Comparative Effectiveness Review
Number 32

Effective Health Care Program

Diagnosis and Treatment of Obstructive Sleep Apnea in Adults

Executive Summary

Background[a]

Obstructive sleep apnea (OSA) is a relatively common disorder in the United States that affects people of all ages, but is most prevalent among the middle-aged and elderly. Affected individuals experience repeated collapse and obstruction of the upper airway during sleep, which results in reduced airflow (hypopnea) or complete airflow cessation (apnea), oxygen desaturation, and arousals from sleep. Adverse clinical outcomes associated with OSA include: cardiovascular disease, hypertension, non-insulin-dependent diabetes, and increased likelihood of motor vehicle and other accidents due to daytime hypersomnolence. Studies estimate the prevalence of OSA at approximately 10 to 20 percent of middle-aged and older adults. Evidence also indicates that these rates are rising, likely due to increasing rates of obesity.

Based on the considerable mortality and morbidity associated with it and its attendant comorbidities, OSA is an important public health issue. Complicating diagnosis and treatment, however, is the great degree of clinical uncertainty that exists regarding the condition, due in large part to inconsistencies in its definition. Ongoing debate surrounds what type and level of respiratory abnormality should be used to define the disorder as well as what the most appropriate diagnostic method is for its detection. In addition, there is no

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at **www.effectivehealthcare.ahrq.gov/reports/final.cfm.**

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care

current established threshold level for the apnea-hypopnea index (AHI) that would indicate the need for treatment. By consensus, people with relatively few apnea or hypopnea events per hour (often <5 or <15) are not formally diagnosed with OSA. Also of concern are the high rates of perioperative and postoperative complications among OSA patients, as are the numbers of asymptomatic and symptomatic individuals who remain undiagnosed and untreated.

Three main categories of outcomes of interest in comparative effectiveness research are clinical (or health) outcomes (i.e., events or conditions that the patient can feel, such as disability or quality of life or death), intermediate or surrogate outcomes (such as laboratory measurements), and adverse events. Objective clinical outcomes relevant to patients with OSA include comorbidities found to be associated with untreated sleep apnea, primarily cardiovascular disease (including congestive heart failure, hypertension, stroke, and myocardial infarction) and non-insulin-dependent diabetes. In addition, mortality due to cardiovascular disease, diabetes, motor vehicle accidents, and other causes represent important adverse outcomes of OSA. Intermediate outcomes of interest in the management of patients with OSA include sleep study measures (e.g., AHI), blood pressure (an intermediate outcome for cardiovascular disease), and hemoglobin A1c (a measure of control of diabetes mellitus).

All interventions have the potential for adverse events. Therefore, it is important to gather information on both the benefits and harms of interventions in order to fully assess the net comparative benefits. Compliance with continuous positive airway pressure (CPAP) and other devices is an important issue related to the effective treatment of OSA. Interventions that have better compliance or that may improve compliance are clearly of interest. Also of relevance is establishing definitive diagnostic standards and measures that would more clearly identify OSA patients, both symptomatic and asymptomatic. Such standards would serve to reduce OSA-related morbidities as well as related health care costs. Studies have found that prior to diagnosis, OSA patients have higher rates of health care use, more frequent and longer hospital stays, and greater health care costs than after diagnosis. Therefore, this review is of additional interest to the requesting organizations and broadly for the identification of diagnostic tests that would contribute to the early and definitive diagnosis of patients with OSA.

Objectives

In response to several nominations received through the Effective Health Care Web site, which were evaluated and found to meet program criteria, the Agency for Healthcare Research and Quality (AHRQ) requested that the Tufts Evidence-based Practice Center (Tufts EPC) conduct a Comparative Effectiveness Review (CER) of studies regarding the diagnosis and treatment of OSA[b]. Key Questions that are clinically relevant for the diagnosis and treatment of OSA were developed with input from domain experts and other stakeholders and from comments received in response to public review. Seven Key Questions are addressed in this report. Three pertain to diagnosis of and screening for OSA (Key Questions 1-3), two address the comparative effectiveness of treatments (Key Questions 5 and 7), and two address associations between baseline patient characteristics and long-term outcomes and treatment compliance (Key Questions 4 and 6).

