to 93%, 72% to 87%, and 71% to 90%, respectively. Using the same cutoffs in a low-risk population, accuracy ranged from 77% to 88%, 68% to 95%, and 88% to 91%, respectively (see supplemental material, **Table S42** through **Table S44**). When the performance of 2–3 channel HSAT was evaluated against unattended in-home PSG, using a cutoff of AHI ≥ 15, accuracy in a high-risk population was 86% (95% CI: 76% to 93%);[53] using a cutoff of AHI ≥ 30, accuracy ranged from 83% to 91% (see supplemental material, **Table S45** and **Table S46**).[53,81]

Six studies evaluated the performance of single channel HSAT against attended or unattended PSG (see supplemental

material, **Table S47** through **Table S50**, and **Table S51** through **Table S53**, respectively).[73,100–103,111]

A single study evaluated the performance of oximetry against unattended in-home PSG.[111] Using a cutoff of AHI ≥ 5, accuracy was 73% (95% CI: 68 to 78%) in a high-risk population, and 79% (95% CI: 74 to 84%) in a low-risk population. Using oximetry to identify OSA at an AHI ≥ 5 cutoff, and assuming a prevalence of 87% in a high-risk population, the findings of the study[111] would result in an estimated average of 274 misdiagnosed patients out of 1,000 tested, and 210 misdiagnosed patients out of 1,000 tested in a low-risk group (assuming a prevalence of 55%). Using a cutoff of AHI ≥ 15 and AHI ≥ 30, oximetry has an accuracy of 86% (95% CI: 83 to 91%) and 74% (95% CI: 71 to 76%) in a high-risk population, and an accuracy of 80% (95% CI: 75 to 84%) and 63% (95% CI: 59 to 67%) in a low-risk population, respectively (see supplemental material, **Table S51** through **Table S53**).

A single study evaluated the performance of PAT, oximetry, and actigraphy against simultaneous unattended in-home PSG and reported a sensitivity of 0.88 (95% CI: 0.47 to 1.00), specificity of 0.87 (95% CI: 0.66 to 0.97) and accuracy of 88% (95% CI: 50 to 100%) in high-risk patients using a cutoff of AHI ≥ 5.[104] These findings would result in 121 misdiagnosed patients out of 1,000 tested in a high-risk population (based on a prevalence of 87%), and 125 misdiagnosed patients out of 1,000 tested in a low-risk population (based on a prevalence of 55%) (see supplemental material, **Table S54**).[104] Two cross-over studies randomized patients to home-based PAT, and in-laboratory simultaneous PSG and PAT.[105,106] For comparison to in-laboratory PSG, only the home-based PAT data were used for this recommendation. A single study that evaluated the performance of the PAT device in the home against in-laboratory PSG using a cutoff of AHI ≥ 5,[106] reported a specificity of 0.43 (95% CI: 0.22 to 0.66). When two studies evaluated the home-based PAT device against in-laboratory PSG at an AHI cutoff of ≥ 15, specificity ranged from 0.77 to 1.00 and sensitivity ranged from 0.92 to 0.96.[105,106] A single study evaluated the PAT device at an AHI cutoff of ≥ 30, and reported a specificity of 0.82 (95% CI: 0.57 to 0.96) and sensitivity of 0.92 (95% CI: 0.62 to 1.00)[105] (see supplemental material, **Table S55** through **Table S57**).

The quality of evidence for diagnostic accuracy was downgraded due to indirectness, imprecision, or inconsistency. The quality ranged from low to high based on different tools and algorithms, diagnostic cutoffs, and risk groups.

The potential consequences for patients classified in true and false positive or negative categories are summarized in the supplemental material, **Table S1**. The TF concluded that the numbers of patients potentially misclassified by HSAT was high enough to be of clinical concern, particularly when tests were inconclusive or negative. In a population that has increased risk of moderate to severe OSA, both the increased likelihood of false negatives and the significant impact of missed diagnoses on patient outcomes can cause significant harm. This reasoning supports required use of a diagnostic test with higher sensitivity (PSG) in this population if HSAT provides a negative or non-diagnostic result.

Clinical Outcomes Assessment: The TF concluded that evaluating the impact of diagnostic accuracy on clinical outcomes is complicated by a number of factors that can cause discordance between tests, including night-to-night variability and inconsistent definitions of respiratory events (e.g., hypopneas) between HSAT and PSG. In addition, there is uncertainty regarding clinical outcomes for patients misclassified by HSAT.

For these reasons, studies that compared clinical outcomes in patients randomized to management pathways based on PSG or HSAT diagnostic assessment, within the same research protocol, provide the best opportunity to assess the acceptability of clinical outcomes using HSAT.

Subjective Sleepiness: A meta-analysis of seven RCTs compared changes in patient-reported sleepiness, using the ESS, in patients diagnosed by HSAT or PSG, followed by PAP titration (see supplemental material, **Figure S18**).[83–85,107–110] The meta-analysis showed a clinically and statistically insignificant difference of 0.38 points (95% CI: −1.07 to 0.32 points) greater improvement in patients randomized to the HSAT pathway versus the attended PSG pathway. This difference indicates that subjective sleepiness is similarly improved in patients who initiate PAP treatment based on diagnosis using either HSAT or PSG. The quality of evidence for subjective sleepiness was high.

Quality of Life: Six RCTs, using various validated instruments (i.e., FOSQ, SAQLI, and SF-36), compared QOL in patients diagnosed by HSAT or PSG, followed by PAP titration.[84,85,107–110] Meta-analysis demonstrated differences in pooled effects between pathways that were not significant (see supplemental material, **Figure S19** through **Figure S23**, and **Table S58**). The quality of evidence ranged from moderate to high based on the measure used to assess QOL. The quality of evidence for the SF-36 physical and mental summary scores was downgraded due to imprecision. The TF considered the overall quality of evidence for QOL to be high as FOSQ and SAQLI measures of QOL were considered more critical for decision-making than the SF-36 measures.

CPAP Adherence: Six RCTs evaluated CPAP adherence (mean hours of use per night); meta-analysis found no significant difference between the two assessment pathways (see supplemental material, **Figure S24**).[83–85,108–110] When determining adherence by number of nights with greater than 4 hours of use, meta-analysis of five RCTs found a clinically insignificant trend towards increased CPAP adherence in the HSAT arm versus the PSG arm (see supplemental material, **Figure S25**).[83–85,107,110] The quality of evidence for CPAP adherence was moderate to high across different AHI cutoffs after being downgraded due to imprecision. The TF determined that the overall quality of evidence across AHI cutoffs was high.

Failure to Complete Diagnostic Algorithm: Among the four RCTs most generalizable to clinical practice, three studies[83–85] required use of PSG if HSAT was inconclusive (did not provide adequate data or showed a low AHI after 1 or 2 unsuccessful attempts) and after 1 or 2 failed

APAP trials (e.g., insufficient use, elevated residual AHI, persistent large leak). Based on data reported by a multicenter RCT there was concern regarding risk of non-completion of diagnostic testing when initial HSAT did not provide a definitive result. Rosen et al. 2012[84] reported that 30% (10/33) of subjects with technically inadequate HSATs and 16% (14/88) of subjects with low AHI on HSAT failed to proceed per protocol to PSG. There was also evidence indicating reduced effectiveness of repeated HSAT attempts for technical failures: 82% (147/180) of initial HSAT attempts were technically acceptable, whereas only 60% (12/20) of second attempts resulted in a technically acceptable study. Although failure to complete the diagnostic algorithm was not originally considered a critical outcome, the TF ultimately determined that it was critical for decisions regarding follow-up for inconclusive HSAT attempts. The quality of evidence regarding performance of PSG after a single inconclusive HSAT (versus multiple attempts) was low.

Overall Quality of Evidence: The TF determined that the critical outcome for diagnostic accuracy assessment was the number of false negative results. The quality of evidence for accuracy was downgraded to moderate due to imprecision, inconsistency, or indirectness. The quality of evidence for the clinical outcomes of sleepiness, quality of life, and CPAP adherence was high. Depression and cardiovascular outcomes were also considered critical outcomes; however, evidence for these outcomes was not available. Therefore, the overall quality of evidence for recommendation 2 is moderate.

In addition to accuracy and clinical outcomes, the TF determined that failure to complete the diagnostic algorithm was a critical outcome for repeat testing after a negative, inconclusive or technically inadequate HSAT. The quality of evidence for performing PSG after a single inconclusive HSAT was determined to be low, as only one study addressed this outcome. Therefore, the overall quality of evidence for recommendation 3 is low.

Resource Use: Though a single night of HSAT is less resource-intensive than a single night of PSG, the relative cost-effectiveness of management pathways that incorporate each of these diagnostic strategies is unclear. Economic analyses have compared the cost-effectiveness of management pathways that incorporate diagnostic strategies using HSAT or PSG.[112–114] All have concluded that PSG is the preferred diagnostic strategy from an economic perspective for adults suspected to have moderate to severe OSA. An important factor in these analyses is the favorable cost-effectiveness of OSA treatment in patients with moderate to severe OSA, particularly when longer time horizons are considered. As a result, diagnostic strategies that lead to increased false negatives, and leave patients untreated, or increase false positives, and unnecessarily treat patients, have less favorable cost-effectiveness. It is important to note that these economic analyses are susceptible to error because of imprecision in modelling of management pathways and limitations in the quality of data available to estimate parameters. The impact of errors can be magnified when extrapolated over long time horizons.

Relative cost-effectiveness of management pathways that use HSAT or PSG for diagnosis can be assessed in the context of a RCT, if resource utilization is measured. Among the four RCTs most generalizable to clinical practice,[83–85,110] only one provided this information.[84] The study reported that in-trial costs were 25% less in the home-arm than the in-laboratory-arm.[84] These estimates were based on the Medicare Fee Schedule for the various study procedures, including office visits and diagnostic testing, and take into account the need to repeat studies.[84] A subsequent cost minimization analysis of this RCT also considered costs from a provider perspective.[115] While provider costs (capital, labor, overhead) were generally less for the home program, this was not true for all modelled scenarios. The provider perspective highlighted the large number of cost components necessary to ensure high quality home-based OSA management, which narrowed the cost difference relative to lab management.

The available studies indicate that the potential cost advantages of HSAT over PSG are not as high as reflected by the cost difference of a single night of testing. Even when HSAT is used in appropriate populations and conditions, additional HSAT and PSG are needed for patients with technically inadequate or inconclusive studies, in order to achieve an accurate diagnosis. In addition, if a home management pathway is used in a manner that results in reduced effectiveness relative to PSG, use of HSAT could in fact be less cost effective than using PSG. Examples of this include use in patient populations with predominantly mild OSA in which there are a higher proportion of negative or indeterminate HSAT results that require follow-up PSG, or use in patients at risk for non-obstructive sleep-related breathing disorders that may not be accurately diagnosed with HSAT. The TF determined that if HSAT is used in the recommended context and management pathway, it would be more cost-effective than if it is used outside this framework.

Benefits versus Harms: Use of HSAT may provide potential benefits to patients with suspected OSA. Such benefits could include convenience, comfort, increased access to testing, and decreased cost. HSAT can be performed in the home environment with fewer attached sensors during sleep. The availability of HSAT for diagnosis may improve access to diagnostic testing in resource-limited settings, or when the patient is unable to leave the home or healthcare setting for testing. In addition, HSAT may be less costly when used appropriately. These benefits must be weighed against the potential for harm. Harms could result from the need for additional diagnostic testing among patients with technically inadequate or inconclusive HSAT findings, or from misdiagnosis and subsequent inappropriate therapy or lack of therapy. As summarized above, the use of HSAT has not been demonstrated to provide inferior clinical benefit, compared to PSG when used in the appropriate context. Therefore, the TF determined that if HSAT is used in the context described in the recommendations and remarks, the risk of harm is minimized and the probability of potential economic benefits increased.

The TF was concerned that, in clinical practice (in contrast to the RCT setting) there would be higher levels of drop out from diagnostic testing, among patients with initial study

attempts that did not result in diagnoses of OSA. In particular, there was concern that patients with false negative HSAT results may not complete additional testing after learning of a negative result, despite the presence of symptoms of OSA. In addition, as described above, HSAT is less accurate than PSG and more likely to result in false negative results. For these reasons, the TF recommends that if the initial HSAT shows a negative or inconclusive result, PSG, rather than a second HSAT, should be performed. There are similar concerns that, following a technically inadequate HSAT, repeat HSAT may be associated with a higher rate of technical failure on the second study, and with increased risk of drop out from the diagnostic process. Therefore, the TF also recommends that if the initial HSAT is technically inadequate, PSG rather than a second HSAT should be performed. On the other hand, the TF recognizes that there may be specific circumstances in which repeat HSAT is appropriate after an initial failed HSAT. These circumstances would include cases in which both of the following are present: the clinician determines that there is a high likelihood of successful recording on a second attempt, and the patient expresses a preference for this approach.

The TF recognizes that HSAT may have value to patients in some contexts beyond what is covered by these recommendations, but has limited the recommendations to apply to situations where there is sufficient evidence to guide evaluation of benefits versus harms.

Patients' Values and Preferences: Individual patient preference for PSG or HSAT will differ depending on circumstances and values. In one of the four RCTs most generalizable to clinical practice, both HSAT and PSG were performed for each patient, and 76% preferred HSAT.[110] This means that a significant percentage (24%) still preferred PSG. Unfortunately, there is insufficient data about diagnostic testing preferences in clinical practice, where preferences may differ from what is seen in the RCT setting. The availability of different options for diagnosis may increase satisfaction, if patient preferences are included in the process of choosing the diagnostic test type. If HSAT is used, the TF determined that patients would value accurate diagnosis, good clinical outcomes, and increased convenience. Based on their clinical judgment, the TF also determined that patients would prefer not having a repeat HSAT if the initial test result is negative, as repeated HSAT would be less likely to produce a definitive result and would unnecessarily inconvenience the patient. In this situation, proceeding directly to PSG, which has greater sensitivity to detect OSA, would be preferred by most patients. The TF also determined that most patients would prefer not to have a repeat HSAT if the initial test was technically inadequate, to avoid inconvenience, but that some patients may desire this option, in specific cases in which there was high likelihood of an adequate result with repeat testing.

Special Considerations: The following sections describe special considerations when using HSAT for the diagnosis of OSA. They provide additional support for, and explanation of the Remarks, and are based on specifications used by studies that support the recommendation statements.

Clinical Population: A review of RCTs that met inclusion criteria indicated that the following criteria should be used to establish the presence of increased risk of moderate to severe OSA and to determine if HSAT use is reasonable: excessive daytime sleepiness occurring on most days, AND the presence of at least two of the following three criteria: habitual loud snoring; witnessed apnea or gasping or choking; or diagnosed hypertension. Among the four RCTs most generalizable to clinical practice, two of the four studies[83,84] required ESS > 12 as an entry criterion: One[110] required at least two out of three criteria (i.e., sleepiness (ESS > 10), witnessed apnea, snoring) for participation; and one, which was performed in a Veteran's Administration population, did not specify any specific entry criteria besides suspected OSA (though the average ESS for participants was elevated at > 12 and 95% were men).[83] In the latter study, 9.9% of individuals in the PSG arm were found to have AHI < 5.[83] In addition to sleepiness, at least two studies in this subset had specific inclusion criteria such as snoring, witnessed apnea, gasping or choking at night, or hypertension.[83,85] One study incorporated neck circumference in the determination of high risk of OSA.[84]

Excluded Patient Populations: Three of the four RCTs most generalizable to clinical practice excluded patients with significant cardiopulmonary disease and other significant sleep disorders.[83,84,110] Two studies excluded patients taking opioids, having uncontrolled psychiatric disorder, neuromuscular disease, and patients with significant safety-related issues related to driving or work. Other notable exclusion criteria, specified by at least one of the studies, included lack of an appropriate living situation, pregnancy, and alcohol abuse. The single study that did not mention exclusion criteria noted that 3 of 148 individuals in the HSAT arm were diagnosed with CSA and 4 of 148 individuals required supplemental oxygen or bi-level PAP and exited the study.[85] In the PSG arm of the study, 6 of 148 individuals were diagnosed with CSA and 12 of 148 required supplemental oxygen or bi-level PAP. Studies outside the four RCTs most generalizable to clinical practice had similar inclusion/exclusion criteria.

Therefore, based on information from three of the four RCTs most generalizable to clinical practice that specified exclusion criteria, and for the reasons discussed above in Resource Use, Benefits and Harms, and Patients' Values and Preferences sections, the TF determined that HSAT should be used in an uncomplicated clinical population. This is defined as the absence of significant cardiopulmonary disease (e.g., heart failure, chronic obstructive pulmonary disease [COPD]), potential respiratory muscle weakness due to neuromuscular conditions, chronic opiate medication use, history of stroke, concern for a significant sleep disorder other than OSA (e.g., CSA, parasomnia, narcolepsy, severe insomnia), and environmental or personal factors that preclude the adequate acquisition and interpretation of data from HSAT.[83,84,110]

Follow-Up: Based on information from the four RCTs most generalizable to clinical practice,[83–85,110] the TF determined that HSAT should be used in the context of an OSA management

pathway that incorporates a PAP therapy initiation protocol for APAP or PSG titration, early follow-up after initiation of therapy, and PSG titration studies for patients failing APAP therapy. All RCTs incorporated early follow-up of APAP titration (within 2–7 days after HSAT) by skilled technical staff.[83–85,110] As described above, the recommendation for using HSAT to diagnose OSA is based on clinically significant improvements in clinical outcomes. Therefore, the TF determined that HSAT should be used in the context of an OSA management pathway that incorporates a PAP therapy initiation protocol and early follow-up after initiation of therapy.

Clinical Expertise: All four RCTs that were most generalizable to clinical practice administered HSAT at academic or tertiary sleep centers with highly skilled sleep medicine providers and technical staff.[83–85,110] HSAT recordings were reviewed by a sleep medicine specialist. One RCT that was not included in this subset (because an overnight oximetry was used as entry criteria) used a simplified nurse-led model of care involving nurse specialists experienced in management of sleep disorders (mean of 8.3 years of experience with CPAP therapy). Therefore, the TF determined that HSAT should be administered by an accredited sleep center under the supervision of a board-certified sleep medicine physician, or a physician who has completed a sleep fellowship, but is awaiting the next opportunity to take the board examination.

Home Sleep Apnea Testing Device: Among the four RCTs that were most generalizable to clinical practice, three used conventional Type 3 devices (nasal pressure, thoracic and abdominal excursion using RIP technology, oxygen saturation, EKG, body position, and oral thermistor in some cases),[84,85,110] and one used a 4-channel device[83] based on PAT with three additional channels (heart rate, pulse oximetry, and actigraphy). The TF determined that testing should be performed using these types of HSAT devices that have been demonstrated to be technically adequate. Additional guidance on technical specifications regarding HSAT is provided in The AASM Manual for the Scoring of Sleep and Associated Events.[24]

Recording Time: In the four RCTs most generalizable to clinical practice, the minimum requirement for an acceptable study was 4 hours of adequate flow and oximetry signals.[83–85,110] Whereas one HSAT study[83] used PAT as a surrogate of flow, two studies recorded nasal pressure flow[85,110] and one study recorded thermistor in addition to nasal pressure flow.[84] The latter three studies also recorded thoracic and abdominal movements.[84,85,110] All of these studies showed at least equivalence of adherence to PAP therapy and functional improvement in the home versus in-laboratory management pathways.[84,85,110] Therefore, the TF determined that a protocol requirement of a minimum of 4 hours of good quality data from HSAT recording, during the habitual sleep period, is warranted to diagnose OSA.

Additionally, nine non-RCT validation studies reported minimum requirements for duration of acceptable signal quality.[35,53,54,81,86,88,93,96,116] The required signals and minimum durations included nasal pressure flow and oximetry for at least 3 hours[88,93,116] or 4 hours[53,81,86,96] and single-channel nasal airflow recording for a minimum of 3 hours[35] or only 2 hours.[54] The diagnostic accuracy of the cardiorespiratory devices compared against PSG for the detection of OSA at different AHI cutoff points was relatively high. One study reported a sensitivity and specificity of 0.88 and 0.84, respectively, for a HSAT AHI cutoff point of ≥ 9 events/h.[53] In a separate study, the sensitivity and specificity for unattended in-home PSG was 0.91 and 0.89 for an AHI cutoff of > 10 events/h, but 0.88 and 0.55, respectively for an AHI cutoff of > 5 events/h.[86] In another study, at an AHI cutoff of > 10 events/h, HSAT had a sensitivity of 0.87, and a specificity of 0.86.[88]

Overall, the body of evidence investigating the minimum number of hours of adequate data on HSAT required to accurately diagnose OSA is very limited. There are no data to suggest that fewer than 4 hours of technically adequate recording compromises the accuracy of test results, and there is no direct evidence on the impact of a minimum number of recording hours of HSAT on clinical outcomes. Based on available indirect evidence, the TF weighed the "risk" of undergoing less than the required duration of good quality HSAT with resultant false negative (or false positive) results, against the "benefit" of potentially increasing the accuracy by performing PSG. Performing PSG in the scenario of a "positive" diagnosis of OSA is less likely to alter clinical decision-making and may, in fact result in unnecessary delays in care with increased cost. Conversely, a "negative" HSAT, in the scenario of a high pretest probability of OSA, will justify PSG even when the test is of adequate quality and duration. The TF believes that the goals of establishing an accurate diagnosis, while minimizing patient inconvenience and cost, align with patient preferences.

Nights of Recording Time: The adequacy of a single night HSAT performed for the diagnosis of OSA in the context of an appropriate clinical population and management pathway is supported by published evidence. Our literature review only identified two studies relevant to the question of whether multiple nights of recording is superior to a single night.[35,73] These studies evaluated the performance of multiple nights (3) of single channel HSAT device (i.e., nasal pressure transducer or oximetry) to the first night of recording. Utilizing PSG as the reference, the studies found that recording over three consecutive nights may decrease the probability of insufficient data and marginally improve accuracy when compared against a single night of recording. However, the TF considered this evidence insufficient to establish the superiority of multiple-night HSAT protocol over a single-night HSAT protocol, as the studies only included a single channel recording and did not evaluate clinically meaningful outcomes or efficiency of care.

A single HSAT recording encompassing multiple nights may have potential advantages or drawbacks relative to only a single night of recording. For example, if multiple-night HSAT improved accuracy or resulted in fewer inconclusive or inadequate studies, patient outcomes or costs might improve. On

the other hand, the potential for multiple-night recordings to increase cost and patient inconvenience must be considered. Insufficient evidence exists to support routine performance of more than a single night's recording for HSAT.

Diagnosis of Obstructive Sleep Apnea in Adults with Comorbid Conditions

Recommendation 4: We recommend that polysomnography, rather than home sleep apnea testing, be used for the diagnosis of OSA in patients with significant cardiorespiratory disease, potential respiratory muscle weakness due to neuromuscular condition, awake hypoventilation or suspicion of sleep-related hypoventilation, chronic opioid medication use, history of stroke or severe insomnia. (STRONG)

Summary

This recommendation is based on the limited data available regarding the validity of HSAT in patients with significant cardiorespiratory disease, neuromuscular disease with respiratory impairment, suspicion of hypoventilation, opioid medication use, history of stroke, or severe insomnia. The overall quality of evidence was very low due to imprecision, indirectness, and risk of bias. The TF considered both the accuracy of HSAT for the detection of OSA, and the concurrent need to detect other forms of sleep-disordered breathing that can occur in these populations (e.g., CSA, hypoventilation and sleep-related hypoxemia). The likelihood of non-obstructive sleep-disordered breathing should be considered by the clinician, when determining which types and severity of cardiorespiratory diseases may be inappropriate for HSAT.

PSG is the gold standard method for the diagnosis of OSA and other forms of sleep-disordered breathing. HSAT has not been adequately validated or demonstrated to provide favorable clinical outcomes and efficient care in the above patient populations, and may result in harm through inaccurate assessment of sleep-disordered breathing.

Based on clinical judgment, the TF determined that the potential harms of using HSAT in the above patient populations outweigh the potential benefits. The TF also determined that patients value accurate OSA diagnosis, favorable clinical outcomes, and the identification of non-obstructive sleep-disordered breathing, and therefore would want to be evaluated by PSG.

Discussion

Four studies examining the validity of HSAT for the diagnosis of OSA in patient populations with significant cardiorespiratory morbidity met our inclusion criteria.[117–120] No RCTs were identified that randomized patients with significant co-morbidities, as outlined above, to management pathways using either PSG or HSAT for diagnosis.

Patients with Comorbid Heart Failure: Our review identified three studies that included patients with heart failure.[117,119,120] A study of 50 patients with stable heart failure (Class 2–4; left ventricular ejection fraction < 40%) evaluated the performance of home oximetry against PSG.[117] Home oximetry was considered positive if the 2% ODI ≥ 10, and the PSG was considered positive if AHI ≥ 15 using a hypopnea criteria that did not require oxygen desaturation or arousal. Home oximetry data was not obtained in 3 patients and oximetry had to be repeated in 2 patients. This study found oximetry to have a sensitivity of 0.85 and specificity of 0.93 in identifying sleep-disordered breathing.[117] The specificity was poor for identifying CSA, based on desaturation/resaturation patterns (specificity of 0.17; sensitivity of 1.0) with 10 of 12 patients with OSA identified as having CSA. A study of 50 patients with heart failure (Class 3; LVEF ≤ 35%) evaluating the performance of an HSAT device that included ECG (2 leads), oximetry, and respiratory impedance sensors against PSG, was able to obtain valid data in 44 patients in the home setting. Sensitivity, specificity and accuracy at AHI ≥ 5 and AHI ≥ 15 cutoffs were 0.92, 0.52, and 0.73 and 0.67, 0.78, and 0.75 respectively.[119] Unfortunately, the performance of the device in distinguishing central from obstructive events was not evaluated.

A study of 100 patients with stable heart failure (mean LVEF ± SD: 34.6% ± 11) evaluated the performance of simultaneous 2-channel HSAT device (nasal pressure flow and oximetry) against unattended in-home PSG.[120] In the 90 patients with valid HSAT recordings, the sensitivity and specificity was 0.98 and 0.60, respectively, using an AHI ≥ 5 cut off (hypopneas required 4% oxygen desaturation for both HSAT and PSG), and 0.93 and 0.92% using an AHI ≥ 15 cutoff. Among these patients, 29% had CSA, 19% had OSA, and 13% had both, based on PSG. The type of sleep apnea could not be determined using the HSAT device. Meta-analysis of these studies (see supplemental material, **Table S61**) found that in a population of 1,000 patients at high risk of moderate to severe OSA (64% prevalence), 45 to 230 more false negative and 18 to 79 more false positives would result from the use of HSAT.[117,119,120] The quality of evidence for was downgraded to low due to imprecision and indirectness.

Patients with Comorbid COPD: Only one study addressed the validity of HSAT (nasal pressure, respiratory excursion (piezoelectric sensor), body position and pulse oximetry) in patients with COPD.[118] Of 72 patients with stable COPD (GOLD stage II and III) and symptoms of OSA, only 26 patients (36%) had HSAT studies of reasonable quality.[118] When comparing HSAT to PSG, the intraclass correlation coefficient was 0.47 (accuracy not provided).[118] Data regarding detection of hypoventilation was not provided. Evidence was downgraded to very low based on imprecision, indirectness, and risk of bias due to significant data loss.