Key Questions

Diagnosis

1. How do different available tests compare in their ability to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? How do these tests compare in different subgroups of patients, based on: race, sex, body mass index, existing non-insulin-dependent diabetes mellitus, existing cardiovascular disease, existing hypertension, clinical symptoms, previous stroke, or airway characteristics?
2. How does phased testing (screening tests or battery followed by full test) compare to full testing alone?
3. What is the effect of preoperative screening for sleep apnea on surgical outcomes?
4 In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index or oxygen desaturation index and other patient characteristics with respect to long-term clinical and functional outcomes?

Treatment

5. What is the comparative effect of different treatments for obstructive sleep apnea in adults?
    a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of obstructive sleep apnea, or other pretreatment factors? Are any of these characteristics or factors predictive of treatment success?
        - Characteristics: Age, sex, race, weight, bed partner, airway, other physical characteristics, and specific comorbidities
        - Obstructive sleep apnea severity or characteristics: Baseline questionnaire (and similar tools) results, formal testing results (including hypoxemia levels), baseline quality of life, positional dependency
        - Other: Specific symptoms
    b. Does the comparative effect of treatments vary based on the definitions of obstructive sleep apnea used by study investigators?
6. In obstructive sleep apnea patients prescribed nonsurgical treatments, what are the associations of pretreatment patient-level characteristics with treatment compliance?
7. What is the effect of interventions to improve compliance with device use (positive airway pressure, oral appliances, positional therapy) on clinical and intermediate outcomes?

Analytic Framework

To guide the development of the Key Questions for the diagnosis and treatment of OSA, we developed an analytic framework (Figure A) that maps the specific linkages associating the populations and subgroups of interest, the interventions (for both diagnosis and treatment), and outcomes of interest (intermediate outcomes, health-related outcomes, compliance, and adverse effects). Specifically, this analytic framework depicts the chain of logic that evidence must support to link the interventions to improved health outcomes.

Methods

Input from Stakeholders

During a topic refinement phase, the initial questions were refined with input from a panel of Key Informants. The Key Informants included experts in sleep medicine, general internal medicine, and psychiatry; a representative from Oregon Division of Medical Assistance programs; a person with OSA; a representative of a sleep apnea advocacy group; and the (AHRQ) Task Order Officer.

After a public review of the proposed Key Questions, the clinical experts from among the Key Informants were reconvened to form the Technical Expert Panel, which served to provide clinical and methodological expertise and input to help refine Key Questions, identify important issues, and define parameters for the review of evidence, including study eligibility criteria.

Data Sources and Selection

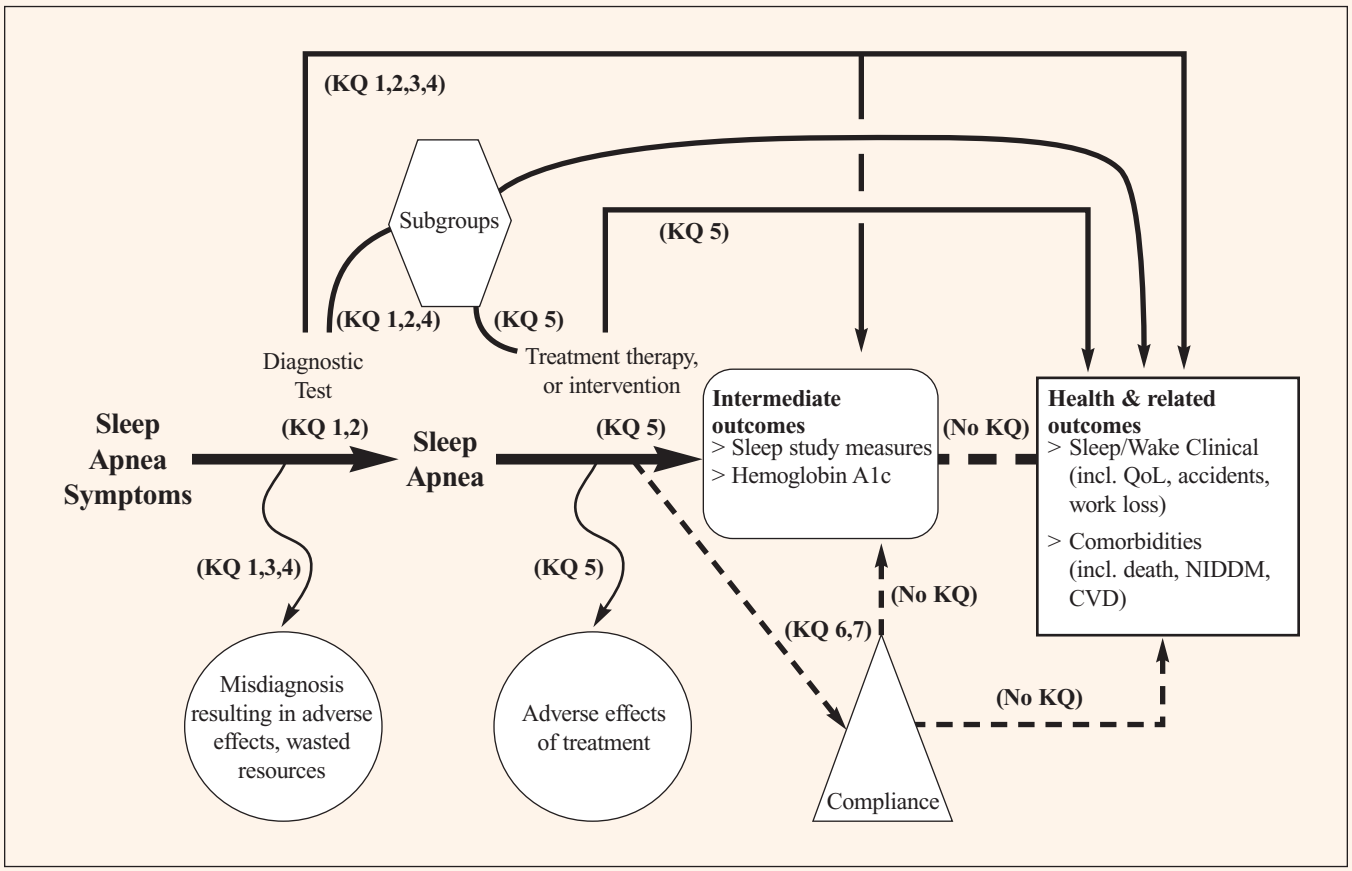
We conducted literature searches of studies in MEDLINE® (inception–September 2010) and the Cochrane Central Register of Controlled Trials (through 3rd quarter 2010). All English-language studies with adult human subjects were screened to identify articles relevant to each Key Question. The search strategy included terms for OSA, sleep apnea diagnostic tests, sleep apnea treatments, and relevant research designs.

The reference lists of related systematic reviews and selected narrative reviews and primary articles were also reviewed, and relevant articles were screened. After screening of the abstracts, full-text articles were retrieved for all potentially relevant articles and rescreened for eligibility.

Data Extraction and Quality Assessment

Study data were extracted into customized forms. Together with information on study design, patient and intervention characteristics, outcome definitions, and study results, the methodological quality of each study was rated from A (highest quality, least likely to have significant bias) to C (lowest quality, most likely to have significant bias).

Figure A. Analytic framework for the diagnosis and treatment of obstructive sleep apnea in adults



CVD, cardiovascular disease; KQ, Key Question; NIDDM, non-insulin-dependent diabetes mellitus; QoL, quality of life.

Data Synthesis and Analysis

For all Key Questions or specific comparisons of interventions with at least two studies, summary tables present the study and baseline patient characteristics, the study quality, and the relevant study results. For each comparison, separate tables include all the studies that reported specific outcomes. For Key Question 1 (diagnosis), we graphically display the Bland-Altman limits of agreement and the sensitivity and specificity of studies comparing portable monitors to polysomnography (PSG). For Key Question 5 (treatment), when there were three or more similar studies evaluating the same outcome, we performed random effects model meta-analyses of the following: the sleep study measures AHI, arousal index, and minimum oxygen saturation; the standard measure of sleepiness, the Epworth Sleepiness Scale (ESS); the quality-of-life measure Functional Outcomes Sleep Questionnaire (FOSQ); and compliance. We performed subgroup meta-analyses based on study design (parallel or crossover), minimum AHI threshold to diagnose OSA, specific intervention (when appropriate), and other factors. Of note, where interventions (either diagnostic tests or treatments) are not discussed, this does not imply that the interventions were excluded from analysis (unless explicitly stated); instead, no studies of these interventions met eligibility criteria.