Patients with Other Comorbidities: No studies were identified that met our inclusion criteria that specifically evaluated the use of HSAT for diagnosis of OSA in patients with history of stroke, chronic opioid medication use, neuromuscular disease with respiratory muscle impairment, high risk of hypoventilation, or severe insomnia. Therefore, the TF concluded that HSAT has not been adequately validated or demonstrated to provide favorable clinical outcomes and efficient care in these patient populations.

Overall Quality of Evidence: The evidence for the use of HSAT in diagnosis of OSA among patients with comorbid heart failure was based on three studies, and this evidence was downgraded to low because of imprecision and indirectness.[117,119,120] The evidence for the use of HSAT in diagnosis of OSA among patients with COPD was based on a single, small study in which the majority of subjects had technically inadequate HSAT data due to recording failure. There was no direct evidence regarding suitability of HSAT for the diagnosis of OSA in patients with neuromuscular disease with respiratory impairment, hypoventilation, chronic opioid medication use, history of stroke, or severe insomnia. The overall quality of evidence for HSAT in patients with comorbid conditions was downgraded to very low due to imprecision, indirectness, and risk of bias.

Benefits versus Harms: Certain patient populations are at increased risk of having forms of SDB other than OSA (e.g., CSA, hypoventilation, and hypoxemia). These forms of SDB can cause significant morbidity and mortality if left untreated. HSAT has not been validated to diagnose some of these types of SDB (CSA, hypoventilation); therefore, the use of HSAT in populations at increased risk for SDB other than OSA increases the likelihood of not detecting these breathing disorders, which could lead to inadequate treatments, increased long-term healthcare costs, morbidity and mortality. In addition, the accuracy of HSAT has not been validated in patients with severe insomnia where it may be compromised leading to similar outcomes. Though the cost of diagnostic PSG is higher than HSAT, the TF determined that the benefits of increased accuracy, use of appropriate therapy, and improved clinical outcomes outweigh this factor. There are, however, instances where PSG cannot be performed for practical reasons (hospitalization, inability of patient to leave home setting or participate in PSG), and use of HSAT may be reasonable, as the alternative is to not addressing SDB at all.

Patients' Values and Preferences: Based on clinical judgment, the TF determined that patients at increased risk for non-OSA SDB would want these breathing disorders to be adequately diagnosed and treated, as therapy of these disorders can result in significant improvement in health and well-being, and would therefore prefer PSG. Similarly, patients with severe insomnia needing evaluation of OSA would prefer PSG. If the optimal diagnostic test (PSG) was not feasible, then they would desire to have other diagnostic tests (i.e., HSAT) available that may aid their clinical provider in providing care for SDB.

Diagnosis of Obstructive Sleep Apnea in Adults Using a Split-Night versus a Full-Night Polysomnography Protocol

Recommendation 5: We suggest that, if clinically appropriate, a split-night diagnostic protocol, rather than a full-night diagnostic protocol for polysomnography be used in the diagnosis of OSA. (WEAK)

Remarks: Clinically appropriate is defined as the absence of conditions identified by the clinician that are likely to interfere with successful diagnosis and treatment using a split-night protocol.

This recommendation is based on a split-night protocol that initiates CPAP titration only when the following criteria are met: (1) a moderate to severe degree of OSA is observed during a minimum of 2 hours of recording time on the diagnostic PSG, AND (2) at least 3 hours are available for CPAP titration.

Summary

This recommendation is based on evidence from nine studies that included typical sleep clinic patients studied for symptoms of OSA. The quality of evidence was determined to be low due to imprecision, indirectness, and risk of bias. In the context of an appropriate protocol, a split-night study has acceptable accuracy to diagnose OSA in an uncomplicated adult patient and may improve efficiency of care when performed in the context of adequate clinical and technical expertise. The split-night protocol potentially provides enhanced efficiency of care by diagnosing OSA and establishing PAP treatment needs within a single night recording.

Many studies included in our review were retrospective case series, in which patients deemed clinically inappropriate for split-night study were unlikely to have been included. Therefore, there may be specific patient characteristics, not yet adequately defined in existing literature, that render patients ill-suited to the shorter diagnostic evaluation or titration period of the split-night study. Examples of such characteristics include severe insomnia, claustrophobia, concern for other forms of sleep-disordered breathing, or concern for non-breathing-related sleep disorders.

A split-night study may be preferred relative to full-night PSG and PAP titration studies due to the convenience and cost savings of completing a diagnostic and titration study during one rather than two separate PSG studies. However, this needs to be balanced with the consequences of potentially inconclusive diagnostic or titration portions of the sleep study. If the diagnostic portion is inconclusive, a second PSG is needed. If the titration portion is inconclusive, a second PAP titration study, or the use of autoadjusting PAP may be needed. Based on clinical judgment, the TF determined that the majority of well-informed patients would choose the split-night protocol over a full-night protocol, when clinically appropriate and feasible (**Figure 2**), and that the benefits of a split-night diagnostic protocol in such circumstances outweigh the potential harms.

Discussion

Our literature search yielded nine studies that met inclusion criteria.[112,121–128] Three focused on the diagnostic accuracy of the initial portion of the PSG recording, against the accuracy of using the same full-night recording,[121,122,127] and a fourth study compared the accuracy of the diagnostic portion of a split-night study against a separate full-night study.[123] Three studies compared success of CPAP titration in those undergoing a split-night study against those undergoing a full-night sleep recording.[125,126,128] One study compared CPAP adherence in those who underwent split-night studies against those who had full-night studies.[124] A study that did not provide data suitable for inclusion in a meta-analysis examined

cost-effectiveness of the split-night study versus the full-night study. Data from this study was considered in the evaluation of resource use.[112]

Diagnostic Accuracy: Four studies that examined diagnostic accuracy and performance characteristics of a split-night protocol used the initial truncated PSG to serve as a representative surrogate of the initial diagnostic portion of a split-night study; the first 2–3 hours of the recording were compared to the full night of sleep recording.[121–123,127] One study found that the 2-hour AHI and 3-hour AHI strongly correlated with the full-night AHI (concordance correlation coefficient = 0.93 and 0.97, respectively).[121] This study reported a sensitivity of 0.80 (95% CI: 0.67 to 0.90) and specificity of 0.93 (95% CI: 0.83 to 0.98) using a cutoff of AHI ≥ 5, and a sensitivity of 0.77 (95% CI: 0.56 to 0.91) and specificity of 0.98 (95% CI: 0.92 to 1.00) using a cutoff of AHI ≥ 15 (see supplemental material, **Table S62** and **Table S63**). When comparing 3 hours of recording versus the full-night recording, excellent consistency of the AHI was observed; there was no significant difference in the AHI derived from the first 3 hours of total sleep time versus the total sleep time (concordance correlation coefficient adjusted for REM and supine sleep of 0.96 and an accuracy of 93%),[121] even in those with a milder degree of OSA (accuracy for AHI cutoffs of ≥ 5, ≥ 10 and ≥ 15 were 95, 97 and 99.5% respectively). One study assessed the diagnostic validity of a 2-hour recording and identified an optimal AHI cutoff of ≥ 30 events/h as providing the highest accuracy (90.9%).[122] This study reported a specificity of 0.90 and a sensitivity of 0.92 (see supplemental material, **Table S64**). Another study showed an AHI Pearson correlation coefficient between a full-night study and the diagnostic portion of the split-night study of 0.63 when the split-night study recording time was ≥ 90 minutes.[123] Finally, a study that compared sleep and respiratory parameters during the first 3 hours of the night against the values recorded during the entire night did not find a significant difference in AHI.[127] Given the lack of definitive data, the TF elected not to designate a specific AHI threshold to inform the decision to initiate PAP titration during a split-night study protocol. The quality of evidence for diagnostic accuracy was downgraded to low due to indirectness, imprecision, and risk of bias.

CPAP Outcomes: Our literature review identified three studies that examined CPAP success in the split-night versus full-night CPAP titration recordings. One study, focused on upper airway resistance syndrome, found no difference in the success rates of CPAP titration, defined as a respiratory effort-related arousal (RERA) index < 5 on the final CPAP setting.[125] A cross-over study involving comparisons of split-night CPAP recordings versus full-night CPAP titration recordings in patients with OSA, showed no significant difference of the AHI, arousal index and the percentage sleep time with oxygen saturation below 90% while on CPAP, though the final CPAP pressure was lower at the end of the split-night titration (8.8 versus 10.3 cm H_2O).[126] One study reported no clinically significant difference in adherence to CPAP treatment in patients undergoing a split-night study (78.7%) versus a full-night study with follow-up titration (77.5%)[124] (see supplemental material, **Figure S26**, **Figure S27**, and **Table S65**). A meta-analysis of two studies (performed by the TF) comparing reduction of AHI after CPAP treatment with split-night PSG against full-night PSG found no clinically significant difference. The quality of evidence for CPAP outcomes was downgraded to low, due to imprecision associated with a limited number of studies and small sample size.

Resource Use: A single cost-effectiveness analysis demonstrated that split-night studies were less costly than full-night studies based on cost per quality of life year (QALY) gained ($1,979 versus $2,092) and would be considered more cost-effective than full-night studies when third-party willingness to pay fell below $11,500 per QALY gained (a level of cost per QALY that would still be considered a good value for payers).[112] However, the TF had low confidence in the certainty of resource use, given the lack of high quality evidence to inform cost effectiveness.

Overall Quality of Evidence: The available studies were methodologically limited due to a number of issues: use of suboptimal study designs (not RCTs), use of the initial portion of a full-night PSG recording as a surrogate for the baseline portion of a split-night study,[121,127] and a lack of consistent use of standard monitoring (e.g., nasal pressure transducer).[121] The overall quality of evidence was determined to be low due to a combination of imprecision, indirectness, and the risk of bias.

Benefits versus Harms: The split-night protocol, in comparison to a full-night baseline assessment followed by a separate PAP titration, has the potential to provide the needed diagnostic information and effective CPAP settings within the same recording. Potential disadvantages of the split-night study include insufficient diagnostic sampling (e.g., limited REM sleep time and limited supine time in those with difficulty initiating sleep), and insufficient time to ascertain appropriate CPAP treatment settings. Based on clinical judgment, the TF determined that there is low certainty that the benefits of a split-night study in comparison to full-night studies exceed the harms.

Patients' Values and Preferences: When comparing the split-night study to the full-night study, existing data are consistent and demonstrate a high level of reproducibility of the standard AHI metric and effective identification of the optimal CPAP pressure. These data also suggest that the two approaches lead to similar follow-up CPAP adherence. Based on their clinical judgment, the TF members determined that the majority of well-informed patients would prefer a split-night protocol over a full-night protocol, when clinically appropriate and feasible (**Figure 2**), due to the lower cost, and the convenience of potentially completing a diagnostic and titration study during one sleep study. However, electing to use a split-night protocol still leaves the possibility that a patient will need to return for a second sleep study, if the diagnostic or titration portions of the split-night study are inconclusive.

Repeat Polysomnography for the Diagnosis of Obstructive Sleep Apnea in Adults

Recommendation 6: We suggest that when the initial polysomnogram is negative and there is still clinical suspicion for OSA, a second polysomnogram be considered for the diagnosis of OSA. (WEAK)

Summary

There was limited evidence from which to assess the efficacy of performing a repeat PSG when the initial PSG is negative. The recommendation is based on evidence from comparisons of a single-night PSG to two-nights of PSG for the diagnosis of OSA. These studies found no consistent differences overall in AHI scores, but potentially significant minorities of patients had results that were different in clinically meaningful ways on the two nights. The certainty in the evidence regarding night-to-night variability of AHI from the meta-analysis started as high, but there was limited evidence from which to assess the efficacy of single-night PSG versus two-night PSG in terms of diagnostic accuracy and clinical outcomes. This led to a downgrading of the overall quality of evidence to very low to reflect the low certainty of the TF that a repeat PSG would improve patient outcomes.

Discussion of a repeat PSG with a patient who has a negative initial PSG is warranted to ensure further testing accords with the patient's values and preferences, given the potential benefits and harms associated with additional testing. Proceeding with a second PSG in patients with a negative initial PSG, in order to establish a diagnosis of OSA, must be balanced against the possibility of a false positive diagnosis, inconvenience to the patient, and the added cost of a second study. Based on their clinical judgment, the TF members determined that the majority of well-informed symptomatic patients would choose a second PSG to diagnose suspected OSA when the initial PSG is negative. The TF also determined that the benefits of a second PSG outweigh the harms; however, the certainty that the benefits outweigh the harms is low.

Discussion

Our literature search identified four observational studies that compared AHI scores between two consecutive nights of PSG.[34,129–131] There was a wide range of OSA severity within the populations included in the four studies (AHI range: 7–34). None of the studies included data on body position during the 2 nights of PSG. One of two studies that reported on sleep architecture changes[130,131] found a statistically significant increase in REM sleep on the second PSG.[131] Only one of the studies indicated that PSG scorers were blinded to the other PSG result.[131]

AHI (Night-to-Night Variability): A meta-analysis of four studies compared AHI data between 2 consecutive nights of PSG[34,129–131] (see supplemental material, **Figure S28** and **Table S66**) and found the mean difference in the AHI between the 2 nights was 0.14 (95% CI: −1.86 to 2.15), which was not statistically or clinically significant. Nonetheless, a subset of individuals had considerable night-to-night variability in their AHIs, which could have potential clinical implications if the AHI crosses a treatment threshold only during the second PSG. Using an AHI cutoff of ≥ 5 to diagnose OSA, three of the studies[34,130,131] identified that 9.9% to 25% of subjects had an AHI < 5 on the first PSG but an AHI ≥ 5 only on the second PSG. Likewise, using an AHI cutoff of ≥ 15 or 20 as a potential treatment threshold, 2 of the studies[34,130] observed that 7.6% and 25% of subjects crossed this threshold only on the second study. OSA severity was also noted to vary in a subset of subjects with 26% to 35% changing the severity classification of their OSA (in either direction) on the 2 nights, though the majority were a shift of a single category (e.g., mild to moderate).[34,130] The quality of evidence for night-to-night variability was high.

Overall Quality of Evidence: The overall quality of evidence for comparing night-to-night AHI variability was originally considered high, due to precise and consistent data across studies.[34,129–131] However, the available literature did not address other clinically meaningful outcomes (e.g., impact on costs, QOL, comorbidities and long-term outcomes) resulting from undergoing a second night of PSG testing. As such, the TF downgraded the overall quality of evidence supporting this recommendation to very low, to reflect the likelihood that future research could result in different estimates of effect for the outcomes of interest, many of which were not available in the current literature.

Benefits versus Harms: A second night of PSG in symptomatic patients allows for the diagnosis of OSA in 8% to 25% of patients with initial false negative studies. Establishing a diagnosis of OSA in these patients allows for treatment that leads to improved symptom control (e.g., less daytime sleepiness), better QOL, and potentially decreased cardiovascular morbidity over time. However, routinely repeating a PSG in patients with an initial negative PSG has potential downsides. There is a risk that repeat testing could lead to false positive cases being identified, and unnecessarily treated. In addition, the routine use of a 2-night study protocol would cause inconvenience to the patient, increased utilization of resources and healthcare costs, and perhaps even delays in the care of other patients awaiting PSG. However, due to the increased likelihood of diagnosing symptomatic patients, and based on their clinical judgment, the TF determined that the benefits of a second PSG outweigh the harms; though the certainty that the benefits outweigh the harms is low.

Patients' Values and Preferences: Patient preference was also considered when weighing the values and trade-offs of a repeat PSG in a patient suspected of having OSA with an initial false negative study. The patient's desire and motivation for further testing can be affected by a variety of factors from the patient's perspective (e.g., QOL, costs) and thus a discussion with the patient is warranted prior to pursuing repeat testing. Based on their clinical judgment, the TF members determined that the majority of well-informed symptomatic patients would choose a second PSG to diagnose suspected OSA, when the initial PSG is negative.

DISCUSSION AND FUTURE DIRECTIONS

This systematic literature review identified many areas that warrant additional study to better inform clinical decision-making and improve patient outcomes.

More accurate and user-friendly clinical screening tools and models are needed to better predict presence and severity of OSA, as well as to improve risk stratification and efficiency of patient management. Identification of biomarkers that detect obstructive sleep-disordered breathing and predict likelihood of adverse clinical outcomes could provide novel information that may improve the diagnosis and management of OSA. These advancements could also improve the efficiency by which conventional sleep apnea tests that measure the physiology of breathing during sleep are used. In addition, these approaches may be useful in situations where conventional tests may not be readily available or logistically feasible to conduct in a timely fashion (e.g., inpatient settings, preoperative clinics).

The current literature is limited, as the majority of study populations included mostly men and had limited ethnic and racial diversity. Therefore, more studies in women and non-Caucasians that elucidate optimal OSA screening methodology, diagnostic approaches and management pathways are needed. These groups may present with different OSA symptoms and have different preferences with regard to, and outcomes in response to, specific OSA diagnostic and management approaches.

For patients scheduled for upper airway surgery for snoring, there is currently insufficient evidence to determine if the diagnostic evaluation of OSA can decrease peri-operative risk and improve surgical outcomes. Because it has been established that questionnaires cannot be used to diagnose OSA, many sleep experts have followed previous guidelines recommending diagnostic testing to evaluate for OSA prior to performing surgery for snoring. Further research to evaluate this protocol would be useful.

While PSG remains the gold standard for the diagnosis of OSA, it involves cumbersome sensors and devices that, if minimized and less obtrusive, could make PSG more tolerable for patients. Newer technology that is less intrusive and more comfortable may influence patient preferences regarding diagnostic approaches. Split-night PSG testing, which may improve the efficiency of PSG, has not been adequately studied. The quality of evidence regarding split-night sleep studies is low and additional research is needed to better determine its overall impact on patient outcomes. Past research often utilized outmoded testing methodology (e.g., they did not use nasal pressure cannulas) or outdated scoring criteria, limiting its relevance. There is also a lack of data on the utility of split-night testing in patients with significant underlying cardiopulmonary disease. Finally, the cost-effectiveness of split-night studies warrants further exploration.

Significant progress has been made in better understanding the accuracy and clinical utility of HSAT, but more is needed. Future research should focus on evaluating HSAT devices in patients with different pretest probabilities for OSA, and in more diverse patient populations, especially those routinely excluded (e.g., at risk for hypoventilation and CSA) from past studies, and in those unable to be studied in the sleep laboratory environment (e.g., due to critical illness, immobility, safety). In addition, the types and numbers of HSAT sensors necessary to adequately diagnose OSA require elucidation. Research should focus on how to better define the optimal physiologic parameters to be measured, particularly concerning the minimal number of parameters necessary and how devices measuring different parameters compare with one another and in different clinical situations. Furthermore, a better understanding of factors associated with inadequate or failed HSAT could help to optimize efficiency of care with regards to choosing the most appropriate diagnostic method for a given patient and clinical situation. Greater study of the cost-effectiveness of home-based management is needed to better define situations in which it may or may not offer value to the healthcare system relative to laboratory-based management. Finally, there is a paucity of data on how patient preferences currently influence clinical decision-making regarding the type of diagnostic testing. The role of patient preference regarding diagnostic pathways (i.e., HSAT versus PSG) and how this may impact outcomes remains to be explored.

More work is needed to determine the duration and number of nights that are optimal for diagnostic testing. For example, when is a second night of PSG indicated in patients suspected of having OSA but who have a negative initial study? Future studies should attempt to determine factors that may predict which patients may benefit from a second night of PSG and measure the impact on clinically meaningful outcomes (e.g., impact on costs, QOL and medical morbidity). Likewise, the duration and number of testing nights required to accurately diagnose or exclude a diagnosis OSA with HSAT is in need of further study. In terms of the minimal duration of HSAT recording time, future comparative effectiveness studies should consider the impact of HSAT duration on clinical accuracy, clinical efficiency, and functional outcomes. Comparative effectiveness studies should also consider the impact of the number of nights of HSAT on clinically meaningful outcomes and efficiency of care (e.g., time to treatment and costs).

Finally, there is a need for controlled trials to determine the role of repeat testing during chronic clinical management. There was insufficient evidence to determine whether, and under what scenarios, repeat PSG or HSAT to confirm severity of OSA or efficacy of therapy improves outcomes relative to clinical follow-up without retesting.

REFERENCES

1. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*. 2005;28(4):499–521.
2. Collop NA, Anderson WM, Boehlecke B, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2007;3(7):737–747.
3. Aurora RN, Chowdhuri S, Ramar K, et al. The treatment of central sleep apnea syndromes in adults: practice parameters with an evidence-based literature review and meta-analyses. *Sleep*. 2012;35(1):17–40.

4. Berry RB, Chediak A, Brown LK, et al. Best clinical practices for the sleep center adjustment of noninvasive positive pressure ventilation (NPPV) in stable chronic alveolar hypoventilation syndromes. *J Clin Sleep Med.* 2010;6(5):491–509.
5. Aurora RN, Bista SR, Casey KR, et al. Updated adaptive servo-ventilation recommendations for the 2012 AASM guideline: "The Treatment of Central Sleep Apnea Syndromes in Adults: Practice Parameters with an Evidence-Based Literature Review and Meta-Analyses." *J Clin Sleep Med.* 2016;12(5):757–761.
6. 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):1006–1014.
7. Barbe F, Duran-Cantolla J, Sanchez-de-la-Torre M, et al. Effect of continuous positive airway pressure on the incidence of hypertension and cardiovascular events in nonsleepy patients with obstructive sleep apnea: a randomized controlled trial. *JAMA.* 2012;307(20):2161–2168.
8. Punjabi NM, Caffo BS, Goodwin JL, et al. Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS Med.* 2009;6(8):e1000132.
9. 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):1046–1053.
10. Young T, Finn L, Peppard PE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep.* 2008;31(8):1071–1078.
11. Ravesloot MJ, van Maanen JP, Hilgevoord AA, van Wagensveld BA, de Vries N. Obstructive sleep apnea is underrecognized and underdiagnosed in patients undergoing bariatric surgery. *Eur Arch Otorhinolaryngol.* 2012;269(7):1865–1871.
12. Johnson KG, Johnson DC. Frequency of sleep apnea in stroke and TIA patients: a meta-analysis. *J Clin Sleep Med.* 2010;6(2):131–137.
13. Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc.* 2008;5(2):136–143.
14. Franklin KA, Lindberg E. Obstructive sleep apnea is a common disorder in the population-a review on the epidemiology of sleep apnea. *J Thorac Dis.* 2015;7(8):1311–1322.
15. Jackson ML, Howard ME, Barnes M. Cognition and daytime functioning in sleep-related breathing disorders. *Prog Brain Res.* 2011;190:53–68.
16. Budhiraja R, Budhiraja P, Quan SF. Sleep-disordered breathing and cardiovascular disorders. *Respir Care.* 2010;55(10):1322–1332; discussion 1330–1332.
17. Lam JC, Mak JC, Ip MS. Obesity, obstructive sleep apnoea and metabolic syndrome. *Respirology.* 2012;17(2):223–236.
18. Kapur V, Blough DK, Sandblom RE, et al. The medical cost of undiagnosed sleep apnea. *Sleep.* 1999;22(6):749–755.
19. Kakkar RK, Berry RB. Positive airway pressure treatment for obstructive sleep apnea. *Chest.* 2007;132(3):1057–1072.
20. Kapur VK. Obstructive sleep apnea: diagnosis, epidemiology, and economics. *Respir Care.* 2010;55(9):1155–1167.
21. Lichstein KL, Justin Thomas S, Woosley JA, Geyer JD. Co-occurring insomnia and obstructive sleep apnea. *Sleep Med.* 2013;14(9):824–829.
22. Iranzo A, Santamaria J. Severe obstructive sleep apnea/hypopnea mimicking REM sleep behavior disorder. *Sleep.* 2005;28(2):203–206.
23. American Academy of Sleep Medicine. *International Classification of Sleep Disorders.* 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.
24. Berry RB, Brooks R, Gamaldo CE, et al.; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications.* Version 2.3. Darien, IL: American Academy of Sleep Medicine; 2016.
25. Ruehland WR, Rochford PD, O'Donoghue FJ, Pierce RJ, Singh P, Thornton AT. The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep.* 2009;32(2):150–157.
26. Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med.* 2000;161(2 Pt 1):369–374.
27. Duce B, Milosavljevic J, Hukins C. The 2012 AASM respiratory event criteria increase the incidence of hypopneas in an adult sleep center population. *J Clin Sleep Med.* 2015;11(12):1425–1431.
28. Collop NA, Tracy SL, Kapur V, et al. Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. *J Clin Sleep Med.* 2011;7(5):531–548.
29. Flemons WW, Littner MR, Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest.* 2003;124(4):1543–1579.
30. Chesson AL, Jr., Ferber RA, Fry JM, et al. The indications for polysomnography and related procedures. *Sleep.* 1997;20(6):423–487.
31. Ferber R, Millman R, Coppola M, et al. Portable recording in the assessment of obstructive sleep apnea. ASDA standards of practice. *Sleep.* 1994;17(4):378–392.
32. Morgenthaler TI, Deriy L, Heald JL, Thomas SM. The evolution of the AASM clinical practice guidelines: another step forward. *J Clin Sleep Med.* 2016;12(1):129–135.
33. 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):485–491.
34. Ahmadi N, Shapiro GK, Chung SA, Shapiro CM. Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. *Sleep Breath.* 2009;13(3):221–226.
35. Rofail LM, Wong KK, Unger G, Marks GB, Grunstein RR. Comparison between a single-channel nasal airflow device and oximetry for the diagnosis of obstructive sleep apnea. *Sleep.* 2010;33(8):1106–1114.
36. Amra B, Nouranian E, Golshan M, Fietze I, Penzel T. Validation of the Persian version of Berlin sleep questionnaire for diagnosing obstructive sleep apnea. *Int J Prev Med.* 2013;4(3):334–339.
37. Bouloukaki I, Komninos ID, Mermigkis C, et al. Translation and validation of Berlin questionnaire in primary health care in Greece. *BMC Pulm Med.* 2013;13:6.
38. Danzi-Soares NJ, Genta PR, Nerbass FB, et al. Obstructive sleep apnea is common among patients referred for coronary artery bypass grafting and can be diagnosed by portable monitoring. *Coron Artery Dis.* 2012;23(1):31–38.
39. Friedman M, Wilson MN, Pulver T, et al. Screening for obstructive sleep apnea/hypopnea syndrome: subjective and objective factors. *Otolaryngol Head Neck Surg.* 2010;142(4):531–535.
40. Kang K, Park KS, Kim JE, et al. Usefulness of the Berlin Questionnaire to identify patients at high risk for obstructive sleep apnea: a population-based door-to-door study. *Sleep Breath.* 2013;17(2):803–810.
41. Laporta R, Anandam A, El-Solh AA. Screening for obstructive sleep apnea in veterans with ischemic heart disease using a computer-based clinical decision-support system. *Clin Res Cardiol.* 2012;101(9):737–744.
42. Pereira EJ, Driver HS, Stewart SC, Fitzpatrick MF. Comparing a combination of validated questionnaires and level III portable monitor with polysomnography to diagnose and exclude sleep apnea. *J Clin Sleep Med.* 2013;9(12):1259–1266.
43. Ulasli SS, Gunay E, Koyuncu T, et al. Predictive value of Berlin Questionnaire and Epworth Sleepiness Scale for obstructive sleep apnea in a sleep clinic population. *Clin Respir J.* 2014;8(3):292–296.
44. Sert Kuniyoshi FH, Zellmer MR, Calvin AD, et al. Diagnostic accuracy of the Berlin Questionnaire in detecting sleep-disordered breathing in patients with a recent myocardial infarction. *Chest.* 2011;140(5):1192–1197.
45. Suksakorn S, Rattanaumpawan P, Banhiran W, Cherakul N, Chotinaiwattarakul W. Reliability and validity of a Thai version of the Berlin questionnaire in patients with sleep disordered breathing. *J Med Assoc Thai.* 2014;97(Suppl 3):S46–S56.
46. Margallo VS, Muxfeldt ES, Guimaraes GM, Salles GF. Diagnostic accuracy of the Berlin questionnaire in detecting obstructive sleep apnea in patients with resistant hypertension. *J Hypertens.* 2014;32(10):2030–2036; discussion 2037.
47. Popevic MB, Milovanovic A, Nagorni-Obradovic L, Nesic D, Milovanovic J, Milovanovic AP. Screening commercial drivers for obstructive sleep apnea: translation and validation of Serbian version of Berlin Questionnaire. *Qual Life Res.* 2016;25(2):343–349.