As per the AHRQ updated methods guide series, we assessed the evidence for each question (or comparison of interventions) based on the risk of bias, study consistency, directness of the evidence, and degree of certainty of the findings. Based on these factors, we graded the overall strength of evidence as high, moderate, low, or insufficient.

When there were substantial differences in conclusions for different outcomes within the same comparison, we also described the evidence supporting each outcome as sufficient, fair, weak, limited, or no evidence.

Results

Key Question 1. How do different available tests compare in their ability to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? How do these tests compare in different subgroups of patients based on: race, sex, body mass index, existing non-insulin-dependent diabetes mellitus, existing cardiovascular disease, existing hypertension, clinical symptoms, previous stroke, or airway characteristics?

Comparison of Portable Devices and Polysomnography

PSG devices are classified as Type I monitors. Portable monitors are classified as either Type II, which record all the same information as PSG; Type III, which do not differentiate between whether the patient is asleep or awake, but have at least two respiratory channels (two airflow channels or one airflow and one effort channel); or Type IV, which fail to fulfill criteria for Type III monitors but usually record more than two bioparameters.

The strength of evidence is moderate, among 15 quality A, 45 quality B, and 39 quality C studies, that Type III and Type IV monitors may have the ability to accurately predict AHI suggestive of OSA with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in PSG. Type III monitors perform better than Type IV monitors at AHI cutoffs of 5, 10, and 15 events/hr. Analysis of difference versus average analyses plots suggest that substantial differences in the measured AHI may be encountered between PSG and both Type III and Type IV monitors. Large differences compared with in-laboratory PSG cannot be excluded for all portable monitors. The evidence is insufficient to adequately compare specific monitors to each other.

No recent studies compared Type II monitors with PSG. A prior Technology Assessment of home diagnosis of OSA concluded that "based on [three quality B studies], type II monitors [used at home] may identify AHI suggestive of OSA with high positive likelihood ratios and low negative likelihood ratios," though "substantial differences in the [measurement of] AHI may be encountered between type II monitors and facility-based PSG."

Comparison of Questionnaires and Polysomnography

Of the six studies reviewed (one quality A, one quality B, four quality C), *the strength of evidence is low* among three studies supporting the use of the Berlin questionnaire in screening for sleep apnea because of the likely selection biases. *The strength of evidence is insufficient* to draw definitive conclusions concerning the use of the STOP, STOP-Bang, ASA Checklist, Epworth Sleepiness Scale, and Hawaii Sleep questionnaires to screen for sleep apnea because each questionnaire was assessed in only a single study.

Clinical Prediction Rules and Polysomnography

The strength of evidence is low among seven studies (three quality A, three quality B, and one quality C) that some clinical prediction rules may be useful in the prediction of a diagnosis of OSA. Ten different clinical prediction rules have been described. Nine clinical prediction rules have been used for the prediction of a diagnosis of OSA (using different criteria). The oropharyngeal morphometric model gave near perfect discrimination (area under the curve [AUC] = 0.996) to predict the diagnosis of OSA, and the pulmonary function data model had 100 percent sensitivity with 84 percent specificity to predict diagnosis of OSA. The remaining models reported lower diagnostic sensitivities and specificities. Each model was deemed useful to predict the diagnoses of OSA by the individual study authors. However, while all the models were internally validated, external validation of these predictive rules has not been conducted in the vast majority of the studies.

Key Question 2. How does phased testing (screening tests or battery followed by full test) compare to full testing alone?

The strength of evidence is insufficient to determine the utility of phased testing, followed by full testing when indicated, to diagnose sleep apnea, as only one study that met our inclusion criteria investigated this question. This prospective quality C study did not fully analyze the phased testing, thus the sensitivity and specificity of

the phased strategy could not be calculated due to a verification bias; not all participants received PSG (full) testing.