48. Khaledi-Paveh B, Khazaie H, Nasouri M, Ghadami MR, Tahmasian M. Evaluation of berlin questionnaire validity for sleep apnea risk in sleep clinic populations. *Basic Clin Neurosci.* 2016;7(1):43–48.
49. Cowan DC, Allardice G, Macfarlane D, et al. Predicting sleep disordered breathing in outpatients with suspected OSA. *BMJ Open.* 2014;4(4):e004519.
50. Pataka A, Daskalopoulou E, Kalamaras G, Fekete Passa K, Argyropoulou P. Evaluation of five different questionnaires for assessing sleep apnea syndrome in a sleep clinic. *Sleep Med.* 2014;15(7):776–781.
51. 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 (Engl).* 2014;127(17):3065–3070.
52. Kim B, Lee EM, Chung YS, Kim WS, Lee SA. The utility of three screening questionnaires for obstructive sleep apnea in a sleep clinic setting. *Yonsei Med J.* 2015;56(3):684–690.
53. Gantner D, Ge JY, Li LH, et al. Diagnostic accuracy of a questionnaire and simple home monitoring device in detecting obstructive sleep apnoea in a Chinese population at high cardiovascular risk. *Respirology.* 2010;15(6):952–960.
54. Simpson L, Hillman DR, Cooper MN, et al. High prevalence of undiagnosed obstructive sleep apnoea in the general population and methods for screening for representative controls. *Sleep Breath.* 2013;17(3):967–973.
55. Nicholl DD, Ahmed SB, Loewen AH, et al. Diagnostic value of screening instruments for identifying obstructive sleep apnea in kidney failure. *J Clin Sleep Med.* 2013;9(1):31–38.
56. Sforza E, Chouchou F, Pichot V, Herrmann F, Barthelemy JC, Roche F. Is the Berlin questionnaire a useful tool to diagnose obstructive sleep apnea in the elderly? *Sleep Med.* 2011;12(2):142–146.
57. Facco FL, Ouyang DW, Zee PC, Grobman WA. Development of a pregnancy-specific screening tool for sleep apnea. *J Clin Sleep Med.* 2012;8(4):389–394.
58. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991;14(6):540–545.
59. Subramanian S, Hesselbacher SE, Aguilar R, Surani SR. The NAMES assessment: a novel combined-modality screening tool for obstructive sleep apnea. *Sleep Breath.* 2011;15(4):819–826.
60. Chen NH, Chen MC, Li HY, Chen CW, Wang PC. A two-tier screening model using quality-of-life measures and pulse oximetry to screen adults with sleep-disordered breathing. *Sleep Breath.* 2011;15(3):447–454.
61. Zou J, Guan J, Yi H, et al. An effective model for screening obstructive sleep apnea: a large-scale diagnostic study. *PLoS One.* 2013;8(12):e80704.
62. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun Y. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. *Br J Anaesth.* 2012;108(5):768–775.
63. Ong TH, Raudha S, Fook-Chong S, Lew N, Hsu AA. Simplifying STOP-BANG: use of a simple questionnaire to screen for OSA in an Asian population. *Sleep Breath.* 2010;14(4):371–376.
64. Sadeghniiat-Haghighi K, Montazeri A, Khajeh-Mehrizi A, et al. The STOP-BANG questionnaire: reliability and validity of the Persian version in sleep clinic population. *Qual Life Res.* 2015;24(8):2025–2030.
65. Alhouqani S, Al Manhali M, Al Essa A, Al-Houqani M. Evaluation of the Arabic version of STOP-Bang questionnaire as a screening tool for obstructive sleep apnea. *Sleep Breath.* 2015;19(4):1235–1240.
66. BaHammam AS, Al-Aqeel AM, Alhedyani AA, Al-Obaid GI, Al-Owais MM, Olaish AH. The validity and reliability of an Arabic version of the STOP-Bang questionnaire for identifying obstructive sleep apnea. *Open Respir Med J.* 2015;9:22–29.
67. Banhiran W, Durongphan A, Saleesing C, Chongkolwatana C. Diagnostic properties of the STOP-Bang and its modified version in screening for obstructive sleep apnea in Thai patients. *J Med Assoc Thai.* 2014;97(6):644–654.
68. Chung F, Chau E, Yang Y, Liao P, Hall R, Mokhlesi B. Serum bicarbonate level improves specificity of STOP-Bang screening for obstructive sleep apnea. *Chest.* 2013;143(5):1284–1293.
69. Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology.* 2008;108(5):812–821.
70. Gurubhagavatula I, Fields BG, Morales CR, et al. Screening for severe obstructive sleep apnea syndrome in hypertensive outpatients. *J Clin Hypertens (Greenwich).* 2013;15(4):279–288.
71. Kushida CA, Efron B, Guilleminault C. A predictive morphometric model for the obstructive sleep apnea syndrome. *Ann Intern Med.* 1997;127(8 Pt 1):581–587.
72. Gurubhagavatula I, Maislin G, Pack AI. An algorithm to stratify sleep apnea risk in a sleep disorders clinic population. *Am J Respir Crit Care Med.* 2001;164(10 Pt 1):1904–1909.
73. Rofail LM, Wong KK, Unger G, Marks GB, Grunstein RR. The utility of single-channel nasal airflow pressure transducer in the diagnosis of OSA at home. *Sleep.* 2010;33(8):1097–1105.
74. Wilson G, Terpening Z, Wong K, et al. Screening for sleep apnoea in mild cognitive impairment: the utility of the multivariable apnoea prediction index. *Sleep Disord.* 2014;2014:945287.
75. Morales CR, Hurley S, Wick LC, et al. In-home, self-assembled sleep studies are useful in diagnosing sleep apnea in the elderly. *Sleep.* 2012;35(11):1491–1501.
76. Chang ET, Yang MC, Wang HM, Lai HL. Snoring in a sitting position and neck circumference are predictors of sleep apnea in Chinese patients. *Sleep Breath.* 2014;18(1):133–136.
77. Sharma SK, Malik V, Vasudev C, et al. Prediction of obstructive sleep apnea in patients presenting to a tertiary care center. *Sleep Breath.* 2006;10(3):147–154.
78. Zerah-Lancner F, Lofaso F, d'Ortho MP, et al. Predictive value of pulmonary function parameters for sleep apnea syndrome. *Am J Respir Crit Care Med.* 2000;162(6):2208–2212.
79. Kolotkin RL, LaMonte MJ, Walker JM, Cloward TV, Davidson LE, Crosby RD. Predicting sleep apnea in bariatric surgery patients. *Surg Obes Relat Dis.* 2011;7(5):605–610.
80. Platt AB, Wick LC, Hurley S, et al. Hits and misses: screening commercial drivers for obstructive sleep apnea using guidelines recommended by a joint task force. *J Occup Environ Med.* 2013;55(9):1035–1040.
81. Chai-Coetzer CL, Antic NA, Rowland LS, et al. A simplified model of screening questionnaire and home monitoring for obstructive sleep apnoea in primary care. *Thorax.* 2011;66(3):213–219.
82. Firat H. Comparison of four established questionnaires to identify highway bus drivers at risk for obstructive sleep apnea in Turkey. *Sleep Biol Rhythms.* 2012;10(3):231–236.
83. Berry RB, Hill G, Thompson L, McLaurin V. Portable monitoring and autotitration versus polysomnography for the diagnosis and treatment of sleep apnea. *Sleep.* 2008;31(10):1423–1431.
84. Rosen CL, Auckley D, Benca R, et al. A multisite randomized trial of portable sleep studies and positive airway pressure autotitration versus laboratory-based polysomnography for the diagnosis and treatment of obstructive sleep apnea: the HomePAP study. *Sleep.* 2012;35(6):757–767.
85. Kuna ST, Gurubhagavatula I, Maislin G, et al. Noninferiority of functional outcome in ambulatory management of obstructive sleep apnea. *Am J Respir Crit Care Med.* 2011;183(9):1238–1244.
86. Campbell AJ, Neill AM. Home set-up polysomnography in the assessment of suspected obstructive sleep apnea. *J Sleep Res.* 2011;20(1 Pt 2):207–213.
87. Gjevre JA, Taylor-Gjevre RM, Skomro R, Reid J, Fenton M, Cotton D. Comparison of polysomnographic and portable home monitoring assessments of obstructive sleep apnea in Saskatchewan women. *Can Respir J.* 2011;18(5):271–274.
88. Masa JF, Corral J, Pereira R, et al. Effectiveness of home respiratory polygraphy for the diagnosis of sleep apnoea and hypopnoea syndrome. *Thorax.* 2011;66(7):567–573.
89. Polese JF, Santos-Silva R, de Oliveira Ferrari PM, Sartori DE, Tufik S, Bittencourt L. Is portable monitoring for diagnosing obstructive sleep apnea syndrome suitable in elderly population? *Sleep Breath.* 2013;17(2):679–686.
90. Santos-Silva R, Sartori DE, Truksinas V, et al. Validation of a portable monitoring system for the diagnosis of obstructive sleep apnea syndrome. *Sleep.* 2009;32(5):629–636.
91. Yin M, Miyazaki S, Ishikawa K. Evaluation of type 3 portable monitoring in unattended home setting for suspected sleep apnea: factors that may affect its accuracy. *Otolaryngol Head Neck Surg.* 2006;134:204–209.

92. Planes C, Leroy M, Bouach Khalil N, et al. Home diagnosis of obstructive sleep apnoea in coronary patients: validity of a simplified device automated analysis. *Sleep Breath.* 2010;14(1):25–32.
93. Masa JF, Corral J, Pereira R, et al. Effectiveness of sequential automatic-manual home respiratory polygraphy scoring. *Eur Respir J.* 2013;41(4):879–887.
94. Garcia-Diaz E, Quintana-Gallego E, Ruiz A, et al. Respiratory polygraphy with actigraphy in the diagnosis of sleep apnea-hypopnea syndrome. *Chest.* 2007;131(3):725–732.
95. Ayappa I, Norman RG, Seelall V, Rapoport DM. Validation of a self-applied unattended monitor for sleep disordered breathing. *J Clin Sleep Med.* 2008;4(1):26–37.
96. Tonelli de Oliveira AC, Martinez D, Vasconcelos LF, et al. Diagnosis of obstructive sleep apnea syndrome and its outcomes with home portable monitoring. *Chest.* 2009;135(2):330–336.
97. Ward KL, McArdle N, James A, et al. A comprehensive evaluation of a two-channel portable monitor to "rule in" obstructive sleep apnea. *J Clin Sleep Med.* 2015;11(4):433–444.
98. Baltzan MA, Verschelden P, Al-Jahdali H, Olha AE, Kimoff RJ. Accuracy of oximetry with thermistor (OxiFlow) for diagnosis of obstructive sleep apnea and hypopnea. *Sleep.* 2000;23(1):61–69.
99. Masdeu MJ, Ayappa I, Hwang D, Mooney AM, Rapoport DM. Impact of clinical assessment on use of data from unattended limited monitoring as opposed to full-in lab PSG in sleep disordered breathing. *J Clin Sleep Med.* 2010;6(1):51–58.
100. Nakano H, Tanigawa T, Ohnishi Y, et al. Validation of a single-channel airflow monitor for screening of sleep-disordered breathing. *Eur Respir J.* 2008;32(4):1060–1067.
101. Ozmen OA, Tuzemen G, Kasapoglu F, et al. The reliability of SleepStrip as a screening test in obstructive sleep apnea syndrome. *Kulak Burun Bogaz Ihtis Derg.* 2011;21(1):15–19.
102. Pang KP, Dillard TA, Blanchard AR, Gourin CG, Podolsky R, Terris DJ. A comparison of polysomnography and the SleepStrip in the diagnosis of OSA. *Otolaryngol Head Neck Surg.* 2006;135(2):265–268.
103. Watkins MR, Talmage JB, Thiese MS, Hudson TB, Hegmann KT. Correlation between screening for obstructive sleep apnea using a portable device versus polysomnography testing in a commercial driving population. *J Occup Environ Med.* 2009;51(10):1145–1150.
104. O'Brien LM, Bullough AS, Shelgikar AV, Chames MC, Armitage R, Chervin RD. Validation of Watch-PAT-200 against polysomnography during pregnancy. *J Clin Sleep Med.* 2012;8(3):287–294.
105. Pittman SD, Ayas NT, MacDonald MM, Malhotra A, Fogel RB, White DP. Using a wrist-worn device based on peripheral arterial tonometry to diagnose obstructive sleep apnea: in-laboratory and ambulatory validation. *Sleep.* 2004;27(5):923–933.
106. Garg N, Rolle AJ, Lee TA, Prasad B. Home-based diagnosis of obstructive sleep apnea in an urban population. *J Clin Sleep Med.* 2014;10(8):879–885.
107. Andreu AL, Chiner E, Sancho-Chust JN, et al. Effect of an ambulatory diagnostic and treatment programme in patients with sleep apnoea. *Eur Respir J.* 2012;39(2):305–312.
108. Antic NA, Buchan C, Esterman A, et al. A randomized controlled trial of nurse-led care for symptomatic moderate-severe obstructive sleep apnea. *Am J Respir Crit Care Med.* 2009;179(6):501–508.
109. Mulgrew AT, Fox N, Ayas NT, Ryan CF. Diagnosis and initial management of obstructive sleep apnea without polysomnography: a randomized validation study. *Ann Intern Med.* 2007;146(3):157–166.
110. Skomro RP, Gjevre J, Reid J, et al. Outcomes of home-based diagnosis and treatment of obstructive sleep apnea. *Chest.* 2010;138(2):257–263.
111. Chung F, Liao P, Elsaid H, Islam S, Shapiro CM, Sun Y. Oxygen desaturation index from nocturnal oximetry: a sensitive and specific tool to detect sleep-disordered breathing in surgical patients. *Anesth Analg.* 2012;114(5):993–1000.
112. Deutsch PA, Simmons MS, Wallace JM. Cost-effectiveness of split-night polysomnography and home studies in the evaluation of obstructive sleep apnea syndrome. *J Clin Sleep Med.* 2006;2(2):145–153.
113. Chervin RD, Murman DL, Malow BA, Totten V. Cost-utility of three approaches to the diagnosis of sleep apnea: polysomnography, home testing, and empirical therapy. *Ann Intern Med.* 1999;130(6):496–505.
114. Pietzsch JB, Garner A, Cipriano LE, Linehan JH. An integrated health-economic analysis of diagnostic and therapeutic strategies in the treatment of moderate-to-severe obstructive sleep apnea. *Sleep.* 2011;34(6):695–709.
115. Kim RD, Kapur VK, Redline-Bruch J, et al. An economic evaluation of home versus laboratory-based diagnosis of obstructive sleep apnea. *Sleep.* 2015;38(7):1027–1037.
116. Masa JF, Corral J, Sanchez de Cos J, et al. Effectiveness of three sleep apnea management alternatives. *Sleep.* 2013;36(12):1799–1807.
117. Series F, Kimoff RJ, Morrison D, et al. Prospective evaluation of nocturnal oximetry for detection of sleep-related breathing disturbances in patients with chronic heart failure. *Chest.* 2005;127(5):1507–1514.
118. Oliveira MG, Nery LE, Santos-Silva R, et al. Is portable monitoring accurate in the diagnosis of obstructive sleep apnea syndrome in chronic pulmonary obstructive disease? *Sleep Med.* 2012;13(8):1033–1038.
119. Abraham WT, Trupp RJ, Phillilps B, et al. Validation and clinical utility of a simple in-home testing tool for sleep-disordered breathing and arrhythmias in heart failure: results of the Sleep Events, Arrhythmias, and Respiratory Analysis in Congestive Heart Failure (SEARCH) study. *Congest Heart Fail.* 2006;12(5):241–247; quiz 248–249.
120. de Vries GE, van der Wal HH, Kerstjens HA, et al. Validity and predictive value of a portable two-channel sleep-screening tool in the identification of sleep apnea in patients with heart failure. *J Card Fail.* 2015;21(10):848–855.
121. Khawaja IS, Olson EJ, van der Walt C, et al. Diagnostic accuracy of split-night polysomnograms. *J Clin Sleep Med.* 2010;6(4):357–362.
122. Chou KT, Chang YT, Chen YM, et al. The minimum period of polysomnography required to confirm a diagnosis of severe obstructive sleep apnoea. *Respirology.* 2011;16(7):1096–1102.
123. Chuang ML, Lin IF, Vintch JR, Liao YF. Predicting continuous positive airway pressure from a modified split-night protocol in moderate to severe obstructive sleep apnea-hypopnea syndrome. *Intern Med.* 2008;47(18):1585–1592.
124. Collen J, Holley A, Lettieri C, Shah A, Roop S. The impact of split-night versus traditional sleep studies on CPAP compliance. *Sleep Breath.* 2010;14(2):93–99.
125. Kristo DA, Shah AA, Lettieri CJ, et al. Utility of split-night polysomnography in the diagnosis of upper airway resistance syndrome. *Sleep Breath.* 2009;13(3):271–275.
126. Yamashiro Y, Kryger MH. CPAP titration for sleep apnea using a split-night protocol. *Chest.* 1995;107(1):62–66.
127. Ciftci B, Ciftci TU, Guven SF. [Split-night versus full-night polysomnography: comparison of the first and second parts of the night]. *Arch Bronconeumol.* 2008;44(1):3–7.
128. Elshaug AG, Moss JR, Southcott AM. Implementation of a split-night protocol to improve efficiency in assessment and treatment of obstructive sleep apnoea. *Intern Med J.* 2005;35(4):251–254.
129. Gouveris H, Selivanova O, Bausmer U, Goepel B, Mann W. First-night-effect on polysomnographic respiratory sleep parameters in patients with sleep-disordered breathing and upper airway pathology. *Eur Arch Otorhinolaryngol.* 2010;267(9):1449–1453.
130. Ma J, Zhang C, Zhang J, et al. Prospective study of first night effect on 2-night polysomnographic parameters in adult Chinese snorers with suspected obstructive sleep apnea hypopnea syndrome. *Chin Med J (Engl).* 2011;124(24):4127–4131.
131. Selwa LM, Marzec ML, Chervin RD, et al. Sleep staging and respiratory events in refractory epilepsy patients: Is there a first night effect? *Epilepsia.* 2008;49(12):2063–2068.

ACKNOWLEDGMENTS

The task force thanks and acknowledges the contributions of Reem Mustafa, MD, MPH, PhD for her work as a methodologist. The task force also thanks and acknowledges the contributions of Lauren Loeding, MPH, for her preliminary work on this guideline.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication January 27, 2017
Submitted in final revised form January 27, 2017
Accepted for publication January 31, 2017
Address correspondence to: Vishesh K. Kapur, MD, MPH, University of Washington, Harborview Medical Center, 325 Ninth Avenue, Box 359762, Seattle, WA 98104; Tel: (206) 744-5703; Fax: (206) 744-5657; Email: research@aasmnet.org

DISCLOSURE STATEMENT

The development of this clinical practice guideline was funded by the American Academy of Sleep Medicine. Dr. Auckley receives royalties from Up-to-Date and is a consultant for the American Board of Internal Medicine. Dr. Chowdhuri has received research support from the Veteran's Health Administration. Dr. Mehra has received research support from Philips Respironics and Resmed in the form of equipment used in clinical research. Dr. Mehra also received grant support from the NIH/NHLBI and royalties from Up-to-Date. Mr. Harrod is an employee of the American Academy of Sleep Medicine. The other authors have indicated no financial conflicts of interest.

Literature Search Terms

sleep apnea, sleep apnoea, OSA, sleep apnea syndromes, ambulatory, monitoring, polysomnography, snoring, otorhinolaryngologic surgical procedures, surgical procedures, surgery, snoring, obstructive sleep apnea, diagnosis, diagnostic, sleep-related breathing disorders, sleep-disordered breathing, portable, home, limited, unattended, non-laboratory, in-home, out of center, monitor, care service, test, testing, sleep study, screening, recording, device, diagnosed, diagnoses, PSG, polysomnogram, respiratory polygraphy, repeat, retest, retesting, home diagnostic test, multichannel recorder, multi-night, split-night, follow-up, two-night, multiple-night

MeSH Terms

sleep apnea syndromes, sleep apnea obstructive, diagnosis, mass screening, probability, predictive value of tests, adult, ambulatory monitoring, polysomnography, follow-up studies, humans, snoring, otorhinolaryngologic surgical procedures, snoring/surgery, tongue/surgery, diagnostic techniques and procedures

Literature Search Limits

January 1, 2005 to June 29, 2016; Human studies, RCTs or observational studies, adults, English language

Inclusion Criteria

Diagnosis of OSA with PSG, HSAT, oximetry, or clinical prediction algorithm; address one of nine PICO questions, adults, outcomes related to accuracy, inconclusive results, complications, quality of life, medical outcomes, adherence, efficiency of diagnosis or access to care

Exclusion Criteria

Treatment paper, no OSA, pediatric subjects, initial sample size > 25 per condition, 50 total for PICO 2, initial sample size > 10 per condition, 20 total for all other PICOs, wrong publication type (review, editorial, methodological, non-RCT or non-observational study), other sleep comorbidities besides OSA, hospitalized or general surgery, diagnostic test not in PICO question, time between HSAT and PSG > 4 weeks, HSAT used in-lab, HSAT used simultaneously with PSG in-lab, MSLT, MWT, and other nap tests performed

Table S1—Summary of Downstream Consequences of OSA Diagnostic Outcomes

True Positive (TP)

- Effective treatment and improved QOL
- Ineffective treatment and worsening of symptoms
- Increased costs due to treatment
- Time for treatment and follow-up
- Psychological distress
- Side-effects of therapy
- Improvement in comorbid conditions (e.g., hypertension)
- Reduced risk of CV events
- Reduced risk of post-CV events
- Reduced risk of motor vehicle accident (MVA)
- Reduced overall health costs

True Negative (TN)

- Confirmation of absence of OSA
- Possible repeat testing if patient deemed at high risk for OSA
- Psychological relief
- Consideration of alternative causes for symptoms
- Saves time and resources
- Focused treatment on true cause of symptoms

False Positive (FP)

- Unnecessary treatment and utilization of resources
- Increased costs due to treatment
- Time for treatment and follow-up
- Psychological distress
- Delay in diagnosis of true condition
- Side-effects of therapy

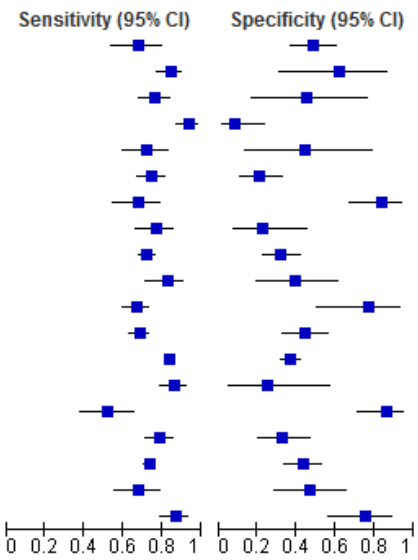
False Negative (FN)

- Absence of necessary treatment
- Reduced QOL
- Psychological distress
- Possible repeat testing if patient deemed at high risk for OSA
- Risk of motor vehicle accident (MVA)
- Risk of hypertension
- Risk of CV events
- Post-MI events
- Post-stroke events
- Death
- Increased costs and utilization of resources due to other condition(s)

Diagnosis of obstructive sleep apnea in adults using clinical tools, questionnaires and predication algorithms

Figure S1—Berlin Questionnaire vs. PSG (AHI ≥ 5)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi 2008	38	38	18	36	0.68 [0.54, 0.80]	0.49 [0.37, 0.61]
Amra 2013	121	5	23	8	0.84 [0.77, 0.90]	0.62 [0.32, 0.86]
Bouloukaki 2013	90	6	28	5	0.76 [0.68, 0.84]	0.45 [0.17, 0.77]
Cowan 2014	89	31	6	3	0.94 [0.87, 0.98]	0.09 [0.02, 0.24]
Danzi-Soares 2012	44	5	17	4	0.72 [0.59, 0.83]	0.44 [0.14, 0.79]
Friedman 2010	120	49	41	13	0.75 [0.67, 0.81]	0.21 [0.12, 0.33]
Kang 2013	44	6	21	30	0.68 [0.55, 0.79]	0.83 [0.67, 0.94]
Khaledi-Paveh 2016	60	17	18	5	0.77 [0.66, 0.86]	0.23 [0.08, 0.45]
Kim 2015	356	66	139	31	0.72 [0.68, 0.76]	0.32 [0.23, 0.42]
Laporta 2012	56	14	12	9	0.82 [0.71, 0.91]	0.39 [0.20, 0.61]
Luo 2014	130	4	65	13	0.67 [0.60, 0.73]	0.76 [0.50, 0.93]
Margallo 2014	236	42	110	34	0.68 [0.63, 0.73]	0.45 [0.33, 0.57]
Pataka 2014	1260	222	241	130	0.84 [0.82, 0.86]	0.37 [0.32, 0.42]
Pereira 2013	100	9	16	3	0.86 [0.79, 0.92]	0.25 [0.05, 0.57]
Popevic 2016	30	6	28	36	0.52 [0.38, 0.65]	0.86 [0.71, 0.95]
Rofail 2010	105	35	28	17	0.79 [0.71, 0.86]	0.33 [0.20, 0.47]
Sarinc Ulasi 2013	986	58	362	44	0.73 [0.71, 0.75]	0.43 [0.33, 0.53]
Sert Kuniyoshi 2011	47	16	22	14	0.68 [0.56, 0.79]	0.47 [0.28, 0.66]
Suksakorn 2014	87	8	13	24	0.87 [0.79, 0.93]	0.75 [0.57, 0.89]



Pooled sensitivity: 0.76 [0.72, 0.80]
Pooled specificity: 0.45 [0.34, 0.56]
LR+: 1.38 [1.15, 1.66]
LR-: 0.53 [0.42, 0.65]
DOR: 2.63 [1.79, 3.86]
Accuracy: 0.70 or **70%**

Figure S2—ROC Curve for Berlin Questionnaire vs. PSG (AHI ≥ 5)

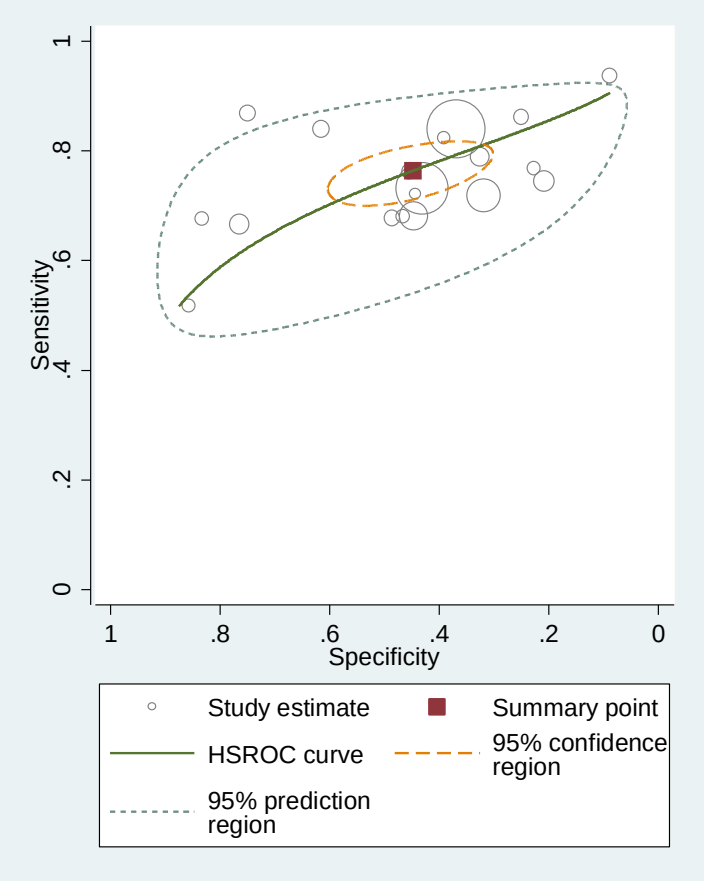


Table S2—Summary of Findings table for Berlin Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Ahmadi 2008 (A); Amra 2013 (B); Bouloukaki 2013 (C); Danzi-Soares 2012 (D); Friedman 2010 (E); Kang 2013 (F); Laporta 2012 (G); Pereira 2013 (H); Rofail 2010 (I); Sarinc Ulasi 2013 (J); Sert Kuniyoshi 2011 (K); Cowan 2014 (L); Khaledi-Paveh 2016 (M); Kim 2015 (N); Luo 2014 (O); Margallo 2014 (P); Pataka 2014 (Q); Popevic 2016 R; Suksakorn 2014 (S)

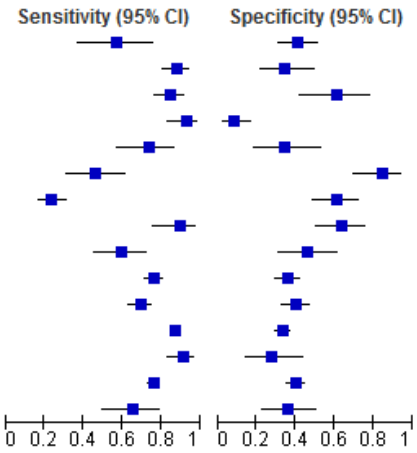
Pooled sensitivity Berlin Questionnaire: 0.76 (95% CI: 0.72 to 0.80) | **Pooled specificity Berlin Questionnaire**: 0.45 (95% CI: 0.34 to 0.56) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 72% (95% CI: 70 to 74%) **Accuracy (low risk):** 62% (95% CI: 60 to 64%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Berlin Questionnaire	Attended PSG	Berlin Questionnaire	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	661 (626 to 696)	870 (870 to 870)	418 (396 to 440)	550 (550 to 550)	6303 (19) [A-S]
		209 fewer TP in Berlin Questionnaire		**132 fewer TP in Berlin Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		209 (174 to 244)	0 (0 to 0)	132 (110 to 154)	0 (0 to 0)	
		209 more FN in Berlin Questionnaire		**132 more FN in Berlin Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	59 (44 to 73)	130 (130 to 130)	202 (153 to 252)	450 (450 to 450)	6303 (19) [A-S]
		71 fewer TN in Berlin Questionnaire		**248 fewer TN in Berlin Questionnaire**		
False positives (patients incorrectly classified as having OSA)		71 (57 to 86)	0 (0 to 0)	248 (198 to 297)	0 (0 to 0)	
		71 more FP in Berlin Questionnaire		**248 more FP in Berlin Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals

Figure S3—Berlin Questionnaire vs. PSG (AHI ≥ 15)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmadi 2008	16	60	12	42	0.57 [0.37, 0.76]	0.41 [0.32, 0.51]
Amra 2013	95	32	13	17	0.88 [0.80, 0.93]	0.35 [0.22, 0.50]
Bouloukaki 2013	83	12	15	19	0.85 [0.76, 0.91]	0.61 [0.42, 0.78]
Cowan 2014	53	66	4	6	0.93 [0.83, 0.98]	0.08 [0.03, 0.17]
Danzi-Soares 2012	28	21	10	11	0.74 [0.57, 0.87]	0.34 [0.19, 0.53]
Firat 2012	21	6	25	33	0.46 [0.31, 0.61]	0.85 [0.69, 0.94]
Friedman 2010	35	29	114	45	0.23 [0.17, 0.31]	0.61 [0.49, 0.72]
Kang 2013	34	23	4	40	0.89 [0.75, 0.97]	0.63 [0.50, 0.75]
Khaledi-Paveh 2016	32	25	22	21	0.59 [0.45, 0.72]	0.46 [0.31, 0.61]
Kim 2015	260	160	83	89	0.76 [0.71, 0.80]	0.36 [0.30, 0.42]
Margallo 2014	160	114	72	76	0.69 [0.63, 0.75]	0.40 [0.33, 0.47]
Pataka 2014	1093	408	168	204	0.87 [0.85, 0.89]	0.33 [0.30, 0.37]
Pereira 2013	80	29	8	11	0.91 [0.83, 0.96]	0.28 [0.15, 0.44]
Sarinc Ulasi 2013	769	261	246	174	0.76 [0.73, 0.78]	0.40 [0.35, 0.45]
Sert Kuniyoshi 2011	30	34	16	19	0.65 [0.50, 0.79]	0.36 [0.23, 0.50]



Pooled sensitivity: 0.75 [0.64, 0.83]
Pooled specificity: 0.42 [0.32, 0.52]
LR+: 1.29 [1.12, 1.48]
LR-: 0.60 [0.44, 0.81]
DOR: 2.16 [1.42, 3.27]
Accuracy: 0.63 or **63%**

Figure S4—ROC Curve for Berlin Questionnaire vs. PSG (AHI ≥ 15)

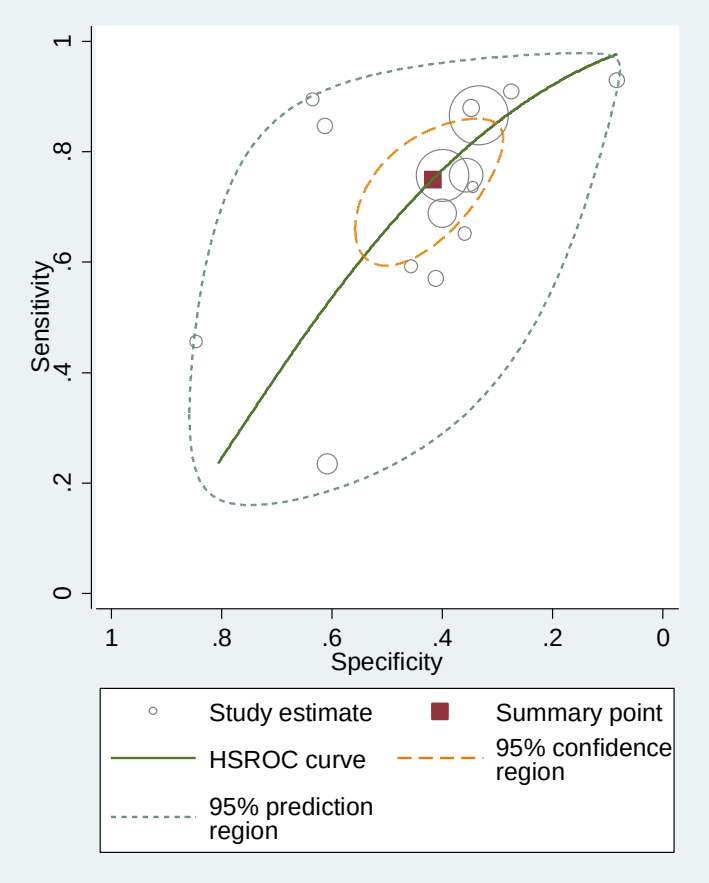


Table S3—Summary of Findings table for Berlin Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Ahmadi 2008 (A); Amra 2013 (B); Bouloukaki 2013 (C); Danzi-Soares 2012 (D); Friedman 2010 (E); Kang 2013 (F); Pereira 2013 (G); Sarinc Ulasi 2013 (H); Sert Kuniyoshi 2011 (I); Cowan 2014 (J); Khaledi-Paveh 2016 (K); Kim 2015 (L); Margallo 2014 (M); Pataka 2014 (N); Firat 2012 (O)

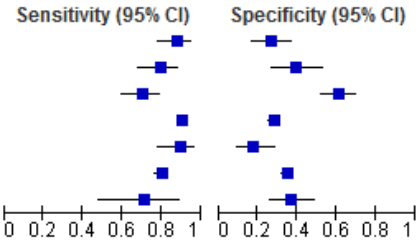
Pooled sensitivity Berlin Questionnaire: 0.75 (95% CI: 0.64 to 0.83) | **Pooled specificity Berlin Questionnaire**: 0.42 (95% CI: 0.32 to 0.52) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 63% (95% CI: 61 to 65%) **Accuracy (low risk):** 50% (95% CI: 50 to 50%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Berlin Questionnaire	Attended PSG	Berlin Questionnaire	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	480 (410 to 531)	640 (640 to 640)	188 (160 to 208)	250 (250 to 250)	5668 (15) [A-O]
		160 fewer TP in Berlin Questionnaire		**62 fewer TP in Berlin Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		160 (109 to 230)	0 (0 to 0)	62 (42 to 90)	0 (0 to 0)	
		209 more FN in Berlin Questionnaire		**62 more FN in Berlin Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	151 (115 to 187)	360 (360 to 360)	315 (240 to 390)	750 (750 to 750)	5668 (15) [A-O]
		209 fewer TN in Berlin Questionnaire		**435 fewer TN in Berlin Questionnaire**		
False positives (patients incorrectly classified as having OSA)		209 (173 to 245)	0 (0 to 0)	435 (360 to 510)	0 (0 to 0)	
		209 more FP in Berlin Questionnaire		**435 more FP in Berlin Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals

Figure S5—Berlin Questionnaire vs. PSG (AHI ≥ 30)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amra 2013	65	61	9	22	0.88 [0.78, 0.94]	0.27 [0.17, 0.37]
Bouloukaki 2013	54	37	14	24	0.79 [0.68, 0.88]	0.39 [0.27, 0.53]
Friedman 2010	67	49	29	78	0.70 [0.60, 0.79]	0.61 [0.52, 0.70]
Pataka 2014	811	681	88	273	0.90 [0.88, 0.92]	0.29 [0.26, 0.32]
Pereira 2013	50	59	6	13	0.89 [0.78, 0.96]	0.18 [0.10, 0.29]
Sarinc Ulasi 2013	487	547	122	294	0.80 [0.77, 0.83]	0.35 [0.32, 0.38]
Sert Kuniyoshi 2011	15	49	6	29	0.71 [0.48, 0.89]	0.37 [0.26, 0.49]



Pooled sensitivity: 0.84 [0.77, 0.89]
Pooled specificity: 0.35 [0.26, 0.44]
DOR: 2.73 [2.11, 3.52]
LR+: 1.28 [1.17, 1.41]
LR-: 0.47 [0.38, 0.58]
Accuracy: 0.56 or **56%**

Figure S6—ROC Curve for Berlin Questionnaire vs. PSG (AHI ≥ 30)

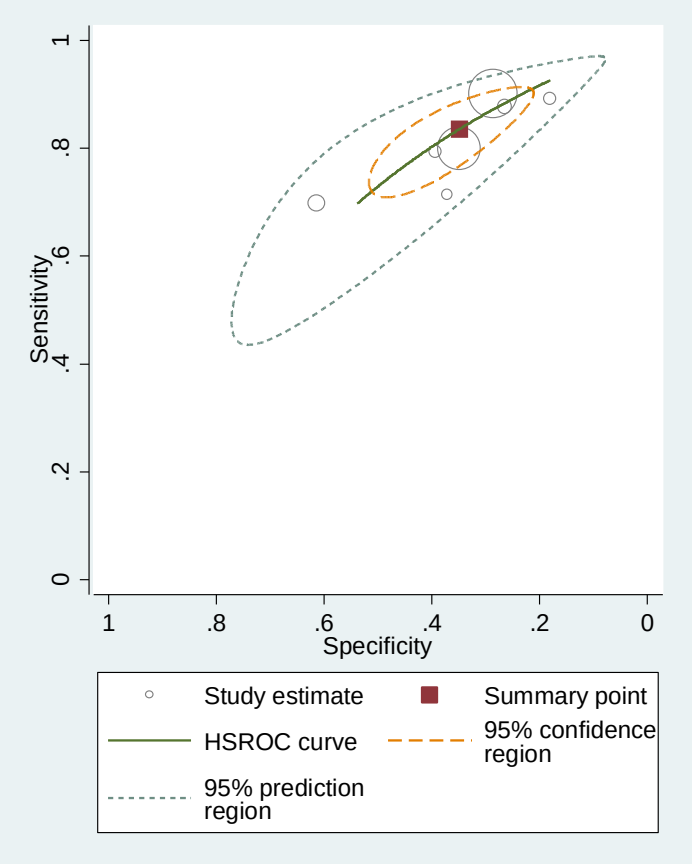


Table S4—Summary of Findings table for Berlin Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Amra 2013 (A); Bouloukaki 2013 (B); Friedman 2010 (C); Pereira 2013 (D); Sarinc Ulasi 2013 (E); Sert Kuniyoshi 2011 (F); Pataka 2014 (G)

Pooled sensitivity Berlin Questionnaire: 0.84 (95% CI: 0.77 to 0.89) | **Pooled specificity Berlin Questionnaire**: 0.35 (95% CI: 0.26 to 0.44) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 53% (95% CI: 52 to 53%) **Accuracy (low risk):** 40% (95% CI: 38 to 42%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Berlin Questionnaire	Attended PSG	Berlin Questionnaire	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	302 (277 to 320)	360 (360 to 360)	84 (77 to 89)	100 (100 to 100)	4039 (7) [A-G]
		58 fewer TP in Berlin Questionnaire		**16 fewer TP in Berlin Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		58 (40 to 83)	0 (0 to 0)	16 (11 to 23)	0 (0 to 0)	
		58 more FN in Berlin Questionnaire		**16 more FN in Berlin Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	224 (166 to 282)	640 (640 to 640)	315 (234 to 396)	900 (900 to 900)	4039 (7) [A-G]
		416 fewer TN in Berlin Questionnaire		**585 fewer TN in Berlin Questionnaire**		
False positives (patients incorrectly classified as having OSA)		416 (358 to 474)	0 (0 to 0)	585 (504 to 666)	0 (0 to 0)	
		416 more FP in Berlin Questionnaire		**585 more FP in Berlin Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals

Table S5—Summary of Findings table for Berlin Questionnaire vs. Home Sleep Apnea Test (HSAT) to diagnose OSA in Suspected Adults (AHI ≥ 5)

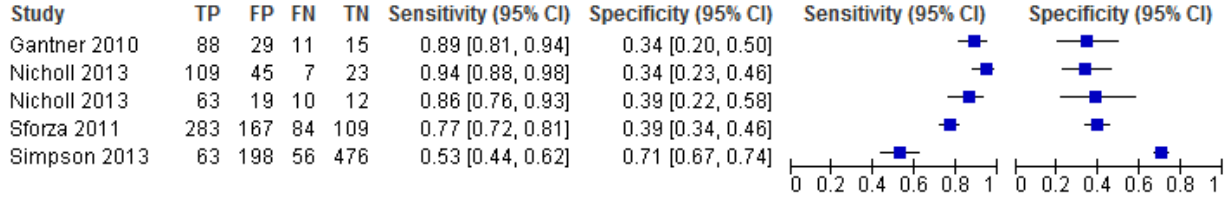
References: Facco 2012 (A)

Single study sensitivity Berlin Questionnaire: 0.39 (95% CI: 0.22 to 0.59) | **Single study specificity Berlin Questionnaire**: 0.68 (95% CI: 0.56 to 0.78) |**Single study sensitivity HSAT**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity HSAT**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy high risk**: 43% (95% CI: 26 to 61%) **Accuracy low risk:** 52% (95% CI: 37 to 68%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Berlin Questionnaire	HSAT	Berlin Questionnaire	HSAT	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	339 (191 to 513)	870 (870 to 870)	215 (121 to 325)	550 (550 to 550)	100 (1) [A]
		531 fewer TP in Berlin Questionnaire		**335 fewer TP in Berlin Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		531 (679 to 357)	0 (0 to 0)	335 (429 to 225)	0 (0 to 0)	
		531 more FN in Berlin Questionnaire		**335 more FN in Berlin Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	88 (73 to 101)	130 (130 to 130)	306 (252 to 351)	450 (450 to 450)	100 (1) [A]
		42 fewer TN in Berlin Questionnaire		**144 fewer TN in Berlin Questionnaire**		
False positives (patients incorrectly classified as having OSA)		42 (57 to 29)	0 (0 to 0)	144 (198 to 99)	0 (0 to 0)	
		42 more FP in Berlin Questionnaire		**144 more FP in Berlin Questionnaire**		

[1]Study consisted of pregnant women only
[2]Wide confidence interval for sensitivity and specificity

Figure S7—Berlin Questionnaire vs. HSAT (AHI ≥ 15)



Pooled sensitivity: 0.76 [0.64, 0.85]
Pooled specificity: 0.44 [0.30, 0.58]
LR+: 1.36 [0.91, 2.02]
LR-: 0.54 [0.26, 1.20]
DOR: 2.31 [1.68, 2.42]
Accuracy: 0.67 or **67%**

Table S6—Summary of Findings table for Berlin Questionnaire vs. HSAT to diagnose OSA in Suspected Adult (AHI ≥ 15)

References: Gantner 2010 (A); Nicholl 2013 (B); Sforza 2011 (C); Simpson 2013 (D)

Pooled sensitivity Berlin Questionnaire: 0.76 (95% CI: 0.44 to 0.85) | **Pooled specificity Berlin Questionnaire**: 0.44 (95% CI: 0.30 to 0.58) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 53% (95% CI: 52 to 53%) **Accuracy (low risk):** 40% (95% CI: 38 to 42%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Berlin Questionnaire	HSAT	Berlin Questionnaire	HSAT	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	486 (282 to 544)	640 (640 to 640)	190 (110 to 213)	250 (250 to 250)	1751 (4) [A-D]
		154 fewer TP in Berlin Questionnaire		**60 fewer TP in Berlin Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		154 (96 to 358)	0 (0 to 0)	60 (37 to 140)	0 (0 to 0)	
		154 more FN in Berlin Questionnaire		**60 more FN in Berlin Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	158 (108 to 209)	360 (360 to 360)	330 (225 to 435)	750 (750 to 750)	1751 (4) [A-D]
		202 fewer TN in Berlin Questionnaire		**420 fewer TN in Berlin Questionnaire**		
False positives (patients incorrectly classified as having OSA)		202 (151 to 252)	0 (0 to 0)	420 (315 to 525)	0 (0 to 0)	
		202 more FP in Berlin Questionnaire		**420 more FP in Berlin Questionnaire**		

[1]Broad range of specificity and sensitivity across studies
[2]Wide confidence intervals

Table S7—Berlin Questionnaire vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Gantner 2010 (A); Nicoll 2013 (B)

Range of sensitivities Berlin Questionnaire: 0.76 to 0.92 | **Range of specificities Berlin Questionnaire**: 0.26 to 0.42
Range of sensitivities HSAT: 1.00 to 1.00 | **Range of specificities HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 44% to 60%
Accuracy (low risk): 31% to 47%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Berlin Questionnaire	HSAT	Berlin Questionnaire	HSAT	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	274 to 331	360 to 360	76 to 92	100 to 100	315 (2) [A,B]
		29 to 86 fewer TP in Berlin Questionnaire		**8 to 24 fewer TP in Berlin Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		86 to 29	0 to 0	24 to 8	0 to 0	
		29 to 86 more FN in Berlin Questionnaire		**8 to 24 more FN in Berlin Questionnaire**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	166 to 269	640 to 640	234 to 378	900 to 900	315 (2) [A,B]
		371 to 474 fewer TN in Berlin Questionnaire		**666 fewer to 522 fewer TN in Berlin Questionnaire**		
False positives (patients incorrectly classified as having OSA)		371 to 474	0 to 0	522 to 666	0 to 0	
		371 to 474 fewer TP in Berlin Questionnaire		**522 to 666 fewer TP in Berlin Questionnaire**		

[1]Indirect evidence as Berlin Questionnaire was not compared against HSAT in any of the PICO questions
[2]Wide range of sensitivity and specificity across studies

Table S8—Summary of Findings table for ESS vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Chen 2011 (A); Danzi-Soares (B); Zou 2013 (C); Sarinc Ulasli 2013 (D); Pataka 2014 (E); Luo 2014 (F)

Range of sensitivities Epworth Sleepiness Scale: 0.27 to 0.72 | Range of specificities Epworth Sleepiness Scale: 0.50 to 0.76 | Range of sensitivities Attended PSG: 1.00 to 1.00 | Range of specificities Attended PSG: 1.00 to 1.00 | **Accuracy (high risk)**: (51% to 52%) **Accuracy (low risk):** (54% to 59%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		ESS	Attended PSG	ESS	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	235 to 626	870 to 870	149 to 396	550 to 550	4724 (6) [A-F]
		244 to 635 fewer TP in ESS		**154 to 401 fewer TP in ESS**		
False negatives (patients incorrectly classified as not having OSA)		244 to 635	0 to 0	154 to 401	0 to 0	
		244 to 635 more FN in ESS		**154 to 401 more FN in ESS**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	65 to 99	130 to 130	225 to 342	450 to 450	4724 (6) [A-F]
		31 to 65 fewer TN in ESS		**108 to 225 fewer TN in ESS**		
False positives (patients incorrectly classified as having OSA)		31 to 65	0 to 0	108 to 225	0 to 0	
		31 to 65 more FP in ESS		**108 to 225 more FP in ESS**		

[1]Wide range of values for sensitivity and specificity across studies
[2]Wide confidence intervals for sensitivity and specificity

Table S9—Summary of Findings table for ESS vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Danzi-Soares (A); Subramanian 2011 (B); Ulasli 2014 (C); Pataka 2014 (D); Luo 2014 (E)

Range of sensitivities ESS: 0.21 to 0.58 | **Range of specificities ESS**: 0.50 to 0.72 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 46% to 48% **Accuracy (low risk):** 54% to 58%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		ESS	Attended PSG	ESS	Attended PSG	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	134 to 371	640 to 640	53 to 143	250 to 250	4093 (5)[A-E]
		269 to 506 fewer TP in ESS		**105 to 197 fewer TP in ESS**		
False negatives (patients incorrectly classified as not having OSA)		269 to 506	0 to 0	105 to 197	0 to 0	
		269 to 506 more FN in ESS		**105 to 197 more FN in ESS**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	180 to 259	360 to 360	375 to 540	750 to 750	4093 (5)[A-E]
		101 to 180 fewer TN in ESS		**210 to 375 fewer TN in ESS**		
False positives (patients incorrectly classified as having OSA)		101 to180	0 to 0	201 to 375	0 to 0	
		101 to 180 more FP in ESS		**201 to 375 more FP in ESS**		

[1]Wide range of values for sensitivity and specificity
[2]Confidence interval for studies is wide

Table S10—Summary of Findings table for ESS vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Ulasli 2014 (A); Pataka 2014 (B); Luo 2014 (C)

Range of sensitivities ESS: 0.53 to 0.63 | **Range of specificities ESS**: 0.54 to 0.62 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 46% to 48% **Accuracy (low risk):** 54% to 58%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		ESS	Attended PSG	ESS	Attended PSG	
True positives (patients with OSA)	⨁⨁⨁⨁ HIGH	191 to 227	360 to 360	53 to 63	100 to 100	3515 (3) [A-C]
		133 to 169 fewer TP in ESS		**37 to 47 fewer TP in ESS**		
False negatives (patients incorrectly classified as not having OSA)		133 to 169	0 to 0	37 to 47	0 to 0	
		133 to 169 more FN in ESS		**105 to 197 more FN in ESS**		
True negatives (patients without OSA)	⨁⨁⨁⨁ HIGH	346 to 397	640 to 640	486 to 558	900 to 900	3515 (3) [A-C]
		243 to 294 fewer TN in ESS		**342 to 414 fewer TN in ESS**		
False positives (patients incorrectly classified as having OSA)		243 to 294	0 to 0	342 to 414	0 to 0	
		243 to 294 more FP in ESS		**342 to 414 more FP in ESS**		

Table S11—Summary of Findings table for ESS vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Facco 2012 (A)

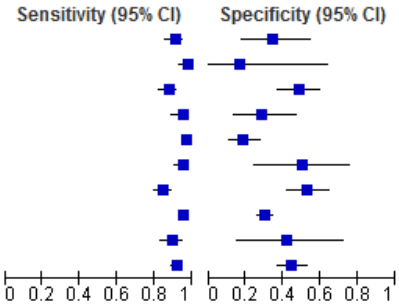
Single study sensitivity ESS: 0.36 (95% CI: 0.19 to 0.57) | **Single study specificity ESS**: 0.77 (95% CI: 0.66 to 0.86) **Single study sensitivity HSAT**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity HSAT**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 41% (95% CI: 25 to 61%) **Accuracy (low risk):** 54% (95% CI: 40 to 70%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		ESS	HSAT	ESS	HSAT	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	313 (165 to 496)	870 (870 to 870)	198 (105 to 314)	550 (550 to 550)	100 (1) [A]
		557 fewer TP in ESS		**352 fewer TP in ESS**		
False negatives (patients incorrectly classified as not having OSA)		557 (374 to 705)	0 (0 to 0)	352 (236 to 445)	0 (0 to 0)	
		557 more FN in ESS		**352 more FN in ESS**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	100 (86 to 112)	130 (130 to 130)	347 (297 to 387)	450 (450 to 450)	100 (1) [A]
		30 fewer TN in ESS		**103 fewer TN in ESS**		
False positives (patients incorrectly classified as having OSA)		30 (18 to 44)	0 (0 to 0)	103 (63 to 153)	0 (0 to 0)	
		30 more FP in ESS		**103 more FP in ESS**		

[1]Study consisted of pregnant women
[2]Wide confidence interval for sensitivity and specificity