Key Question 3. What is the effect of preoperative screening for sleep apnea on surgical outcomes?

The strength of evidence is insufficient regarding postoperative outcomes with mandatory screening for sleep apnea. Two quality C prospective studies assessed the effect of preoperative screening for sleep apnea on surgical outcomes. One study found no significant differences in outcomes between patients undergoing bariatric surgery who had mandatory PSG or PSG based on clinical parameters. The second study found that general surgery patients willing to undergo preoperative PSG were more likely to have perioperative complications, particularly cardiopulmonary complications, possibly suggesting that patients willing to undergo PSG are more ill than other patients.

Key Question 4. In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index or oxygen desaturation index, and other patient characteristics with respect to long-term clinical and functional outcomes?

The strength of evidence is high from four studies (three quality A, one quality B) indicating that an AHI >30 events/hr is an independent predictor of all-cause mortality; although one study found that this was true only in men under age 70. All other outcomes were analyzed by only one or two studies. Thus, only a *low strength of evidence* exists that a high AHI (>30 events/hr) is associated with incident diabetes. This association, however, may be confounded by obesity, which may result in both OSA and diabetes. *The strength of evidence is insufficient* regarding the association between AHI and other clinical outcomes. The two studies of cardiovascular mortality did not have consistent findings, and the two studies of hypertension had unclear conclusions. One study of nonfatal cardiovascular disease found a significant association with baseline AHI (as they did for cardiovascular mortality). One study each found no association between AHI and stroke or long-term quality of life.

Key Question 5. What is the comparative effect of different treatments for obstructive sleep apnea in adults?

a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of obstructive sleep apnea, or other pretreatment factors? Are any of these characteristics or factors predictive of treatment success?

- **Characteristics: age, sex, race, weight, bed partner, airway, other physical characteristics, and specific comorbidities**
- **Obstructive sleep apnea severity or characteristics: baseline questionnaire (and similar tools) results, formal testing results (including hypoxemia levels), baseline quality of life, positional dependency**
- **Other: specific symptoms**

b. Does the comparative effect of treatments vary based on the definitions of obstructive sleep apnea used by study investigators?

With some exceptions for studies of surgical interventions, we reviewed only randomized controlled trials (RCT) of interventions used specifically for the treatment of obstructive sleep apnea (OSA).

Comparison of Continuous Positive Airway Pressure and Control

There are 22 trials (11 each of quality B and C) that provide sufficient evidence supporting large improvements in sleep measures with continuous positive airway pressure (CPAP) compared with control. There is only weak evidence that demonstrated no consistent benefit in improving quality of life, neurocognitive measures, or other intermediate outcomes. Despite no evidence or weak evidence for an effect of CPAP on clinical outcomes, given the large magnitude of effect on the intermediate outcomes AHI and ESS, the strength of evidence that CPAP is an effective treatment to alleviate sleep apnea signs and symptoms was rated *moderate*.

Comparison of CPAP and Sham CPAP

There are 24 trials (5 quality A, 13 quality B, 6 quality C) that provide sufficient evidence supporting large improvements in sleep measures with CPAP compared

with sham CPAP, but weak evidence of possibly no difference between CPAP and sham CPAP in improving quality of life, neurocognitive measures, or other intermediate outcomes. Despite no evidence or weak evidence for an effect of CPAP on clinical outcomes, given the large magnitude of effect on the intermediate outcomes of AHI, ESS, and arousal index, the strength of evidence that CPAP is an effective treatment for the relief of signs and symptoms of sleep apnea was rated *moderate*.

Comparison of Oral and Nasal CPAP

Three small trials (one quality B, two quality C) with inconsistent results preclude any substantive conclusions concerning the efficacy of oral (or full face mask) versus nasal CPAP in improving compliance in patients with OSA. Largely due to small sample size, the reported effect estimates in the studies reviewed were generally imprecise. Thus, overall, *the strength of evidence is insufficient* regarding differences in compliance or other outcomes between oral and nasal CPAP.