Figure S8—STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alhouqani 2015	149	19	15	10	0.91 [0.85, 0.95]	0.34 [0.18, 0.54]
BaMammam 2015	92	5	2	1	0.98 [0.93, 1.00]	0.17 [0.00, 0.64]
Banhiran 2014	194	42	28	39	0.87 [0.82, 0.91]	0.48 [0.37, 0.60]
Cowan 2014	92	23	5	9	0.95 [0.88, 0.98]	0.28 [0.14, 0.47]
Kim 2015	480	77	17	18	0.97 [0.95, 0.98]	0.19 [0.12, 0.28]
Luo 2014	186	8	10	8	0.95 [0.91, 0.98]	0.50 [0.25, 0.75]
Ong 2010	201	38	37	43	0.84 [0.79, 0.89]	0.53 [0.42, 0.64]
Pataka 2014	1311	329	70	143	0.95 [0.94, 0.96]	0.30 [0.26, 0.35]
Pereira 2013	104	7	12	5	0.90 [0.83, 0.95]	0.42 [0.15, 0.72]
Sadeghniiat-Haghighi 2015	404	90	36	73	0.92 [0.89, 0.94]	0.45 [0.37, 0.53]



Pooled sensitivity: 0.93 [0.90, 0.95]
Pooled specificity: 0.36 [0.29, 0.44]
LR+: 1.46 [1.32, 1.62]
LR-: 0.19 [0.16, 0.23]
DOR: 7.72 [6.35, 9.39]
Accuracy: 0.80 or **80%**

Figure S9—ROC Curve for STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

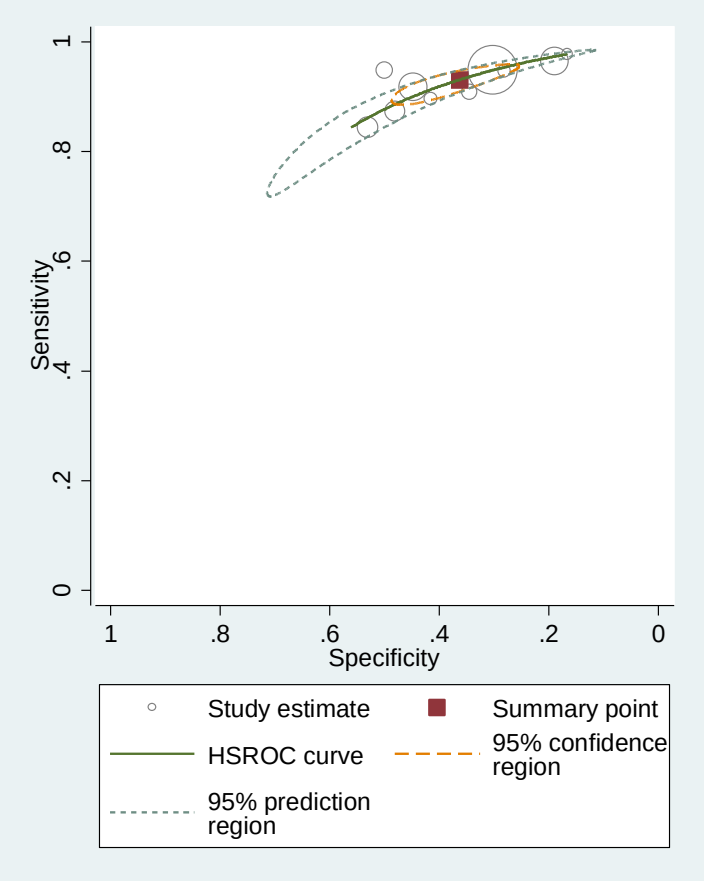


Table S12—Summary of Findings table for STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Alouqani 2015 (A); BaHammam 2015 (B); Banhiran 2014 (C); Cowan 2014 (D); Kim 2015 (E); Luo 2014 (F); Ong 2010 (G); Pataka 2014 (H); Pereira 2013 (I); Sadeghniiat-Highighi 2015 (J)

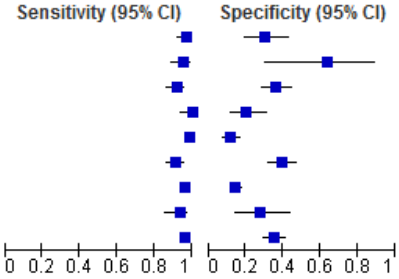
Pooled sensitivity STOP-BANG Questionnaire: 0.93 (95% CI: 0.90 to 0.95) | **Pooled specificity STOP-BANG Questionnaire**: 0.36 (95% CI: 0.29 to 0.44) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 53% (95% CI: 52 to 53%) **Accuracy (low risk):** 40% (95% CI: 38 to 42%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		STOP- BANG Questionnaire	Attended PSG	STOP- BANG Questionnaire	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	809 (783 to 827)	870 (870 to 870)	512 (495 to 523)	550 (550 to 550)	4432 (10) [A-J]
		61 fewer TP in STOP-BANG Questionnaire		**38 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		61 (43 to 87)	0 (0 to 0)	38 (27 to 55)	0 (0 to 0)	
		61 more FN in STOP-BANG Questionnaire		**38 more FN in STOP-BANG Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	47 (38 to 57)	130 (130 to 130)	162 (130 to 198)	450 (450 to 450)	4432 (10) [A-J]
		83 fewer TN in STOP-BANG Questionnaire		**288 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		83 (73 to 92)	0 (0 to 0)	288 (252 to 320)	0 (0 to 0)	
		83 more FP in STOP-BANG Questionnaire		**288 more FP in STOP-BANG Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals for specificity

Figure S10—STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alhouqani 2015	120	48	4	21	0.97 [0.92, 0.99]	0.30 [0.20, 0.43]
BaHammam 2015	85	4	4	7	0.96 [0.89, 0.99]	0.64 [0.31, 0.89]
Banhiran 2014	139	97	12	55	0.92 [0.87, 0.96]	0.36 [0.29, 0.44]
Cowan 2014	56	58	0	15	1.00 [0.94, 1.00]	0.21 [0.12, 0.32]
Kim 2015	337	219	6	30	0.98 [0.96, 0.99]	0.12 [0.08, 0.17]
Ong 2010	148	97	14	63	0.91 [0.86, 0.95]	0.39 [0.32, 0.47]
Pataka 2014	1200	514	51	88	0.96 [0.95, 0.97]	0.15 [0.12, 0.18]
Pereira 2013	82	29	6	11	0.93 [0.86, 0.97]	0.28 [0.15, 0.44]
Sadeghniiat-Haghighi 2015	306	185	12	100	0.96 [0.94, 0.98]	0.35 [0.30, 0.41]



Pooled sensitivity: 0.95 [0.94, 0.97]
Pooled specificity: 0.27 [0.20, 0.36]
LR+: 1.31 [1.18, 1.45]
LR-: 0.17 [0.12, 0.23]
DOR: 7.86 [5.37, 11.49]
Accuracy: 0.68 or **68%**

Figure S11—STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

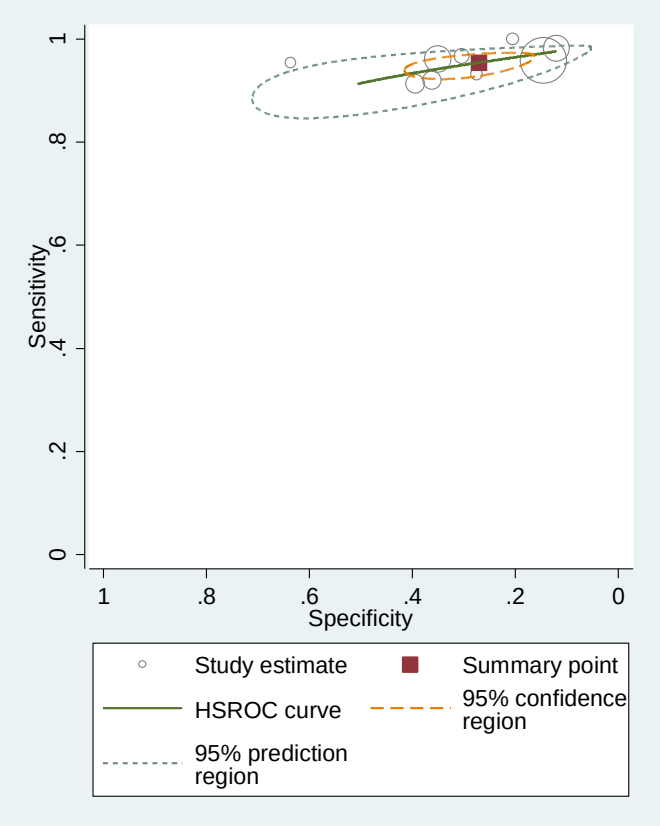


Table S13—Summary of Findings table for STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Alouqani 2015 (A); BaHammam 2015 (B); Banhiran 2014 (C); Cowan 2014 (D); Kim 2015 (E); Ong 2010 (F); Pataka 2014 (G); Pereira 2013 (H); Sadeghniiat-Highighi 2015 (I)

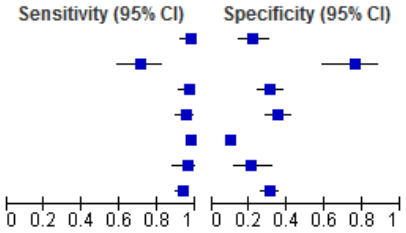
Pooled sensitivity STOP-BANG Questionnaire: 0.95 (95% CI: 0.94 to 0.97) | **Pooled specificity STOP-BANG Questionnaire**: 0.27 (95% CI: 0.20 to 0.36) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 70% (95% CI: 69 to 72%) **Accuracy (low risk):** 43% (95% CI: 42 to 44%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		STOP- BANG Questionnaire	Attended PSG	STOP- BANG Questionnaire	Attended PSG	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	614 (602 to 621)	640 (640 to 640)	240 (235 to 243)	250 (250 to 250)	4223 (9) [A-I]
		26 fewer TP in STOP-BANG Questionnaire		**10 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		26 (19 to 38)	0 (0 to 0)	10 (7 to 15)	0 (0 to 0)	
		26 more FN in STOP-BANG Questionnaire		**10 more FN in STOP-BANG Questionnaire**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	90 (65 to 122)	360 (360 to 360)	188 (135 to 255)	750 (750 to 750)	4223 (9) [A-I]
		270 fewer TN in STOP-BANG Questionnaire		**562 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		270 (238 to 295)	0 (0 to 0)	562 (495 to 615)	0 (0 to 0)	
		270 more FP in STOP-BANG Questionnaire		**562 more FP in STOP-BANG Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals for specificity

Figure S12—STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alhouqani 2015	85	83	2	23	0.98 [0.92, 1.00]	0.22 [0.14, 0.31]
BaHammam 2015	45	9	18	28	0.71 [0.59, 0.82]	0.76 [0.59, 0.88]
Banhiran 2014	94	142	3	64	0.97 [0.91, 0.99]	0.31 [0.25, 0.38]
Ong 2010	106	135	5	73	0.95 [0.90, 0.99]	0.35 [0.29, 0.42]
Pataka 2014	890	852	18	93	0.98 [0.97, 0.99]	0.10 [0.08, 0.12]
Pereira 2013	54	57	2	15	0.96 [0.88, 1.00]	0.21 [0.12, 0.32]
Sadeghniiat-Haghighi 2015	189	278	13	123	0.94 [0.89, 0.97]	0.31 [0.26, 0.35]



Pooled sensitivity: 0.94 [0.90, 0.97]
Pooled specificity: 0.30 [0.17, 0.46]
LR+: 1.34 [1.12, 1.61]
LR-: 0.18 [0.14, 0.24]
DOR: 7.37 [5.37, 10.1]
Accuracy: 0.54 or **54%**

Figure S13—ROC Curve for STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

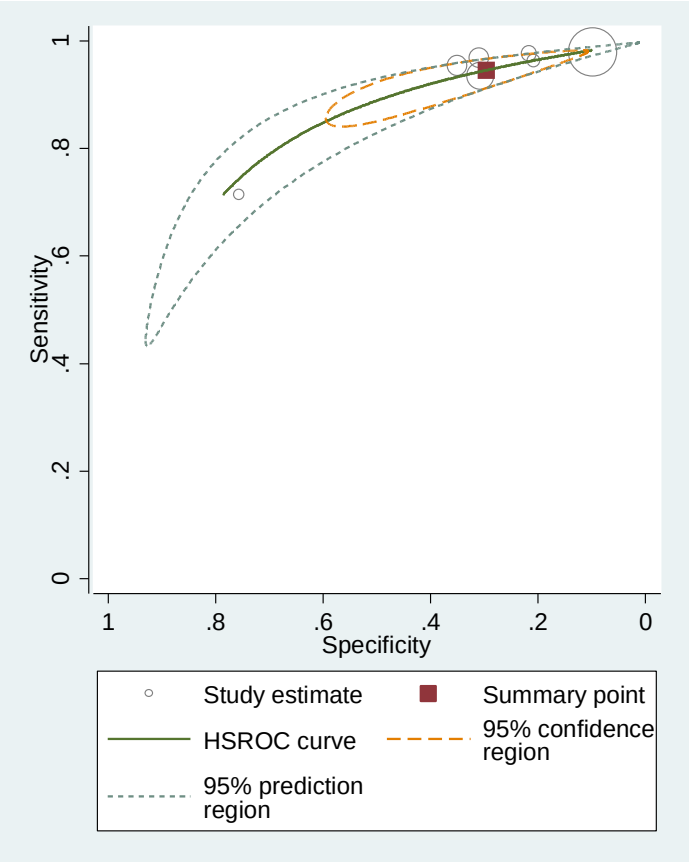


Table S14—Summary of Findings table for STOP-BANG Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Alouqani 2015 (A); BaHammam 2015 (B); Banhiran 2014 (C); Ong 2010 (D); Pataka 2014 (E); Pereira 2013 (F); Sadeghniiat-Highighi 2015 (G)

Pooled sensitivity STOP-BANG Questionnaire: 0.94 (95% CI: 0.90 to 0.97) | **Pooled specificity STOP-BANG Questionnaire**: 0.30 (95% CI: 0.17 to 0.46) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 53% (95% CI: 53 to 55%) **Accuracy (low risk):** 36% (95% CI: 33 to 40%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		STOP- BANG Questionnaire	Attended PSG	STOP- BANG Questionnaire	Attended PSG	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	338 (324 to 349)	360 (360 to 360)	94 (90 to 97)	100 (100 to 100)	3449 (7) [A-G]
		22 fewer TP in STOP-BANG Questionnaire		**6 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		22 (11 to 36)	0 (0 to 0)	6 (3 to 10)	0 (0 to 0)	
		22 more FN in STOP-BANG Questionnaire		**6 more FN in STOP-BANG Questionnaire**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	192 (109 to 294)	640 (640 to 640)	270 (153 to 414)	900 (900 to 900)	3449 (7) [A-G]
		448 fewer TN in STOP-BANG Questionnaire		**630 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		448 (346 to 531)	0 (0 to 0)	630 (486 to 747)	0 (0 to 0)	
		270 more FP in STOP-BANG Questionnaire		**630 more FP in STOP-BANG Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals for specificity

Table S15—Summary of Findings table for STOP-BANG Questionnaire vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Chung 2013 (A)

Single study sensitivity STOP- BANG Questionnaire: 0.87 (95% CI: 0.80 to 0.92) | **Single study specificity STOP- BANG Questionnaire**: 0.33 (95% CI: 0.21 to 0.48) **Single study sensitivity HSAT**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity HSAT**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 80% (95% CI: 72 to 86%) **Accuracy (low risk):** 63% (95% CI: 53 to 72%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		STOP- BANG Questionnaire	HSAT	STOP- BANG Questionnaire	HSAT	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	757 (696 to 800)	870 (870 to 870)	479 (440 to 506)	550 (550 to 550)	192 (1) [A]
		113 fewer TP in STOP- BANG Questionnaire		**71 fewer TP in STOP- BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		113 (174 to 70)	0 (0 to 0)	71 (44 to 110)	0 (0 to 0)	
		113 more FN in STOP- BANG Questionnaire		**71 more FN in STOP- BANG Questionnaire**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	43 (27 to 62)	130 (130 to 130)	149 (94 to 216)	450 (450 to 450)	192 (1) [A]
		87 fewer TN in STOP- BANG Questionnaire		**301 fewer TN in STOP- BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		87 (68 to 103)	0 (0 to 0)	301 (234 to 356)	0 (0 to 0)	
		87 more FP in STOP- BANG Questionnaire		**301 more FP in STOP- BANG Questionnaire**		

Table S16—Summary of Findings table for STOP-BANG Questionnaire vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Chung 2013 (A); Nicholl 2013 (B)

Range of sensitivities STOP-BANG Questionnaire: 0.88 to 0.94 | **Range of specificities STOP-BANG Questionnaire**: 0.24 to 0.31
Range of sensitivities HSAT: 1.00 to 1.00 | **Range of specificities HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 65% to 71% **Accuracy (low risk):** 40% to 47%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		STOP- BANG Questionnaire	HSAT	STOP- BANG Questionnaire	HSAT	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	563 to 602	640 to 640	220 to 235	250 to 250	364 (2) [A,B]
		38 to 77 fewer TP in STOP-BANG Questionnaire		**15 to 30 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		38 to 77	0 to 0	15 to 30	0 to 0	
		38 to 77 more FN in STOP-BANG Questionnaire		**15 to 30 more FN in STOP- BANG Questionnaire**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	86 to 112	360 to 360	180 to 232	750 to 750	364 (2) [A,B]
		248 to 274 fewer TN in STOP-BANG Questionnaire		**570 to 518 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		248 to 274	0 to 0	518 to 570	0 to 0	
		248 to 274 more FP in STOP-BANG Questionnaire		**518 to 570 more FP in STOP-BANG Questionnaire**		

[1]Indirect evidence as one of the two studies consisted of chronic kidney disease and end-stage renal disease patients.

Table S17—Summary of Findings table for STOP-BANG Questionnaire vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Chung 2013 (A); Nicholl 2013 (B)

Range of sensitivities STOP- BANG Questionnaire: 0.88 to 1.00 | **Range of specificities STOP- BANG Questionnaire**: 0.20 to 0.53 **Range of sensitivities HSAT**: 1.00 to 1.00 | **Range of specificities HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 44% to 70% **Accuracy (low risk):** 27% to 58%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		STOP- BANG Questionnaire	HSAT	STOP- BANG Questionnaire	HSAT	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	317 to 360	360 to 360	88 to 100	100 to 100	364 (2) [A,B]
		0 to 43 fewer TP in STOP-BANG Questionnaire		**0 to 12 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		0 to 43	0 to 0	0 to 12	0 to 0	
		0 to 43 more FN in STOP-BANG Questionnaire		**0 more to 12 fewer FN in STOP- BANG Questionnaire**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	128 to 339	640 to 640	180 to 477	900 to 900	364 (2) [A,B]
		301 to 512 fewer TN in STOP- BANG Questionnaire		**423 to 720 fewer TN in STOP- BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		301 to 512	0 to 0	423 to 720	0 to 0	
		301 to 512 more FP in STOP- BANG Questionnaire		**423 to 720 more FP in STOP-BANG Questionnaire**		

[1]Indirect evidence as one study consisted of pregnant women, and the other study consisted of chronic kidney disease and end-stage renal disease patients
[2]Braod range of specificity across studies

Table S18—Summary of Findings table for STOP-BANG Questionnaire vs. PSG or HSAT to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Chung 2013 (A); Nicholl 2013 (B)

Range of sensitivities STOP- BANG Questionnaire: 0.18 to 0.90 | **Range of specificities STOP- BANG Questionnaire**: 0.28 to 0.88 **Range of sensitivities PSG or HSAT**: 1.00 to 1.00 | **Range of specificities PSG or HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 19% to 90% **Accuracy (low risk):** 22% to 89%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		STOP- BANG Questionnaire	PSG or HSAT	STOP- BANG Questionnaire	PSG or HSAT	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	157 to 783	870 to 870	99 to 495	550 to 550	364 (2) [A,B]
		87 to 713 fewer TP in STOP-BANG Questionnaire		**55 to 451 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		87 to 713	0 to 0	55 to 451	0 to 0	
		87 to 713 more FN in STOP-BANG Questionnaire		**55 to 451 more FN in STOP-BANG Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	36 to 114	130 to 130	126 to 396	450 to 450	364 (2) [A,B]
		16 to 94 fewer TN in STOP-BANG Questionnaire		**54 to 324 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		16 to 94	0 to 0	54 to 324	0 to 0	
		16 to 94 more FP in STOP-BANG Questionnaire		**54 to 324 more FP in STOP-BANG Questionnaire**		

[1]Indirect evidence as one study consisted of pregnant women, and the other study consisted of chronic kidney disease and end-stage renal disease patients
[2]Broad range of specificity across studies

Table S19—Summary of Findings table for STOP-BANG Questionnaire vs. PSG or HSAT to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Chung 2012 (A); Chung 2013 (B)

Range of sensitivities STOP- BANG Questionnaire: 0.10 to 0.95 | **Range of specificities STOP- BANG Questionnaire**: 0.11 to 0.88
Range of sensitivities PSG or HSAT: 1.00 to 1.00 | **Range of specificities PSG or HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 10% to 92% **Accuracy (low risk):** 11% to 90%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		STOP- BANG Questionnaire	PSG or HSAT	STOP- BANG Questionnaire	PSG or HSAT	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	64 to 608	640 to 640	25 to 238	250 to 250	1056 (2) [A,B]
		32 to 576 fewer TP in STOP-BANG Questionnaire		**12 to 225 fewer TP in STOP-BANG Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		32 to 576	0 to 0	12 to 225	0 to 0	
		32 to 576 more FN in STOP-BANG Questionnaire		**12 to 225 more FN in STOP-BANG Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	40 to 317	360 to 360	83 to 660	750 to 750	1056 (2) [A,B]
		43 to 320 fewer TN in STOP-BANG Questionnaire		**90 to 667 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		43 to 320	0 to 0	90 to 667	0 to 0	
		43 to 320 more FP in STOP-BANG Questionnaire		**90 to 667 more FP in STOP-BANG Questionnaire**		

[1]Indirect evidence as one study consisted of pregnant women
[2]Wide range of specificity and sensitivity

Table S20—Summary of Findings table for STOP-BANG Questionnaire vs. PSG or HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Chung 2012 (A); Chung 2013 (B)

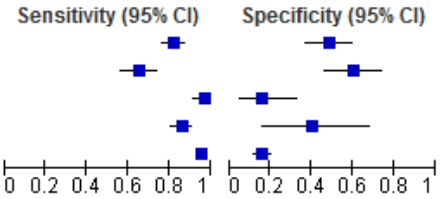
Range of sensitivities STOP- BANG Questionnaire: 0.28 to 1.00 | **Range of specificities STOP- BANG Questionnaire**: 0.17 to 0.88 **Range of sensitivities PSG or HSAT**: 1.00 to 1.00 | **Range of specificities PSG or HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 21% to 92% **Accuracy (low risk):** 18% to 89%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		STOP- BANG Questionnaire	PSG or HSAT	STOP- BANG Questionnaire	PSG or HSAT	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	101 to 360	360 to 360	28 to 100	100 to 100	1056 (2) [A,B]
		0 to 259 fewer TP in STOP-BANG Questionnaire		**0 to 72 fewer TP in STOP-Bang Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		0 to 259	0 to 0	0 to 72	0 to 0	
		0 to 259 more FN in STOP-BANG Questionnaire		**0 to 72 more FN in STOP-Bang Questionnaire**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	109 to 563	640 to 640	153 to 792	900 to 900	1056 (2) [A,B]
		77 to 531 fewer TN in STOP-BANG Questionnaire		**108 to 747 fewer TN in STOP-BANG Questionnaire**		
False positives (patients incorrectly classified as having OSA)		77 to 531	0 to 0	108 to 747	0 to 0	
		77 to 531 more FP in STOP-BANG Questionnaire		**108 to 747 more FP in STOP-BANG Questionnaire**		

[1]Indirect evidence as one study consisted of pregnant women
[2]Wide range of specificity and sensitivity

Figure S14—STOP Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Banhiran 2014	182	42	40	39	0.82 [0.76, 0.87]	0.48 [0.37, 0.60]
Chung 2008	81	21	43	32	0.65 [0.56, 0.74]	0.60 [0.46, 0.74]
Cowan 2014	94	27	3	5	0.97 [0.91, 0.99]	0.16 [0.05, 0.33]
Luo 2014	170	9	27	6	0.86 [0.81, 0.91]	0.40 [0.16, 0.68]
Pataka 2014	1427	296	74	56	0.95 [0.94, 0.96]	0.16 [0.12, 0.20]



Pooled sensitivity: 0.88 [0.77, 0.94]
Pooled specificity: 0.33 [0.18, 0.52]
LR+: 1.31 [1.10, 1.57]
LR-: 0.36 [0.27, 0.47]
DOR: 3.68 [2.80, 4.83]
Accuracy: 0.78 or **78%**

Figure S15—ROC Curve for STOP Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

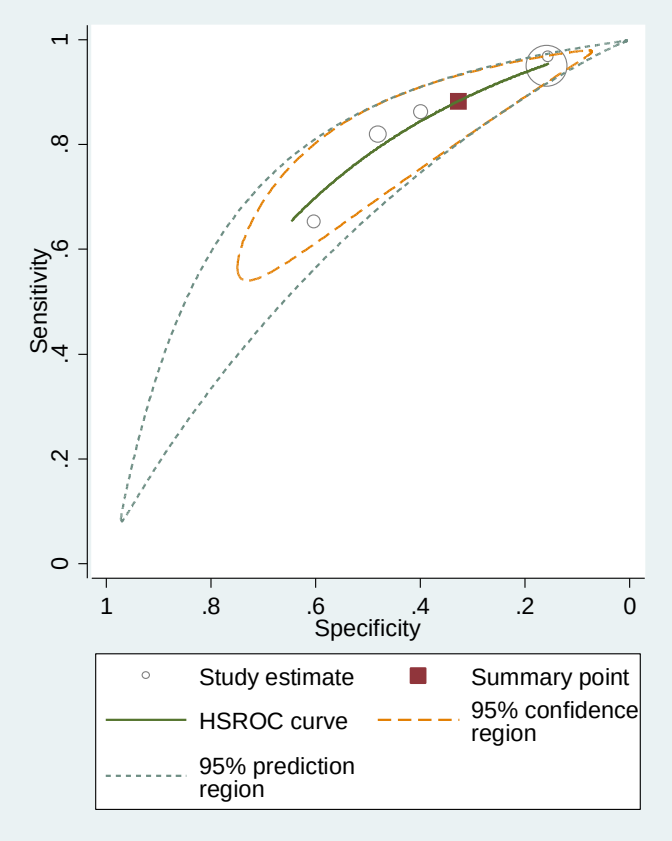


Table S21—Summary of Findings table for STOP Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Chung 2008 (A); Cowan 2014 (B); Pataka 2014 (C); Luo 2014 (D); Banhiran 2014 (E)

Pooled sensitivity STOP Questionnaire: 0.88 (95% CI: 0.77 to 0.94) | **Pooled specificity STOP Questionnaire**: 0.33 (95% CI: 0.18 to 0.52) **Pooled sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 81% (95% CI: 74 to 86%) **Accuracy (low risk):** 63% (95% CI: 60 to 67%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		STOP Questionnaire	Attended PSG	STOP Questionnaire	Attended PSG	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	766 (670 to 818)	870 (870 to 870)	484 (424 to 517)	550 (550 to 550)	2674 (5) [A-E]
		104 fewer TP in STOP Questionnaire		**66 fewer TP in STOP Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		104 (52 to 200)	0 (0 to 0)	66 (33 to 126)	0 (0 to 0)	
		104 more FN in STOP Questionnaire		**66 more FN in STOP Questionnaire**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	43 (23 to 68)	130 (130 to 130)	149 (81 to 234)	450 (450 to 450)	2674 (5) [A-E]
		87 fewer TN in STOP Questionnaire		**301 fewer TN in STOP Questionnaire**		
False positives (patients incorrectly classified as having OSA)		87 (62 to 107)	0 (0 to 0)	301 (216 to 369)	0 (0 to 0)	
		87 more FP in STOP Questionnaire		**301 more FP in STOP Questionnaire**		

[1]Broad range of specificity across studies
[2]Wide confidence intervals for specificity

Table S22—Summary of Findings table for STOP Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Chung 2008 (A); Pataka 2014 (B); Banhiran 2014 (C); Luo 2014 (D); Cowan (E)

Range of sensitivities STOP Questionnaire: 0.62 to 0.98 | Range of specificities STOP Questionnaire: 0.10 to 0.63 | Range of sensitivities Attended PSG: 1.00 to 1.00 | Range of specificities Attended PSG: 1.00 to 1.00 | **Accuracy (high risk)**: 60% to 79% | **Accuracy (low risk):** 45% to 48%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		STOP Questionnaire	Attended PSG	STOP Questionnaire	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	397 to 627	640 to 640	155 to 245	250 to 250	2674 (5) [A-E]
		13 to 243 fewer TP in STOP Questionnaire		**5 to 95 fewer TP in STOP Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		13 to 243	0 to 0	5 to 95	0 to 0	
		13 to 243 more FN in STOP Questionnaire		**5 to 95 more FN in STOP Questionnaire**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	36 to 227	360 to 360	75 to 473	750 to 750	2674 (5) [A-E]
		133 to 324 fewer TN in STOP Questionnaire		**277 to 675 fewer TN in STOP Questionnaire**		
False positives (patients incorrectly classified as having OSA)		103 to 324	0 to 0	277 to 675	0 to 0	
		133 to 324 more FP in STOP Questionnaire		**277 to 675 more FP in STOP Questionnaire**		

[1]Wide range of values for sensitivity and specificity

Table S23—Summary of Findings table for STOP Questionnaire vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Pataka 2014 (A); Banhiran 2014 (B); Luo 2014 (C)

Range of sensitivities STOP Questionnaire: 0.91 to 0.97 | Range of specificities STOP Questionnaire: 0.11 to 0.36 | Range of sensitivities Attended PSG: 1.00 to 1.00 | Range of specificities Attended PSG: 1.00 to 1.00 | **Accuracy (high risk)**: 48% to 49% | **Accuracy (low risk):** 25% to 34%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		STOP Questionnaire	Attended PSG	STOP Questionnaire	Attended PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	328 to 349	360 to 360	91 to 97	100 to 100	2368 (3) [A-C]
		11 to 32 fewer TP in STOP Questionnaire		**3 to 9 fewer TP in STOP Questionnaire**		
False negatives (patients incorrectly classified as not having OSA)		11 to 32	0 to 0	3 to 9	0 to 0	
		11 to 32 more FN in STOP Questionnaire		**3 to 9 more FN in STOP Questionnaire**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	70 to 230	640 to 640	99 to 324	900 to 900	2368 (3) [A-C]
		410 to 570 fewer TN in STOP Questionnaire		**576 to 801 fewer TN in STOP Questionnaire**		
False positives (patients incorrectly classified as having OSA)		410 to 570	0 to 0	576 to 801	0 to 0	
		410 to 570 more FP in STOP Questionnaire		**576 to 801 more FP in STOP Questionnaire**		

[1]Wide range of values for specificity

Table S24—Summary of Findings table for Morphometric Model vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Gurubhagavatula 2013 (A); Kushida 1997 (B)

Range of sensitivities Morphometric Model: 0.88 to 0.98 | **Range of specificities Morphometric Model**: 0.11 to 0.31
Range of sensitivities Attended PSG: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Morphometric Model	Attended PSG	Morphometric Model	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	766 to 853	870 to 870	484 to 539	550 to 550	350 (2) [A,B]
		17 to 104 fewer TP in Morphometric Model		**11 to 66 fewer TP in Morphometric Model**		
False negatives (patients incorrectly classified as not having OSA)		17 to 104	0 to 0	11 to 66	0 to 0	
		17 to 104 more FN in Morphometric Model		**11 to 66 more FN in Morphometric Model**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	14 to 40	130 to 130	49 to 139	450 to 450	350 (2) [A,B]
		90 to 116 fewer TN in Morphometric Model		**311 to 401 fewer TN in Morphometric Model**		
False positives (patients incorrectly classified as having OSA)		90 to 116	0 to 0	311 to 401	0 to 0	
		90 to 116 more FP in Morphometric Model		**311 to 401 more FP in Morphometric Model**		

[1]Wide range of values for specificity

Table S25—Summary of Findings table for Adjusted Neck Circumference vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Nicholl 2013 (A)

Range of sensitivities Adjusted Neck Circumference: 0.34 to 0.93 | **Range of specificities Adjusted Neck Circumference**: 0.37 to 0.94 **Range of sensitivities HSAT**: 1.00 to 1.00 | **Range of specificities HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 35% to 93% **Accuracy (low risk):** 36% to 94%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Adjusted Neck Circumference	HSAT	Adjusted Neck Circumference	HSAT	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	218 to 595	640 to 640	85 to 233	250 to 250	172 (1) [A]
		45 to 422 fewer TP in Adjusted Neck Circumference		**17 to 165 fewer TP in Adjusted Neck Circumference**		
False negatives (patients incorrectly classified as not having OSA)		45 to 422	0 to 0	17 to 165	0 to 0	
		45 to 422 more FN in Adjusted Neck Circumference		**17 to 165 more FN in Adjusted Neck Circumference**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	133 to 338	360 to 360	277 to 705	750 to 750	172 (1) [A]
		22 to 227 fewer TN in Adjusted Neck Circumference		**45 to 473 fewer TN in Adjusted Neck Circumference**		
False positives (patients incorrectly classified as having OSA)		22 to 227	0 to 0	45 to 473	0 to 0	
		22 to 227 more FP in Adjusted Neck Circumference		**45 to 473 more FP in Adjusted Neck Circumference**		

[1]Wide range of values for specificity and sensitivity

Table S26—Summary of Findings table for Adjusted Neck Circumference vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Nicholl 2013 (A); Gurubhagavatula 2013 (B)

Range of sensitivities Adjusted Neck Circumference: 0.40 to 0.96 | **Range of specificities Adjusted Neck Circumference**: 0.32 to 0.92 **Range of sensitivities HSAT**: 1.00 to 1.00 | **Range of specificities HSAT**: 1.00 to 1.00 **Accuracy (high risk)**: 35% to 94% **Accuracy (low risk):** 33% to 92%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Adjusted Neck Circumference	HSAT	Adjusted Neck Circumference	HSAT	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1,2]	144 to 346	360 to 360	40 to 96	100 to 100	422 (2) [A,B]
		14 to 216 fewer TP in Adjusted Neck Circumference		**4 to 60 fewer TP in Adjusted Neck Circumference**		
False negatives (patients incorrectly classified as not having OSA)		14 to 216	0 to 0	4 to 60	0 to 0	
		14 to 216 more FN in Adjusted Neck Circumference		**4 to 60 more FN in Adjusted Neck Circumference**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	205 to 589	640 to 640	288 to 828	900 to 900	422 (2) [A,B]
		51 to 435 fewer TN in Adjusted Neck Circumference		**72 to 612 fewer TN in Adjusted Neck Circumference**		
False positives (patients incorrectly classified as having OSA)		51 to 435	0 to 0	72 to 612	0 to 0	
		51 to 435 more FP in Adjusted Neck Circumference		**72 to 612 more FP in Adjusted Neck Circumference**		

[1]Wide range of values for specificity and sensitivity

Table S27—Summary of Findings table for Multivariable Apnea Prediction (MAP) vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Gurubhagavatula 2001 (A); Gurubhagavatula 2013 (B); Rofail 2010 (C); Wilson 2014 (D)

Range of sensitivities MAP: 0.68 to 0.85 | **Range of specificities MAP**: 0.56 to 0.92 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 66% to 81% **Accuracy (low risk):** 63% to 79%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		MAP	Attended PSG	MAP	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	592 to 739	870 to 870	374 to 468	550 to 550	683 (4) [A-D]
		131 to 278 fewer TP in MAP		**82 to 176 fewer TP in MAP**		
False negatives (patients incorrectly classified as not having OSA)		131 to 278	0 to 0	82 to 170	0 to 0	
		131 to 278 more FN in MAP		**82 to 176 more FN in MAP**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	73 to 120	130 to 130	252 to 414	450 to 450	683 (4) [A-D]
		10 to 57 fewer TN in MAP		**36 to 198 fewer TN in MAP**		
False positives (patients incorrectly classified as having OSA)		10 to 57	0 to 0	36 to 198	0 to 0	
		10 to 57 more FP in MAP		**36 to 198 more FP in MAP**		

[1]Wide range of values for specificity and sensitivity

Table S28—MAP vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Gurubhagavatula 2001 (A); Gurubhagavatula 2013 (B); Morales 2012 (C); Wilson 2014 (D)

Range of sensitivities MAP: 0.80 to 0.90 | **Range of specificities MAP**: 0.44 to 0.72 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 58% to 70% **Accuracy (low risk):** 50% to 50%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		MAP	Attended PSG	MAP	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	288 to 360	360 to 360	80 to 100	100 to 100	436 (4) [A-D]
		0 to 72 fewer TP in MAP		**10 to 20 fewer TP in MAP**		
False negatives (patients incorrectly classified as not having OSA)		0 to 72	0 to 0	0 to 20	0 to 0	
		36 to 72 more FN in MAP		**0 to 20 more FN in MAP**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	122 to 461	640 to 640	171 to 648	900 to 900	436 (4) [A-D]
		179 to 518 fewer TN in MAP		**252 to 729 fewer TN in MAP**		
False positives (patients incorrectly classified as having OSA)		179 to 518	0 to 0	252 to 729	0 to 0	
		179 to 518 more FP in MAP		**252 to 729 more FP in MAP**		

[1]Wide range of values for specificity

Table S29—Summary of Findings table for Prediction Models vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Chang 2014 (A); Zou 2013 (B)

Range of sensitivities Prediction Models: 0.33 to 0.90 | **Range of specificities Prediction Models**: 0.50 to 1.00 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 58% to 80% **Accuracy (low risk):** 61% to 88%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Prediction Models	Attended PSG	Prediction Models	Attended PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	287 to 783	870 to 870	182 to 495	550 to 550	1089 (2) [A,B]
		87 to 583 fewer TP in Prediction Models		**55 to 368 fewer TP in Prediction Models**		
False negatives (patients incorrectly classified as not having OSA)		87 to 583	0 to 0	55 to 368	0 to 0	
		87 to 583 more FN in Prediction Models		**55 to 368 more FN in Prediction Models**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	65 to 130	130 to 130	225 to 450	450 to 450	1089 (2) [A,B]
		0 to 65 fewer TN in Prediction Models		**0 to 225 fewer TN in Prediction Models**		
False positives (patients incorrectly classified as having OSA)		0 to 65	0 to 0	0 to 225	0 to 0	
		0 to 65 more FP in Prediction Models		**0 to 225 more FP in Prediction Models**		

[1]Wide values of sensitivity and specificity

Table S30—Summary of Findings table for Prediction Models vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Sharma 2006 (A); Zerah-Lancner 2000 (B)

Range of sensitivities Prediction Models: 0.82 to 1.00 | **Range of specificities Prediction Models**: 0.84 to 0.91 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 85% to 96% **Accuracy (low risk):** 85% to 92%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Prediction Models	Attended PSG	Prediction Models	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	525 to 640	640 to 640	205 to 250	250 to 250	287 (2) [A,B]
		0 to 115 fewer TP in Prediction Models		**0 to 45 fewer TP in Prediction Models**		
False negatives (patients incorrectly classified as not having OSA)		0 to 115	0 to 0	0 to 45	0 to 0	
		0 to 115 more FN in Prediction Models		**0 to 45 more FN in Prediction Models**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	302 to 328	360 to 360	630 to 683	750 to 750	287 (2) [A,B]
		32 to 58 fewer TN in Prediction Models		**67 to 120 fewer TN in Prediction Models**		
False positives (patients incorrectly classified as having OSA)		32 to 58	0 to 0	67 to 120	0 to 0	
		32 to 58 more FP in Prediction Models		**67 to 120 more FP in Prediction Models**		

Table S31—Summary of Findings table for Prediction Models vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Platt 2013 (A); Morales 2012 (B); Kolotkin 2011 (C)

Range of sensitivities Prediction Models: 0.76 to 0.97 | **Range of specificities Prediction Models**: 0.19 to 0.75 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 58% to 70% **Accuracy (low risk):** 51% to 52%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Prediction Models	HSAT	Prediction Models	HSAT	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	274 to 349	360 to 360	76 to 97	100 to 100	697 (3) [A-C]
		11 to 86 fewer TP in Prediction Models		**3 to 24 fewer TP in Prediction Models**		
False negatives (patients incorrectly classified as not having OSA)		11 to 86	0 to 0	3 to 24	0 to 0	
		11 to 86 more FN in Prediction Models		**3 to 24 more FN in Prediction Models**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	122 to 480	640 to 640	171 to 675	900 to 900	697 (3) [A-C]
		160 to 518 fewer TN in Prediction Models		**225 to 729 fewer TN in Prediction Models**		
False positives (patients incorrectly classified as having OSA)		160 to 518	0 to 0	225 to 729	0 to 0	
		160 to 518 more FP in Prediction Models		**225 to 729 more FP in Prediction Models**		

[1]Wide range of values for specificity

Table S32—Summary of Findings table for OSA 50 vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Firat 2012 (A)

Single study sensitivity OSA 50: 0.63 (95% CI: 0.49 to 0.77) | **Single study specificity OSA 50**: 0.82 (95% CI: 0.70 to 0.94) **Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 70% (95% CI: 57 to 83%) **Accuracy (low risk):** 77% (95% CI: 65 to 90%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		OSA 50	Attended PSG	OSA 50	Attended PSG	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	403 (314 to 493)	640 (640 to 640)	158 (123 to 193)	250 (250 to 250)	85 (1) [A]
		237 fewer TP in OSA 50		**92 fewer TP in OSA 50**		
False negatives (patients incorrectly classified as not having OSA)		237 (147 to 326)	0 (0 to 0)	92 (57 to1 27)	0 (0 to 0)	
		237 more FN in OSA 50		**92 more FN in OSA 50**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	295 (252 to 338)	360 (360 to 360)	615 (525 to 705)	750 (750 to 750)	85 (1) [A]
		65 fewer TN in OSA 50		**135 fewer TN in OSA 50**		
False positives (patients incorrectly classified as having OSA)		65 (22 to 108)	0 (0 to 0)	135 (45 to 225)	0 (0 to 0)	
		65 more FP in OSA 50		**135 more FP in OSA 50**		

[1]Indirect evidence as study only included highway bus drivers
[2]Wide confidence intervals for sensitivity and specificity

Table S33—Summary of Findings table for OSA 50 vs. HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Chai-Coetzer 2011 (A)

Single study sensitivity OSA 50: 0.88 (95% CI: 0.60 to 0.98) | **Single study specificity OSA 50**: 0.82 (95% CI: 0.70 to 0.90)
Single study sensitivity HSAT: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity HSAT**: 1.00 (95% CI: 1.00 to 1.00)
Accuracy (high risk): 84% (95% CI: 66 to 93%) **Accuracy (low risk):83% (95% CI:** 69 to 91%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		OSA 50	HSAT	OSA 50	HSAT	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	317 (216 to 353)	360 (360 to 360)	88 (60 to 98)	100 (100 to 100)	78 (1) [A]
		43 fewer TP in OSA 50		**12 fewer TP in OSA 50**		
False negatives (patients incorrectly classified as not having OSA)		43 (7 to144)	0 (0 to 0)	12 (2 to 40)	0 (0 to 0)	
		43 more FN in OSA 50		**12 more FN in OSA 50**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	525 (448 to 576)	640 (640 to 640)	738 (630 to 810)	900 (900 to 900)	78 (1) [A]
		115 fewer TN in OSA 50		**162 fewer TN in OSA 50**		
False positives (patients incorrectly classified as having OSA)		115 (64 to192)	0 (0 to 0)	162 (90 to 270)	0 (0 to 0)	
		115 more FP in OSA 50		**162 more FP in OSA 50**		

[1]Wide confidence intervals for sensitivity and specificity

Table S34—Summary of Findings table for Clinical Decision Support System vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: LaPorta 2012 (A)

Single study sensitivity Clinical Decision Support System: 0.98 (95% CI: 0.92 to 1.00) | **Single study specificity Clinical Decision Support System**: 0.87 (95% CI: 0.66 to 0.97) **Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 97% (95% CI: 89 to 100%) **Accuracy (low risk):** 93% (95% CI: 80 to 99%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Clinical Decision Support System	Attended PSG	Clinical Decision Support System	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	853 (800 to 870)	870 (870 to 870)	539 (506 to 550)	550 (550 to 550)	91 (1) [A]
		17 fewer TP in Clinical Decision Support System		**11 fewer TP in Clinical Decision Support System**		
False negatives (patients incorrectly classified as not having OSA)		17 (0 to 70)	0 (0 to 0)	11 (0 to44)	0 (0 to 0)	
		17 more FN in Clinical Decision Support System		**11 more FN in Clinical Decision Support System**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	113 (86 to 126)	130 (130 to 130)	391 (297 to 436)	450 (450 to 450)	91 (1) [A]
		17 fewer TN in Clinical Decision Support System		**59 fewer TN in Clinical Decision Support System**		
False positives (patients incorrectly classified as having OSA)		17 (4 to 44)	0 (0 to 0)	59 (14 to 153)	0 (0 to 0)	
		17 more FP in Clinical Decision Support System		**59 more FP in Clinical Decision Support System**		

[1]Indirect evidence as study only included patients with ischemic heart disease
[2]Wide confidence intervals for specificity

Table S35—Summary of Findings table for (Obstructive Sleep Apnea Hypopnea Syndrome) OSAHS Score vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Friedman 2010 (A)

Single study sensitivity OSAHS Score: 0.86 (95% CI: 0.80 to 0.91) | **Single study specificity OSAHS Score**: 0.47 (95% CI: 0.34 to 0.56) |**Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 81% (95% CI: 74 to 86%) **Accuracy (low risk):** 68% (95% CI: 59 to 75%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		OSAHS Score	Attended PSG	OSAHS Score	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	748 (696 to 792)	870 (870 to 870)	473 (440 to 501)	550 (550 to 550)	223 (1) [A]
		122 fewer TP in OSAHS Score		**77 fewer TP in OSAHS Score**		
False negatives (patients incorrectly classified as not having OSA)		122 (78 to174)	0 (0 to 0)	77 (49 to110)	0 (0 to 0)	
		122 more FN in OSAHS Score		**77 more FN in OSAHS Score**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	61 (44 to 73)	130 (130 to 130)	211 (153 to 252)	450 (450 to 450)	223 (1) [A]
		69 fewer TN in OSAHS Score		**239 fewer TN in OSAHS Score**		
False positives (patients incorrectly classified as having OSA)		69 (57 to 86)	0 (0 to 0)	239 (198 to 297)	0 (0 to 0)	
		69 more FP in OSAHS Score		**239 more FP in OSAHS Score**		

[1]Wide confidence intervals for specificity

Table S36—Summary of Findings table for Kushida Index vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Kushida 1997 (A)

Single study sensitivity Kushida Index: 0.98 (95% CI: 0.95 to 0.99) | **Single study specificity Kushida Index**: 1.00 (95% CI: 0.92 to 1.00) | **Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 98% (95% CI: 95 to 99%) **Accuracy (low risk):** 99% (95% CI: 94 to 100%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Kushida Index	Attended PSG	Kushida Index	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	853 (827 to 861)	870 (870 to 870)	539 (523 to 545)	550 (550 to 550)	301 (1) [A]
		17 fewer TP in Kushida Index		**11 fewer TP in Kushida Index**		
False negatives (patients incorrectly classified as not having OSA)		17 (9 to43)	0 (0 to 0)	11 (5 to 27)	0 (0 to 0)	
		17 more FN in Kushida Index		**11 more FN in Kushida Index**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	130 (120 to 130)	130 (130 to 130)	450 (414 to 450)	450 (450 to 450)	301 (1) [A]
		0 fewer TN in Kushida Index		**0 fewer TN in Kushida Index**		
False positives (patients incorrectly classified as having OSA)		0 (0 to10)	0 (0 to 0)	0 (0 to 36)	0 (0 to 0)	
		0 fewer FP in Kushida Index		**0 fewer FP in Kushida Index**		

Home sleep apnea testing for the diagnosis of obstructive sleep apnea in adults

Table S37—Summary of Findings table for Type 2 HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Campbell 2011 (A); Banhiran 2014 (B)

Range of sensitivities Type 2 HSAT: 0.88 to 0.97 | **Range of specificities Type 2 HSAT**: 0.50 to 0.56
Range of sensitivities Attended: 1.00 to 1.00 | **Range of specificities Attended**: 1.00 to 1.00 **Accuracy (high risk)**: 84% to 91% **Accuracy (low risk):** 73% to 77%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Type 2 HSAT	Attended	Type 2 HSAT	Attended	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	766 to 844	870 to 870	484 to 534	550 to 550	116 (2) [A,B]
		26 to 104 fewer TP in Type 2 HSAT		**16 to 66 fewer TP in Type 2 HSAT**		
False negatives (patients incorrectly classified as not having OSA)		26 to 104	0 to 0	16 to 66	0 to 0	
		26 to 104 more FN in Type 2 HSAT		**16 to 66 more FN in Type 2 HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	65 to 73	130 to 130	225 to 252	450 to 450	116 (2) [A,B]
		57 to 65 fewer TN in Type 2 HSAT		**198 to 225 fewer TN in Type 2 HSAT**		
False positives (patients incorrectly classified as having OSA)		57 to 65	0 to 0	198 to 225	0 to 0	
		57 to 65 more FP in Type 2 HSAT		**198 to 225 more FP in Type 2 HSAT**		

Table S38—Summary of Findings table for Type 2 HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Campbell 2011 (A); Banhiran 2014 (B)

Range of sensitivities Type 2 HSAT: 0.94 to 0.95 | **Range of specificities Type 2 HSAT**: 0.76 to 0.77
Range of sensitivities Attended: 1.00 to 1.00 | **Range of specificities Attended**: 1.00 to 1.00 **Accuracy (high risk)**: 88% to 88% **Accuracy (low risk):** 81% to 81%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Type 2 HSAT	Attended	Type 2 HSAT	Attended	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	602 to 608	640 to 640	235 to 238	250 to 250	116 (2) [A,B]
		32 to 38 fewer TP in Type 2 HSAT		**12 to 15 fewer TP in Type 2 HSAT**		
False negatives (patients incorrectly classified as not having OSA)		32 to 38	0 to 0	12 to 15	0 to 0	
		32 to 38 more FN in Type 2 HSAT		**12 to 15 more FN in Type 2 HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	274 to 277	360 to 360	570 to 578	250 to 250	116 (2) [A,B]
		83 to 86 fewer TN in Type 2 HSAT		**172 to 180 fewer TN in Type 2 HSAT**		
False positives (patients incorrectly classified as having OSA)		83 to 86	0 to 0	172 to 180	0 to 0	
		83 to 86 more FP in Type 2 HSAT		**172 to 180 more FP in Type 2 HSAT**		

Table S39—Summary of Findings table for Type 3 HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Gjevre 2011 (A); Masa 2011 Thorax (B); Polese 2012 (C); Santos-Silva 2009 (D); Yin 2006 (E); Planes 2010 (F); Masa 2013 (G)

Range of sensitivities Type 3 HSAT: 0.90 to 1 | **Range of specificities Type 3 HSAT**: 0.30 to 0.67
Range of sensitivities Attended: 1.00 to 1.00 | **Range of specificities Attended**: 1.00 to 1.00 **Accuracy (high risk)**: 84% to 91% **Accuracy (low risk):** 70% to 78%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Type 3 HSAT	Attended	Type 3 HSAT	Attended	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	783 to 870	870 to 870	495 to 550	550 to 550	1001 (7) [A-G]
		0 to 87 fewer TP in Type 3 HSAT		**0 to 55 fewer TP in Type 3 HSAT**		
False negatives (patients incorrectly classified as not having OSA)		0 to 87	0 to 0	0 to 55	0 to 0	
		0 to 87 more FN in Type 3 HSAT		**0 to 55 more FN in Type 3 HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	39 to 87	130 to 130	135 to 302	450 to 450	1001 (7) [A-G]
		43 to 91 fewer TN in Type 3 HSAT		**148 to 315 fewer TN in Type 3 HSAT**		
False positives (patients incorrectly classified as having OSA)		43 to 91	0 to 0	180 to 315	0 to 0	
		43 to 91 more FP in Type 3 HSAT		**148 to 315 more FP in Type 3 HSAT**		

[1]Wide range of values for specificity

Table S40—Summary of Findings table for Type 3 HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

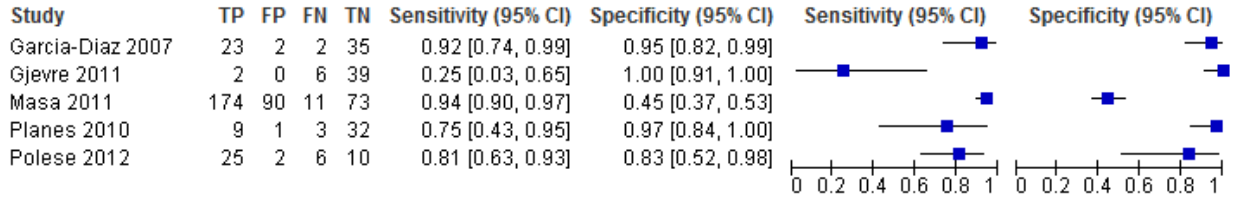
References: Garcia-Diaz 2007 (A); Gjevre 2011 (B); Polese 2012 (C); Santo Silva 2009 (D); Yin 2006 (E); Planes 2010 (F)

Range of sensitivities Type 3 HSAT: 0.62 to 0.94 | **Range of specificities Type 3 HSAT**: 0.25 to 0.97
Range of sensitivities Attended PSG: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 65% to 91% **Accuracy (low risk):** 59% to 90%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Type 3 HSAT	Attended PSG	Type 3 HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	397 to 602	640 to 640	155 to 235	250 to 250	457 (6) [A-F]
		38 to 243 fewer TP in Type 3 HSAT		**15 to 95 fewer TP in Type 3 HSAT**		
False negatives (patients incorrectly classified as not having OSA)		38 to 243	0 to 0	15 to 95	0 to 0	
		38 to 243 more FN in Type 3 HSAT		**15 to 95 more FN in Type 3 HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	90 to 349	360 to 360	188 to 728	750 to 750	457 (6) [A-F]
		11 to 270 fewer TN in Type 3 HSAT		**22 to 562 fewer TN in Type 3 HSAT**		
False positives (patients incorrectly classified as having OSA)		11 to 270	0 to 0	22 to 562	0 to 0	
		11 to 270 more FP in Type 3 HSAT		**22 to 562 more FP in Type 3 HSAT**		