Comparison of Autotitrating CPAP and Fixed CPAP

The strength of evidence is moderate that autotitrating CPAP (autoCPAP) and fixed pressure CPAP result in similar levels of compliance (hours used per night) and treatment effects for patients with OSA. Twenty-one studies (1 quality A, 10 quality B, 10 quality C) comprising an experimental population of over 800 patients provided evidence that autoCPAP reduces sleepiness as measured by ESS by approximately 0.5 points more than fixed CPAP. The two devices were found to result in similar compliance and changes in AHI from baseline, quality of life, and most other sleep study measures. However, there is also evidence that minimum oxygen saturation improves more with fixed CPAP than with autoCPAP, although by only about one percent. Evidence is limited regarding the relative effect of fixed CPAP and autoCPAP on blood pressure. There were no data on objective clinical outcomes.

Comparison of Bilevel CPAP and Fixed CPAP

The strength of evidence is insufficient regarding any difference in compliance or other outcomes between bilevel CPAP and fixed CPAP. Five small, highly clinically heterogeneous trials (one quality B, four quality C) with largely null findings did not support any substantive differences in the efficacy of bilevel CPAP versus fixed CPAP in the treatment of patients with OSA. Largely due to small sample sizes, the studies mostly had imprecise estimates of the comparative effects.

Comparison of Flexible Bilevel CPAP and Fixed CPAP

The strength of evidence is insufficient regarding the relative merits of flexible bilevel CPAP and fixed CPAP as there was only one quality B study that investigated this comparison. This study found that flexible bilevel CPAP may yield increased compliance (use ≥4 hr/night) compared with fixed CPAP.

Comparison of C-Flex™ and Fixed CPAP

No statistically significant differences in compliance or other outcomes were found between C-Flex and fixed CPAP. *The strength of evidence is low* for this finding because of the mixed quality (Bs and Cs) of the four primary studies.

Comparison of Humidification in CPAP

The strength of evidence is insufficient to determine whether there is a difference in compliance or other outcomes between positive airway pressure treatment with and without humidification. Five trials examined different aspects of humidified CPAP treatment for patients with OSA. While some studies reported a benefit of added humidity in CPAP treatment in improving patient compliance, this effect was not consistent across all the studies. Overall, the studies were clinically heterogeneous, small, and of quality B (three studies) or C (two studies).

Comparison of Mandibular Advancement Devices and No Treatment or Inactive Oral Devices

The strength of evidence is moderate to show that the use of mandibular advancement devices (MAD) improves sleep apnea signs and symptoms. Five trials (four quality B, one quality C) compared MAD with no treatment, using a variety of different types of MAD, and found significant improvements with MAD in AHI, ESS, and other sleep study measures. Any differences in quality of life measures or neurocognitive tests were equivocal between treatment groups. No trial evaluated objective clinical outcomes. Another five trials (four quality B, one quality C) compared the effects of MAD with inactive oral devices and reported similar findings.

Comparison of Different Oral Devices

The strength of evidence is insufficient to draw conclusions with regard to the relative efficacy of different types of oral MAD in patients with OSA because the reviewed studies were generally small, and each was concerned with a unique comparison. Five studies (four quality B, one quality C) with unique comparisons found little to no differences between different types and methods of use of MAD or other oral devices in sleep study or sleepiness measures. No study evaluated objective clinical outcomes. Only one study evaluated compliance; no significant differences were observed. One trial found that a greater degree of mandibular advancement resulted in an increased number of patients achieving an AHI <10 events/hr; however, the mean AHI was similar between treatment groups.

Comparison of Mandibular Advancement Devices and CPAP

The strength of evidence is moderate that CPAP is superior to MAD in improving sleep study measures. Ten mostly quality B trials overall found that CPAP resulted in greater reductions in AHI and arousal index, and increases in minimum oxygen saturation. The evidence regarding the relative effects on ESS were too heterogeneous to allow conclusions. In a single study, patients were more compliant with MAD than CPAP (hours used per night and nights used). No study evaluated objective clinical outcomes. *The strength of evidence is insufficient* to address which patients might benefit most from either treatment.