[1]Wide range of values for specificity and sensitivity

Figure S16—Type 3 HSAT vs. PSG (AHI ≥ 30)



Pooled sensitivity: 0.87 [0.77, 0.93]
Pooled specificity: 0.88 [0.59, 0.97]
DOR: 49.0 [13.9, 172.2]
LR+: 7.06 [1.88, 26.6]
LR-: 0.14 [0.08, 0.25]
Accuracy: 0.77 or **77%**

Figure S17—ROC Curve for Type 3 HSAT vs. PSG (AHI ≥ 30)

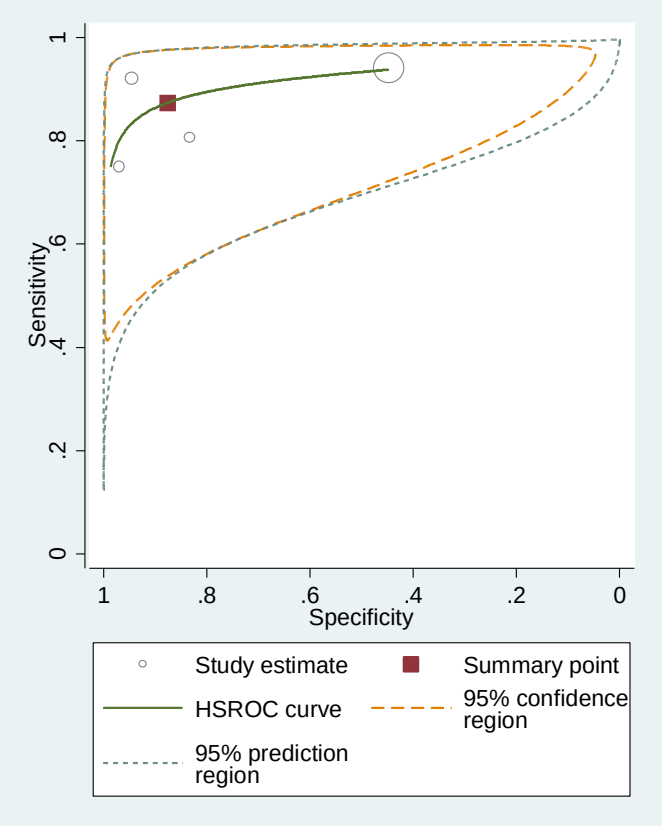


Table S41—Summary of Findings table for Type 3 HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Garcia-Diaz 2007 (A); Gjevre 2011 (B); Masa 2011 (C); Planes 2010 (D); Polese 2012 (E)

Pooled sensitivity Type 3 HSAT: 0.87 (95% CI: 0.77 to 0.93) | **Pooled specificity Type 3 HSAT**: 0.88 (95% CI: 0.59 to 0.97)
Pooled sensitivity Attended PSG: 1.00 (95% CI: 1.00 to 1.00) | **Pooled specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00)
Accuracy (high risk): 88% (95% CI: 81 to 94%) **Accuracy (low risk):** 88% (95% CI: 71 to 95%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Type 3 HSAT	Attended PSG	Type 3 HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	313 (277 to 335)	360 (360 to 360)	87 (77 to 93)	100 (100 to 100)	545 (5) [A-E]
		47 fewer TP in Type 3 HSAT		**13 fewer TP in Type 3 HSAT**		
False negatives (patients incorrectly classified as not having OSA)		47 (25 to 83)	0 (0 to 0)	13 (7 to 23)	0 (0 to 0)	
		47 more FN in Type 3 HSAT		**13 more FN in Type 3 HSAT**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	563 (378 to 621)	640 (640 to 640)	792 (531 to 873)	900 (900 to 900)	545 (5) [A-E]
		77 fewer TN in Type 3 HSAT		**108 fewer TN in Type 3 HSAT**		
False positives (patients incorrectly classified as having OSA)		77 (19 to 262)	0 (0 to 0)	108 (27 to 369)	0 (0 to 0)	
		77 more FP in Type 3 HSAT		**108 more FP in Type 3 HSAT**		

[1]Wide confidence intervals for sensitivity and specificity
[2]Wide range of values for sensitivity and specificity

Table S42—Summary of Findings table for 2-3 Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Ayappa 2008 (A); Tonelli de Oliveria 2009 (B); Ward 2015 (C)

Range of sensitivities 2-3 Channel HSAT: 0.80 to 0.96 | **Range of specificities 2-3 Channel HSAT**: 0.65 to 0.83 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 81% to 93% **Accuracy (low risk):** 77% to 88%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		2-3 Channel HSAT	Attended PSG	2-3 Channel HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	696 to 835	870 to 870	440 to 528	550 to 550	292 (3) [A-C]
		35 to 174 fewer TP in 2-3 Channel HSAT		**22 to 110 fewer TP in 2-3 Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		35 to 174	0 to 0	22 to 110	0 to 0	
		35 to 174 more FN in 2-3 Channel HSAT		**22 to 110 more FN in 2-3 Channel HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	85 to 108	130 to 130	293 to 373	450 to 450	292 (3) [A-C]
		22 to 45 fewer TN in 2-3 Channel HSAT		**77 to 157 fewer TN in 2-3 Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		22 to 45	0 to 0	77 to 157	0 to 0	
		22 to 45 more FP in 2-3 Channel HSAT		**77 to 157 more FP in 2-3 Channel HSAT**		

[1]Wide range of sensitivity

Table S43—Summary of Findings table for 2-3 Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Ayappa 2008 (A); Baltzan 2000 (B); Masdeu 2010 (C); Tonelli de Oliveria 2009 (D); Ward 2015 (E)

Range of sensitivities 2-3 Channel HSAT: 0.66 to 0.88 | **Range of specificities 2-3 Channel HSAT**: 0.62 to 1.00 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 72% to 87% **Accuracy (low risk):** 68% to 95%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		2-3 Channel HSAT	Attended PSG	2-3 Channel HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	422 to 563	640 to 640	165 to 220	250 to 250	443 (5) [A-E]
		77 to 218 fewer TP in 2-3 Channel HSAT		**30 to 85 fewer TP in 2-3 Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		77 to 218	0 to 0	30 to 85	0 to 0	
		77 to 218 more FN in 2-3 Channel HSAT		**30 to 85 more FN in 2-3 Channel HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	223 to 360	360 to 360	465 to 750	750 to 750	443 (5) [A-E]
		0 to 137 fewer TN in 2-3 Channel HSAT		**0 to 285 fewer TN in 2-3 Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		22 to 137	0 to 0	0 to 285	0 to 0	
		0 to 137 more FP in 2-3 Channel HSAT		**0 to 285 more FP in 2-3 Channel HSAT**		

[1]Wide range of sensitivity and specificity

Table S44—Summary of Findings table for 2-3 Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Tonelli de Oliveria 2009 (A); Ward 2015 (B)

Range of sensitivities 2-3 Channel HSAT: 0.78 to 0.90 | **Range of specificities 2-3 Channel HSAT**: 0.92 to 0.98 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 71% to 90% **Accuracy (low risk):** 88% to 91%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		2-3 Channel HSAT	Attended PSG	2-3 Channel HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	155 to 288	360 to 360	43 to 80	100 to 100	225 (2) [A-B]
		72 to 205 fewer TP in 2-3 Channel HSAT		**20 to 57 fewer TP in 2-3 Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		72 to 205	0 to 0	20 to 57	0 to 0	
		72 to 205 more FN in 2-3 Channel HSAT		**20 to 57 more FN in 2-3 Channel HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	589 to 627	640 to 640	828 to 882	900 to 900	225 (2) [A-B]
		13 to 51 fewer TN in 2-3 Channel HSAT		**18 to 72 fewer TN in 2-3 Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		13 to 51	0 to 0	18 to 72	0 to 0	
		13 to 51 more FP in 2-3 Channel HSAT		**18 to 72 more FP in 2-3 Channel HSAT**		

Table S45—Summary of Findings table for 2-3 Channel HSAT vs. In-home PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Gantner 2010 (A)

Single study sensitivity 2-3 Channel HSAT: 0.88 (95% CI: 0.80 to 0.93) | **Single study specificity 2-3 Channel HSAT**: 0.84 (95% CI: 0.69 to 0.93) **Single study sensitivity In-home PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity In-home PSG**: 1.00 (95% CI: 1.00 to 1.00) 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 86% (95% CI: 76 to 93%) **Accuracy (low risk):** 85% (95% CI: 72 to 93%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		2-3 Channel HSAT	In-home PSG	2-3 Channel HSAT	In-home PSG	
True positives (patients with OSA)	⊕⊕◯◯ LOW[1,2]	563 (512 to 595)	640 (640 to 640)	220 (200 to 233)	250 (250 to 250)	143 (1) [A]
		77 fewer TP in 2-3 Channel HSAT		**30 fewer TP in 2-3 Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		77 (45 to128)	0 (0 to 0)	30 (50 to 17)	0 (0 to 0)	
		77 more FN in 2-3 Channel HSAT		**30 more FN in 2-3 Channel HSAT**		
True negatives (patients without OSA)	⊕⊕◯◯ LOW[1,2]	302 (248 to 335)	360 (360 to 360)	630 (518 to 698)	750 (750 to 750)	143 (1) [A]
		58 fewer TN in 2-3 Channel HSAT		**120 fewer TN in 2-3 Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		58 (25 to 112)	0 (0 to 0)	120 (52 to 232)	0 (0 to 0)	
		58 more FP in 2-3 Channel HSAT		**120 more FP in 2-3 Channel HSAT**		

[1]Indirect evidence as study only included Chinese population at high cardiovascular risk
[2]Wide confidence interval for specificity

Table S46—Summary of Findings table for 2-3 Channel HSAT vs. in-home PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Chai-Coetzer 2010 (A); Gantner 2010 (B)

Range of sensitivities 2-3 Channel HSAT: 0.84 to 0.97 | **Range of specificities 2-3 Channel HSAT**: 0.82 to 0.87 **Range of sensitivities In-home PSG**: 1.00 to 1.00 | **Range of specificities In-home PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 83% to 91% **Accuracy (low risk):** 82% to 88%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		2-3 Channel HSAT	In-home PSG	2-3 Channel HSAT	In-home PSG	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	302 to 349	360 to 360	84 to 97	100 to 100	300 (2) [A,B]
		11 to 58 fewer TP in 2-3 Channel HSAT		**3 to 16 fewer TP in 2-3 Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		11 to 58	0 to 0	3 to 16	0 to 0	
		11 to 58 more FN in 2-3 Channel HSAT		**3 to 16 more FN in 2-3 Channel HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	525 to 557	640 to 640	738 to 783	900 to 900	300 (2) [A,B]
		83 to 115 fewer TN in 2-3 Channel HSAT		**117 to 162 fewer TN in 2-3 Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		83 to 115	0 to 0	117 to 162	0 to 0	
		83 to 115 more FP in 2-3 Channel HSAT		**117 to 162 more FP in 2-3 Channel HSAT**		

Table S47—Summary of Findings table for Single Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Nakano 2008 (A)

Single study sensitivity Single Channel HSAT: 0.96 (95% CI: 0.91 to 1.00) | **Single study specificity Single Channel HSAT**: 0.82 (95% CI: 0.60 to 1.00) **Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 94% (95% CI: 87 to 100%) **Accuracy (low risk):** 90% (95% CI: 77 to 100%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Single Channel HSAT	Attended PSG	Single Channel HSAT	Attended PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	835 (792 to 870)	870 (870 to 870)	528 (501 to 550)	550 (550 to 550)	100 (1) [A]
		35 fewer TP in Single Channel HSAT		**22 fewer TP in Single Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		35 (0 to78)	0 (0 to 0)	22 (0 to 49)	0 (0 to 0)	
		35 more FN in Single Channel HSAT		**22 more FN in Single Channel HSAT**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	107 (78 to 130)	130 (130 to 130)	369 (270 to 450)	450 (450 to 450)	100 (1) [A]
		23 fewer TN in Single Channel HSAT		**81 fewer TN in Single Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		23 (0 to 52)	0 (0 to 0)	81 (0 to180)	0 (0 to 0)	
		23 more FP in Single Channel HSAT		**81 more FP in Single Channel HSAT**		

[1]Wide confidence interval for specificity

Table S48—Summary of Findings table for Single Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Nakano 2008 (A); Ozmen 2011 (B); Pang 2006 (C); Watkins 2009 (D)

Range of sensitivities Single-Channel HSAT: 0.55 to 0.91 | **Range of specificities Single-Channel HSAT**: 0.70 to 0.82 **Range of sensitivities Attended PSG**: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 60% to 88% **Accuracy (low risk):** 66% to 84%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Single-Channel HSAT	Attended PSG	Single-Channel HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	352 to 582	640 to 640	138 to 228	250 to 250	235 (4) A-D
		58 to 288 fewer TP in Single-Channel HSAT		**22 to 112 fewer TP in Single-Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		58 to 288	0 to 0	22 to 112	0 to 0	
		58 to 288 more FN in Single-Channel HSAT		**22 to 112 more FN in Single-Channel HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	252 to 295	360 to 360	525 to 615	750 to 750	235 (4) A-D
		65 to108 fewer TN in Single-Channel HSAT		**135 to 225 fewer TN in Single-Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		65 to 108	0 to 0	135 to 225	0 to 0	
		65 to108 more FP in Single-Channel HSAT		**135 to 225 more FP in Single-Channel HSAT**		

[1]Wide range of values for sensitivity

Table S49—Summary of Findings table for Single Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Nakano 2008 (A)

Single study sensitivity Single Channel HSAT: 0.89 (95% CI: 0.80 to 0.97) | **Single study specificity Single Channel HSAT**: 0.96 (95% CI: 0.90 to 1.00) **Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 93% (95% CI: 86 to 99%) **Accuracy (low risk):** 95% (95% CI: 88 to 100%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Single Channel HSAT	Attended PSG	Single Channel HSAT	Attended PSG	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	320 (288 to 349)	360 (360 to 360)	89 (80 to 97)	100 (100 to 100)	100 (1) [A]
		40 fewer TP in Single Channel HSAT		**11 fewer TP in Single Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		40 (11 to 72)	0 (0 to 0)	11 (3 to 20)	0 (0 to 0)	
		40 more FN in Single Channel HSAT		**11 more FN in Single Channel HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	614 (576 to 640)	640 (640 to 640)	864 (810 to 900)	900 (900 to 900)	100 (1) [A]
		26 fewer TN in Single Channel HSAT		**36 fewer TN in Single Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		26 (0 to 64)	0 (0 to 0)	36 (0 to 90)	0 (0 to 0)	
		26 more FP in Single Channel HSAT		**36 more FP in Single Channel HSAT**		

Table S50—Summary of Findings table for Other Single-Channel HSAT vs. PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Rofail 2010 (A)

Single study sensitivity Other Single-Channel HSAT: 0.80 (95% CI: 0.67 to 0.93) | **Single study specificity Other Single-Channel HSAT**: 0.87 (95% CI: 0.77 to 0.97) **Single study sensitivity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Attended PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 81% (95% CI: 68 to 94%) **Accuracy (low risk):** 83% (95% CI: 72 to 95%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Other Single-Channel HSAT	Attended PSG	Other Single-Channel HSAT	Attended PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	696 (583 to 809)	870 (870 to 870)	440 (369 to 512)	550 (550 to 550)	92 (1) [A]
		174 fewer TP in Other Single-Channel HSAT		**110 fewer TP in Other Single-Channel HSAT**		
False negatives (patients incorrectly classified as not having OSA)		174 (61 to 287)	0 (0 to 0)	110 (38 to 181)	0 (0 to 0)	
		174 more FN in Other Single-Channel HSAT		**110 more FN in Other Single-Channel HSAT**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	113 (100 to 126)	130 (130 to 130)	391 (347 to 436)	450 (450 to 450)	92 (1) [A]
		17 fewer TN in Other Single-Channel HSAT		**59 fewer TN in Other Single-Channel HSAT**		
False positives (patients incorrectly classified as having OSA)		17 (4 to 30)	0 (0 to 0)	59 (14 to 103)	0 (0 to 0)	
		17 more FP in Other Single-Channel HSAT		**59 more FP in Other Single-Channel HSAT**		

[1]Wide confidence intervals for specificity and sensitivity

Table S51—Summary of Findings table for Oximetry vs. In-home PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Chung 2012 (A)

Single study sensitivity Oximetry : 0.70 (95% CI: 0.66 to 0.76) | **Single study specificity Oximetry** : 0.90 (95% CI: 0.85 to 0.94)
Single study sensitivity In-home PSG: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity In-home PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 73% (95% CI: 68 to 78%) **Accuracy (low risk):** 79% (95% CI: 74 to 84%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Oximetry	In-home PSG	Oximetry	In-home PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	609 (574 to 661)	870 (870 to 870)	385 (363 to 418)	550 (550 to 550)	243 (1) [A]
		261 fewer TP in Oximetry		**165 fewer TP in Oximetry**		
False negatives (patients incorrectly classified as not having OSA)		261 (209 to 296)	0 (0 to 0)	165 (132 to187)	0 (0 to 0)	
		261 more FN in Oximetry		**165 more FN in Oximetry**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	117 (111 to 122)	130 (130 to 130)	405 (382 to 423)	450 (450 to 450)	243 (1) [A]
		13 fewer TN in Oximetry		**45 fewer TN in Oximetry**		
False positives (patients incorrectly classified as having OSA)		13 (8 to19)	0 (0 to 0)	45 (27 to 68)	0 (0 to 0)	
		13 more FP in Oximetry		**45 more FP in Oximetry**		

[1]Indirect evidence as study only included patients scheduled for inpatient surgery

Table S52—Summary of Findings table for Oximetry vs. In-home PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Chung 2012 (A)

Single study sensitivity Oximetry: 0.93 (95% CI: 0.90 to 0.97) | **Single study specificity Oximetry**: 0.75 (95% CI: 0.70 to 0.80)
Single study sensitivity In-home PSG: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity In-home PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 86% (95% CI: 83 to 91%) **Accuracy (low risk):** 80% (95% CI: 75 to 84%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Oximetry	In-home PSG	Oximetry	In-home PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	595 (576 to 621)	640 (640 to 640)	233 (225 to 243)	250 (250 to 250)	243 (1) [A]
		45 fewer TP in Oximetry		**17 fewer TP in Oximetry**		
False negatives (patients incorrectly classified as not having OSA)		45 (19 to 64)	0 (0 to 0)	17 (7 to 25)	0 (0 to 0)	
		45 more FN in Oximetry		**17 more FN in Oximetry**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	270 (252 to 288)	360 (360 to 360)	563 (525 to 600)	750 (750 to 750)	243 (1) [A]
		90 fewer TN in Oximetry		**187 fewer TN in Oximetry**		
False positives (patients incorrectly classified as having OSA)		90 (72 to108)	0 (0 to 0)	187 (150 to 225)	0 (0 to 0)	
		90 more FP in Oximetry		**187 more FP in Oximetry**		

[1]Indirect evidence as study included patients scheduled for inpatient surgery

Table S53—Summary of Findings table for Oximetry vs. In-home PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Chung 2012 (A)

Single study sensitivity Oximetry: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity Oximetry**: 0.59 (95% CI: 0.54 to 0.63)
Single study sensitivity In-home PSG: 1.00 (95% CI: 1.00 to 1.00) | **Single study specificity In-home PSG**: 1.00 (95% CI: 1.00 to 1.00) **Accuracy (high risk)**: 74% (95% CI: 71 to 76%) **Accuracy (low risk):** 63% (95% CI: 59 to 67%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Oximetry	In-home PSG	Oximetry	In-home PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	360 (360 to 360)	360 (360 to 360)	100 (100 to 100)	100 (100 to 100)	243 (1) [A]
		0 fewer TP in Oximetry		**0 fewer TP in Oximetry**		
False negatives (patients incorrectly classified as not having OSA)		0 (0 to 0)	0 (0 to 0)	0 (0 to 0)	0 (0 to 0)	
		0 fewer FN in Oximetry		**0 fewer FN in Oximetry**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	378 (346 to 403)	640 (640 to 640)	531 (486 to 567)	900 (900 to 900)	243 (1) [A]
		262 fewer TN in Oximetry		**369 fewer TN in Oximetry**		
False positives (patients incorrectly classified as having OSA)		262 (237 to 294)	0 (0 to 0)	369 (333 to 414)	0 (0 to 0)	
		262 more FP in Oximetry		**369 more FP in Oximetry**		

[1]Indirect evidence as study includes patients scheduled for inpatient surgery

Table S54—Summary of Findings table for Watch-Peripheral Arterial Tone (Watch-PAT) vs. In-Home PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: O'Brien 2012 (A)

Single study sensitivity Watch-PAT: 0.88 (95% CI: 0.47 to 1.00) | **Single study specificity Watch-PAT**: 0.87 (95% CI: 0.66 to 0.97) **Accuracy (high risk)**: 88% (95% CI: 50 to 100%) **Accuracy (low risk):** 88% (95% CI: 55 to 99%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Watch-PAT	In-Home PSG	Watch-PAT	In-Home PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	766 (409 to 870)	870 (870 to 870)	484 (259 to 550)	550 (550 to 550)	31 (1) [A]
		104 fewer TP in Watch-PAT		**66 fewer TP in Watch-PAT**		
False negatives (patients incorrectly classified as not having OSA)		104 (0 to 461)	0 (0 to 0)	66 (0 to 291)	0 (0 to 0)	
		104 more FN in Watch-PAT		**66 more FN in Watch-PAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	113 (86 to 126)	130 (130 to 130)	391 (297 to 436)	450 (450 to 450)	31 (1) [A]
		17 fewer TN in Watch-PAT		**59 fewer TN in Watch-PAT**		
False positives (patients incorrectly classified as having OSA)		17 (4 to 44)	0 (0 to 0)	59 (14 to 153)	0 (0 to 0)	
		17 more FP in Watch-PAT		**59 more FP in Watch-PAT**		

[1]Wide confidence intervals for sensitivity and specificity

Table S55—Summary of Findings table for Watch-PAT vs. In-lab PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Garg 2014 (A)

Single study sensitivity Watch-PAT: 0.96 (95% CI: 0.85 to 0.99) | **Single study specificity Watch-PAT**: 0.43 (95% CI: 0.22 to 0.66) **Accuracy (high risk)**: 89% (95% CI: 77 to 95%) **Accuracy (low risk):** 72% (95% CI: 57 to 84%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Watch-PAT	In-Home PSG	Watch-PAT	In-Home PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	835 (739 to 861)	870 (870 to 870)	528 (468 to 545)	550 (550 to 550)	75 (1) [A]
		35 fewer TP in Watch-PAT		**22 fewer TP in Watch-PAT**		
False negatives (patients incorrectly classified as not having OSA)		35 (9 to 131)	0 (0 to 0)	22 (5 to 82)	0 (0 to 0)	
		35 more FN in Watch-PAT		**22 more FN in Watch-PAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	56 (29 to 86)	130 (130 to 130)	193 (99 to 297)	450 (450 to 450)	75 (1) [A]
		74 fewer TN in Watch-PAT		**257 fewer TN in Watch-PAT**		
False positives (patients incorrectly classified as having OSA)		74 (44 to 101)	0 (0 to 0)	257 (153 to 351)	0 (0 to 0)	
		74 more FP in Watch-PAT		**257 more FP in Watch-PAT**		

[1]Wide confidence intervals for specificity

Table S56—Summary of Findings table for Watch-PAT vs. In-lab PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Pittman 2004 (A); Garg 2014 (B)

Range of sensitivities Watch-PAT: 0.92 to 0.96 | **Range of specificities Watch-PAT**: 0.77 to 1.00 **Accuracy (high risk)**: 84% to 97% **Accuracy (low risk):** 82% to 99%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Watch-PAT	In-Lab PSG	Watch-PAT	In-Lab PSG	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	589 to 614	640 to 640	230 to 240	250 to 250	104 (2) [A,B]
		26 to 51 fewer TP in Watch-PAT		**10 to 37 fewer TP in Watch-PAT**		
False negatives (patients incorrectly classified as not having OSA)		26 to 51	0 to 0	10 to 20	0 to 0	
		26 to 51 more FN in Watch-PAT		**10 to 20 more FN in Watch-PAT**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	277 to 360	360 to 360	578 to 750	750 to 750	104 (2) [A,B]
		0 to 83 fewer TN in Watch-PAT		**0 to 172 more TN in Watch-PAT**		
False positives (patients incorrectly classified as having OSA)		0 to 83	0 to 0	0 to 172	0 to 0	
		0 to 83 more FP in Watch-PAT		**0 to 172 more FP in Watch-PAT**		

[1]Wide range of values for specificity

Table S57—Summary of Findings table for Watch-PAT vs. In-lab PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Pittman 2004 (A)

Single study sensitivity Watch-PAT: 0.92 (95% CI: 0.62 to 1.00) | **Single study specificity Watch-PAT**: 0.82 (95% CI: 0.57 to 0.96) **Accuracy (high risk)**: 83% (95% CI: 58 to 97%) **Accuracy (low risk):** 83% (95% CI: 58 to 96%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Watch-PAT	In-Lab PSG	Watch-PAT	In-Lab PSG	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	331 (223 to 360)	360 (360 to 360)	92 (62 to 100)	100 (100 to 100)	29 (1) [A]
		29 fewer TP in Watch-PAT		**8 fewer TP in Watch-PAT**		
False negatives (patients incorrectly classified as not having OSA)		29 (0 to 137)	0 (0 to 0)	8 (0 to 38)	0 (0 to 0)	
		29 more FN in Watch-PAT		**8 more FN in Watch-PAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	525 (365 to 614)	360 (360 to 360)	738 (513 to 864)	100 (100 to 100)	29 (1) [A]
		165 more TN in Watch-PAT		**638 more TN in Watch-PAT**		
False positives (patients incorrectly classified as having OSA)		115 (26 to 275)	0 (0 to)	162 (36 to 87)	0 (0 to 0)	
		115 more FP in Watch-PAT		**162 more FP in Watch-PAT**		

[1]Wide confidence intervals for sensitivity and specificity

Figure S18—HSAT vs. Attended PSG (ESS)

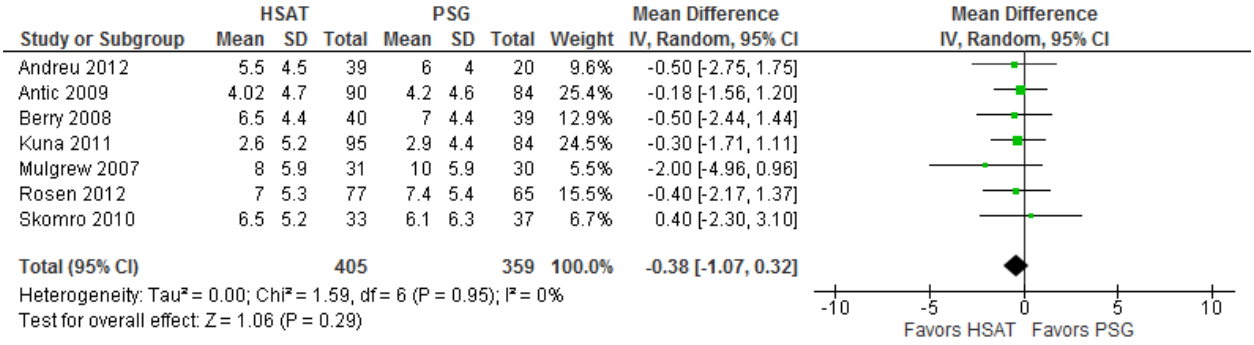


Figure S19—HSAT vs. PSG (QOL; FOSQ)

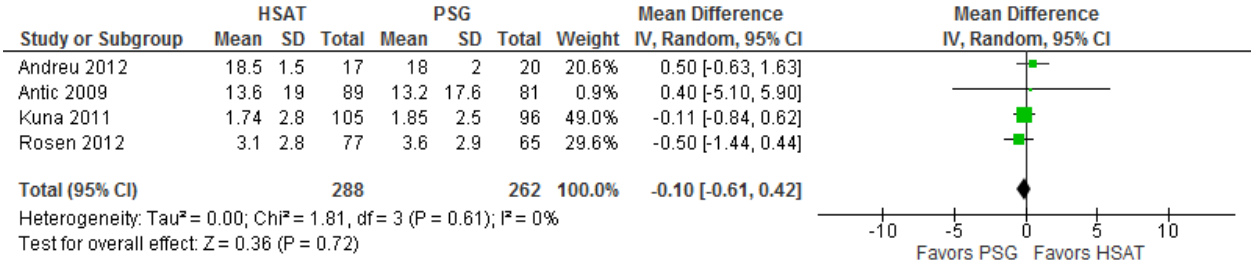


Figure S20—HSAT vs. PSG (QOL; SAQLI)

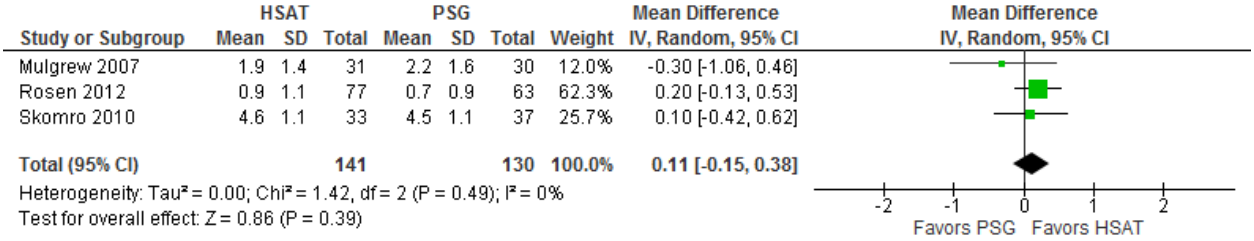


Figure S21—HSAT vs. PSG (QOL; SF-36 Vitality Score)

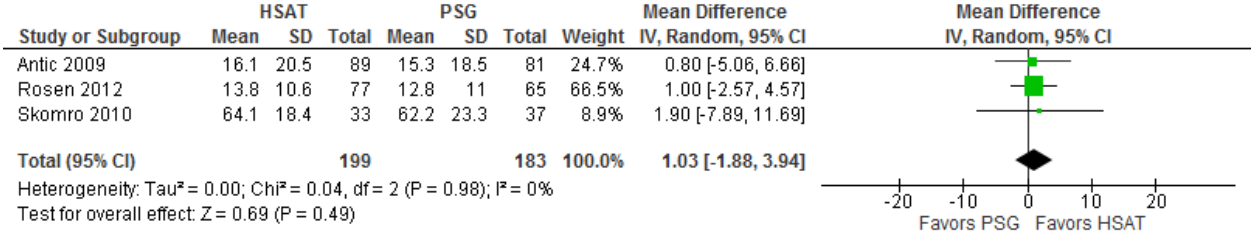


Figure S22—HSAT vs. PSG (QOL; SF-12/36 Physical Component Summary Score)

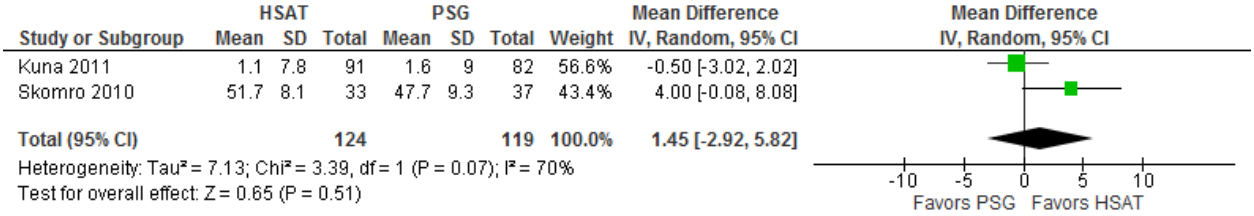


Figure S23—HSAT vs. PSG (QOL; SF-12/36 Mental Component Summary Score)

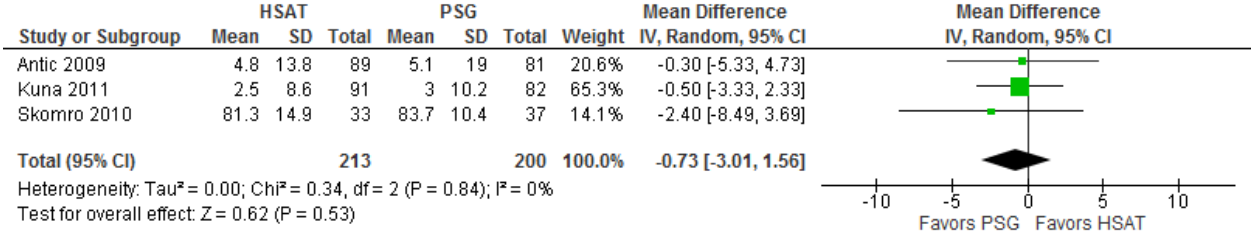


Figure S24—HSAT vs. PSG (CPAP Adherence, h/night)

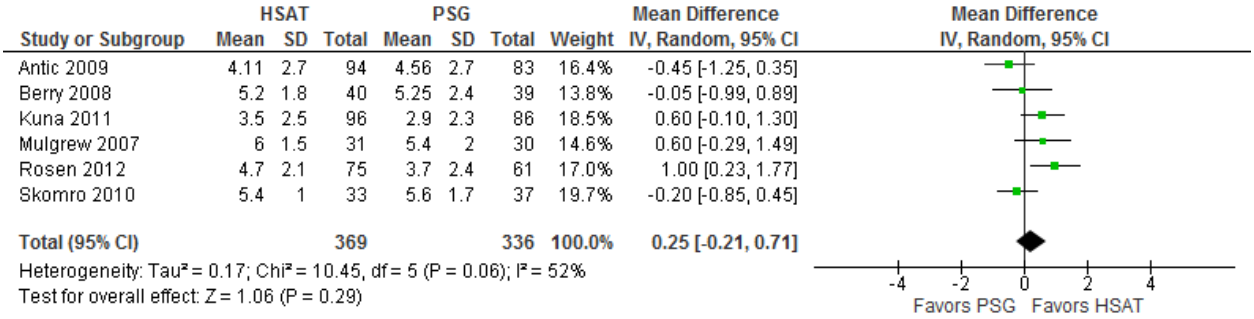


Figure S25—HSAT vs. PSG (CPAP Adherence, no. nights > 4 h)

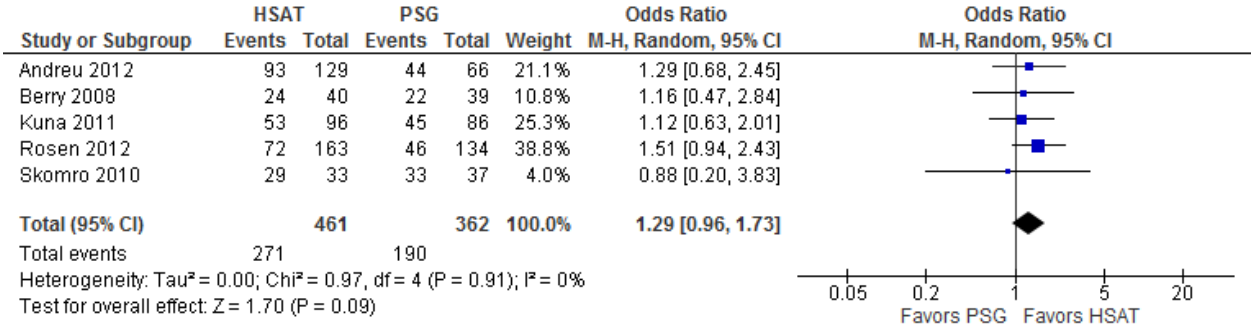


Table S58—HSAT compared to PSG for adults suspected of OSA

References: Andreu 2012 (A); Antic 2009 (B); Berry 2008 (C); Kuna 2011 (D); Mulgrew 2007 (E); Rosen 2012 (F); Skomro 2010 (G)

Patient or population: adults suspected of OSA
Setting: Home, lab
Intervention: HSAT
Comparison: Attended PSG

Outcomes	Quality of the evidence (GRADE)	Anticipated absolute effects* (95% CI) MD between HSAT and PSG	№ of participants (studies)	Comments
Sleepiness* (ESS)	⊕⊕⊕⊕ HIGH	The mean difference in sleepiness (ESS) after treatment was 0.38 less (1.07 less to 0.32 less) with HSAT	764 (7 RCTs) [A-G]	
QOL (FOSQ)*	⊕⊕⊕⊕ HIGH	The mean difference in QOL (FOSQ) after treatment was 0.10 lower (0.42 higher to 0.61 lower) with HSAT	550 (4 RCTs) [A,B,D,F]	
QOL (SAQLI)*	⊕⊕⊕⊕ HIGH	The mean difference in QOL (SAQLI) after treatment was 0.11 greater (0.15 lower to 0.38 greater) with HSAT	271 (3 RCTs) [E,F,G]	
QOL (SF-36 Vitality Score)*	⊕⊕⊕⊕ HIGH	The mean difference in QOL (SF-36 Vitality Score) after treatment was 1.03 greater (1.88 lower to 3.94 greater) with HSAT	382 (3 RCTs) [B,F,G]	
QOL (SF-12/SF-36 Physical Component Summary)*	⊕⊕⊕◯ MODERATE[1]	The mean difference in QOL (SF-12/SF-36 Physical Component Summary) after treatment was 1.45 greater (2.92 fewer to 5.82 greater) with HSAT	243 (2 RCTs) [D,G]	
QOL (SF-12/SF-36 Mental Component Summary)*	⊕⊕⊕◯ MODERATE[1]	The mean difference in QOL (SF-12/SF-36 Mental Component Summary) after treatment was 0.73 lower (1.56 greater to 3.01 lower) with HSAT	413 (3 RCTs) [B,D,G]	
CPAP Adherence (h/night)*	⊕⊕⊕◯ MODERATE[2]	The mean CPAP Adherence (h/night) in the intervention group was 0.25 h more (0.21 less to 0.71 more) with HSAT	705 (6 RCTs) [B-G]	
		Relative Effect Baseline Risk / Comparative risk		
Compliance (No. of nights > 4 h)*	⊕⊕⊕⊕ HIGH	525 per 1000 / **588 per 1000** (515 to 656) **OR 1.29** (0.96 to 1.73)	823 (5 RCTs) [A,C,D,F,G]	

*Critical Outcomes
[1]Quality of evidence for QOL as measured by the SF-36 was downgraded due to imprecision (i.e., 95% CI of mean difference crosses clinical decision threshold of 3 points for SF-36 physical and mental component summary scores and the origin of the plot)
[2]Quality of evidence for adherence was downgraded due to imprecision (i.e., 95% CI of mean difference crosses clinical decision threshold of 0.5 h/night and the origin of the plot)

Table S59—Summary of Findings table for Multiple-night HSAT vs. Single-night HSAT to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Rofail 2010 (A)

Single study sensitivity Multiple-night HSAT: 0.80 (95% CI: 0.67 to 0.93) | **Single study specificity Multiple-night HSAT**: 0.87 (95% CI: 0.77 to 0.97) **Multiple-night HSAT Accuracy (high risk)**: 81% (95% CI: 68 to 94%) **Multiple-night HSAT Accuracy (low risk):** 83% (95% CI: 72 to 95%) **Single study sensitivity Single-night HSAT**: 0.75 (95% CI: 0.63 to 0.85) | **Single study specificity Single-night HSAT**: 0.79 (95% CI: 0.61 to 0.97) **Single-night HSAT Accuracy (high risk)**: 76% (95% CI: 63 to 86%) **Single-night HSAT Accuracy (low risk):** 77% (95% CI: 62 to 90%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Multiple-night HSAT	Single-night HSAT	Multiple-night HSAT	Single-night HSAT	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	696 (583 to 809)	653 (548 to 739)	440 (369 to 512)	413 (347 to 468)	92 (1) [A]
		43 more TP in Multiple-night HSAT		**27 more TP in Multiple-night HSAT**		
False negatives (patients incorrectly classified as not having OSA)		174 (61 to 287)	217 (131 to 322)	110 (38 to 181)	137 (82 to 203)	
		43 fewer FN in Multiple-night HSAT		**27 fewer FN in Multiple-night HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	113 (100 to 126)	103 (79 to 126)	391 (347 to 436)	356 (274 to 436)	92 (1) [A]
		10 more TN in Multiple-night HSAT		**35 more TN in Multiple-night HSAT**		
False positives (patients incorrectly classified as having OSA)		17 (4 to 30)	27 (4 to 51)	59 (14 to 103)	94 (14 to 176)	
		10 fewer FP in Multiple-night HSAT		**35 fewer FP in Multiple-night HSAT**		

[1]Wide confidence intervals for sensitivity and specificity

Table S60—Summary of Findings table for Multiple-night HSAT vs. Single-night HSAT to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Rofail 2010 (A)

Single study sensitivity Multiple-night HSAT: 0.90 (95% CI: 0.83 to 0.98) | **Single study specificity Multiple-night HSAT**: 0.85 (95% CI: 0.78 to 0.89) **Multiple-night HSAT Accuracy (high risk)**: 87% (95% CI: 80 to 92%) **Multiple-night HSAT Accuracy (low risk):** 86% (95% CI: 78 to 90%) **Single study sensitivity Single-night HSAT**: 0.90 (95% CI: 0.84 to 0.98) | **Single study specificity Single-night HSAT**: 0.83 (95% CI: 0.76 to 0.87) **Single-night HSAT Accuracy (high risk)**: 86% (95% CI: 77 to 91%) **Single-night HSAT Accuracy (low risk):** 84% (95% CI: 78 to 88%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Multiple-night HSAT	Single-night HSAT	Multiple-night HSAT	Single-night HSAT	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	324 (299 to 353)	324 (302 to 353)	90 (83 to 98)	90 (84 to 98)	92 (1) [A]
		0 fewer TP in Multiple-night HSAT		**0 fewer TP in Multiple-night HSAT**		
False negatives (patients incorrectly classified as not having OSA)		36 (7 to 61)	36 (7 to 58)	10 (2 to17)	10 (2 to16)	
		0 fewer FN in Multiple-night HSAT		**0 fewer FN in Multiple-night HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	544 (499 to 570)	531 (486 to 557)	765 (702 to 801)	747 (684 to 783)	92 (1) [A]
		13 more TN in Multiple-night HSAT		**18 more TN in Multiple-night HSAT**		
False positives (patients incorrectly classified as having OSA)		96 (70 to 141)	109 (83 to154)	135 (99 to198)	153 (117 to 216)	
		13 fewer FP in Multiple-night HSAT		**18 fewer FP in Multiple-night HSAT**		

Diagnosis of obstructive sleep apnea in adults with comorbid conditions

Table S61—Summary of Findings table for HSAT vs. PSG to diagnose OSA in Suspected Adults with comorbid conditions (AHI ≥ 15)

References: Abraham 2006 (A); Series 2005 (B); de Vries 2015 (C)

Range of sensitivities HSAT: 0.64 to 0.93 | **Range of specificities HSAT**: 0.78 to 0.95
Range of sensitivities Attended PSG: 1.00 to 1.00 | **Range of specificities Attended PSG**: 1.00 to 1.00 **Accuracy (high risk)**: 69% to 89% **Accuracy (low risk):** 74% to 92%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		HSAT	Attended PSG	HSAT	Attended PSG	
True positives (patients with OSA)	⨁⨁◯◯ LOW[1,2]	410 to 595	640 to 640	160 to 233	250 to 250	190 (3 [A-C])
		45 to 230 fewer TP in HSAT		**17 to 90 fewer TP in HSAT**		
False negatives (patients incorrectly classified as not having OSA)		45 to 230	0 to 0	17 to 90	0 to 0	
		45 to 230 more FN in HSAT		**17 to 90 more FN in HSAT**		
True negatives (patients without OSA)	⨁⨁◯◯ LOW[1,2]	281 to 342	360 to 360	585 to 712	750 to 750	190 (3) [A-C]
		18 to 79 fewer TN in HSAT		**38 to 165 fewer TN in HSAT**		
False positives (patients incorrectly classified as having OSA)		18 to 79	0 to 0	38 to 165	0 to 0	
		18 to 79 more FP in HSAT		**38 to 165 more FP in HSAT**		

[1]Wide range of values for sensitivity and specificity
[2]Indirectness as study populations not representative of all comorbid conditions typically associated with OSA

Diagnosis of obstructive sleep apnea in adults using a split-night versus a full-night polysomnography protocol

Table S62—Summary of Findings table for Split-night PSG vs. Full-night PSG to diagnose OSA in Suspected Adults (AHI ≥ 5)

References: Khawja 2010 (A)

Single study sensitivity split-night HSAT: 0.80 (95% CI: 0.67 to 0.90) | **Single study specificity split-night HSAT**: 0.93 (95% CI: 0.83 to 0.98) **Accuracy (high risk)**: 82% (95% CI: 69 to 91%) **Accuracy (low risk):** 86% (95% CI: 74 to 94%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 87%		Prevalence 55%		
		Split-night HSAT	Full-night HSAT	Split-night HSAT	Full-night HSAT	
True positives (patients with OSA)	⨁⨁⨁◯ MODERATE[1]	699 (583 to 783)	870 (870 to 870)	442 (369 to 495)	550 (550 to 550)	114 (1) [A]
		171 fewer TP in split-night HSAT		**108 fewer TP in split-night HSAT**		
False negatives (patients incorrectly classified as not having OSA)		171 (87 to 287)	0 (0 to 0)	108 (55 to 181)	0 (0 to 0)	
		171 fewer FN in split-night HSAT		**108 more FN in split-night HSAT**		
True negatives (patients without OSA)	⨁⨁⨁◯ MODERATE[1]	121 (108 to 127)	130 (130 to 130)	419 (373 to 441)	450 (450 to 450)	114 (1) [A]
		9 fewer TN in split-night HSAT		**31 fewer TN in split-night HSAT**		
False positives (patients incorrectly classified as having OSA)		9 (3 to 22)	0 (0 to 0)	31 (9 to 77)	0 (0 to 0)	
		9 more FP in split-night HSAT		**31 more FP in split-night HSAT**		

[1]Wide confidence intervals for sensitivity

Table S63—Summary of Findings table for Split-night PSG vs. Full-night PSG to diagnose OSA in Suspected Adults (AHI ≥ 15)

References: Khawja 2010 (A)

Single study sensitivity split-night HSAT: 0.77 (95% CI: 0.56 to 0.91) | **Single study specificity split-night HSAT**: 0.98 (95% CI: 0.92 to 1.00) **Accuracy (high risk)**: 85% (95% CI: 69 to 94%) **Accuracy (low risk):** 93% (95% CI: 83 to 98%)

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested (95% CI)				Number of participants (studies)
		Prevalence 64%		Prevalence 25%		
		Split-night HSAT	Full-night HSAT	Split-night HSAT	Full-night HSAT	
True positives (patients with OSA)	⊕⊕⊕◯ MODERATE[1]	493 (358 to 582)	640 (640 to 640)	193 (140 to 228)	250 (250 to 250)	114 (1) [A]
		147 fewer TP in split-night HSAT		**57 fewer TP in split-night HSAT**		
False negatives (patients incorrectly classified as not having OSA)		147 (58 to 282)	0 (0 to 0)	57 (22 to 110)	0 (0 to 0)	
		147 more FN in split-night HSAT		**57 more FN in split-night HSAT**		
True negatives (patients without OSA)	⊕⊕⊕◯ MODERATE[1]	353 (331 to 359)	360 (360 to 360)	735 (690 to 748)	750 (750 to 750)	114 (1) [A]
		7 fewer TN in split-night HSAT		**15 fewer TN in split-night HSAT**		
False positives (patients incorrectly classified as having OSA)		7 (1 to 29)	0 (0 to 0)	15 (2 to 60)	0 (0 to 0)	
		7 more FP in split-night HSAT		**15 more FP in split-night HSAT**		

[1]Wide confidence intervals for sensitivity

Table S64—Summary of Findings table for Split-night PSG vs. Full-night PSG to diagnose OSA in Suspected Adults (AHI ≥ 30)

References: Chou 2011 (A)

Single study sensitivity split-night HSAT: 0.90 (95% CI: not available) | **Single study specificity split-night HSAT**: 0.92 (95% CI: not available) **Accuracy (high risk)**: 91% **Accuracy (low risk):** 92%

Test result	Quality of the Evidence (GRADE)	Number of results per 1000 patients tested				Number of participants (studies)
		Prevalence 36%		Prevalence 10%		
		Split-night HSAT	Full-night HSAT	Split-night HSAT	Full-night HSAT	
True positives (patients with OSA)	⊕⊕⊕⊕ HIGH	324	360	90	100	198 (1) [A]
		36 fewer TP in split-night HSAT		**10 fewer TP in split-night HSAT**		
False negatives (patients incorrectly classified as not having OSA)		36	0	10	0	
		36 more FN in split-night HSAT		**10 more FN in split-night HSAT**		
True negatives (patients without OSA)	⊕⊕⊕⊕ HIGH	589	640	828	900	198 (1) [A]
		51 fewer TN in split-night HSAT		**72 fewer TN in split-night HSAT**		
False positives (patients incorrectly classified as having OSA)		51	0	72	0	
		51 more FP in split-night HSAT		**72 more FP in split-night HSAT**		

Figure S26—Split-night PSG vs. Full-night PSG (Adherence, h/night)

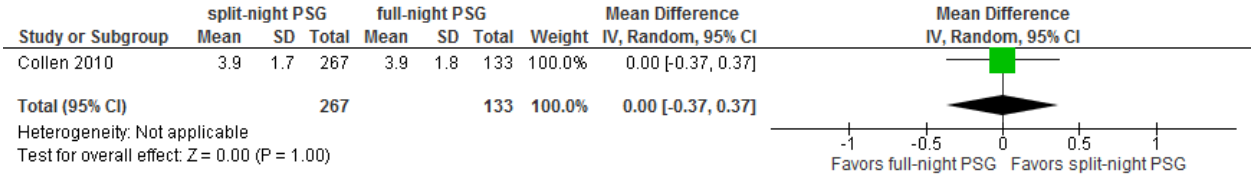


Figure S27—Split-night PSG vs. Full-night PSG (AHI after CPAP)

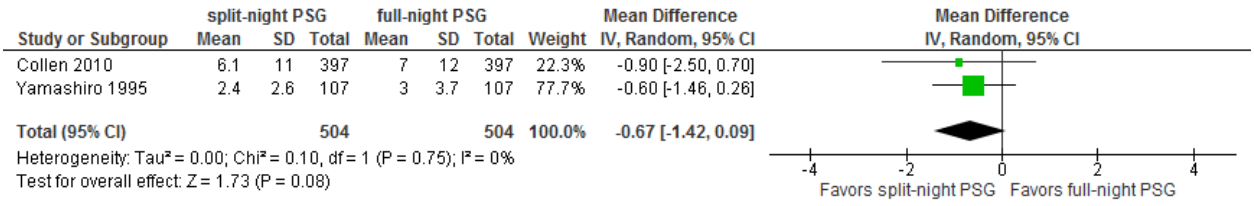


Table S65—Summary of Findings table for split-night PSG vs. full-night PSG for the improvement in clinical outcomes of Adults suspected of OSA

References: Collen 2010 (A); Yamashiro 1995 (B)

Patient or population: adults suspected of OSA
Setting: in-lab
Intervention: split-night PSG
Comparison: full-night PSG

Outcomes	Quality of the evidence (GRADE)	Anticipated absolute effects* (95% CI) MD between HSAT and PSG	№ of participants (studies)	Comments
AHI*	⨁⨁◯◯ LOW[1]	The mean difference in AHI after treatment was 0.67 lower (1.42 lower to 0.09 higher) with split-night	504 (2 RCTs) [A,B]	
CPAP Adherence (h/night) *	⨁⨁◯◯ LOW[1]	The mean CPAP Adherence (h/night) in the split-night PSG group was 0.00 greater (0.37 fewer to 0.37 greater) with split-night PSG	400 (1 RCT) [A]	

*Critical Outcomes
[1]Downgraded due to imprecision associated with a limited number of studies and small sample size

Repeat polysomnography for the diagnosis of obstructive sleep apnea in adults

Figure S28—Two-night PSG vs. Single-night PSG (night-to-night variability in AHI)

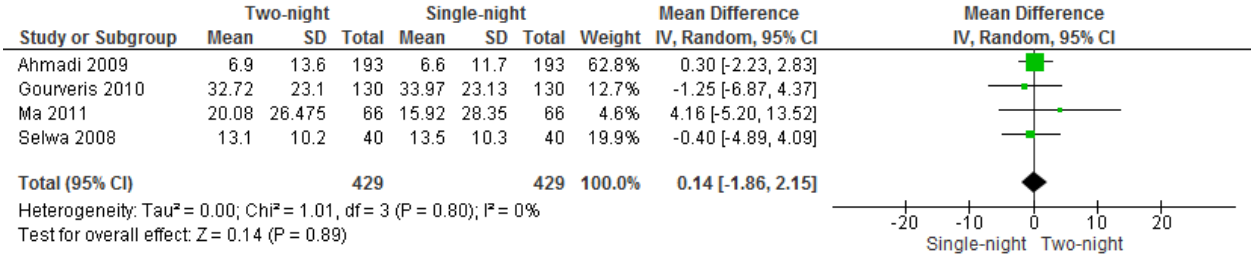


Table S66—Summary of Findings table for Two-night PSG vs. Single-night PSG for the improvement in clinical outcomes of Adults suspected of OSA

References: Ahmadi 2009 (A); Gourveris 2010 (B); Ma 2011 (C); Selwa 2008 (D)

Patient or population: Adults suspected of OSA
Setting: Attended in-lab
Intervention: Two-night PSG
Comparison: Single-night PSG

Outcomes	Quality of the evidence (GRADE)	Anticipated absolute effects* (95% CI) Single-night PSG vs. second-night PSG	№ of participants (studies)
AHI	⊕⊕⊕⊕ HIGH	The mean difference in AHI (variability) was 0.14 events/h lower (-1.86 greater to 2.15 lower) with a single-night PSG.	858 (4 RCTs) [A-D]