Comparison of Surgery and Control

The strength of evidence is insufficient to evaluate the relative efficacy of surgical interventions for the treatment of OSA. Six trials and one nonrandomized prospective study with unique interventions compared surgery with control treatment for the management of patients with OSA. Three studies were rated quality A, one quality B, and three quality C. The results were inconsistent across studies as to which outcomes were improved with surgery compared with no or sham surgery.

Comparison of Surgery and CPAP

The strength of evidence is insufficient to determine the relative merits of surgical treatments versus CPAP. Of 12 studies (1 quality A, 11 quality C) comparing surgical modalities with CPAP, only two were RCTs, and they compared CPAP with uvulopalatopharyngoplasty (UPPP), removal of the soft tissue at the back of the throat, the uvula, and soft palate. While one of these trials found that CPAP resulted in a higher mortality benefit, the other found no difference between groups. Due to the heterogeneity of interventions and outcomes examined, the variability of findings across studies, and the inherent bias of all but one study regarding which patients received surgery, it is not possible at this time to draw useful conclusions comparing surgical interventions with CPAP in the treatment of patients with OSA. The quality A trial was the only unbiased comparison of surgery and CPAP (patients had previously received neither treatment). It did not find statistically significant differences in ESS and quality of life measures between patients with mild to moderate OSA who had temperature-controlled radiofrequency tissue volume reduction of the soft palate and those who had CPAP at 2 months followup. Likewise, the other trial, comparing maxillomandibular advancement osteotomy and CPAP, did not find statistically significant differences in AHI and ESS in patients with severe OSA. For the nonrandomized studies, comparisons between surgery and CPAP are difficult to interpret since baseline patient characteristics (including sleep apnea severity) differed significantly between groups, particularly in regards to what previous treatments patients had. The reported findings on sleep study and quality of life measures were heterogeneous across studies.

Comparison of Surgery and Mandibular Advancement Devices

The strength of evidence is insufficient regarding the relative merit of MAD versus surgery in the treatment of OSA, as there was only one study (quality B) that examined this question. A statistically significant improvement in AHI was observed in the MAD group compared with the surgery group. No study evaluated objective clinical outcomes.

Comparison of Other Treatments

The strength of evidence is low to show that some intensive weight loss programs may be effective treatment for OSA in obese patients. Three trials (one quality A, two quality B) compared weight loss interventions with control interventions. All three trials found significant relative reductions in AHI with diet. Other outcomes were inconsistent.

The strength of evidence is insufficient to determine the effects of other potential treatments for OSA. Twenty-one studies evaluated other interventions including atrial overdrive pacing, eight different drugs, palatal implants, oropharyngeal exercises, a tongue-retaining device, a positional alarm, combination tongue-retaining device and positional alarm, bariatric surgery, nasal dilator strips, acupuncture, and auricular plaster. All of these interventions were evaluated by one or two studies only. The findings were heterogeneous. No study evaluated objective clinical outcomes.

Key Question 6. In OSA patients prescribed nonsurgical treatments, what are the associations of pretreatment patient-level characteristics with treatment compliance?

Across five studies (one quality A, one quality B, three quality C), *the strength of evidence is moderate* that more severe OSA as measured by higher AHI is associated with greater compliance with CPAP use. Each study measured compliance differently, including thresholds of 1, 2, or 3 hours of use per night or as a continuous variable, and undefined "objective compliance" measured by the device. *The strength of evidence is moderate* that a higher ESS score is also associated with improved compliance. *There are low strengths of evidence* that younger age, snoring, lower CPAP pressure, higher BMI, higher mean oxygen saturation, and the sleepiness domain on the Grenoble Sleep Apnea Quality of Life test are each possible independent predictors of compliance. It is important to note, however, that selective reporting, particularly of nonreporting of nonsignificant associations, cannot be ruled out. The heterogeneity of analyzed and reported potential predictors greatly limits these conclusions. Differences across studies as to which variables were independent predictors may be due to the adjustment for different variables, in addition to differences in populations, outcomes, CPAP machines, and CPAP training and followup. One quality C study of mandibular advancement devices failed to identify potential predictors of compliance.

Key Question 7. What is the effect of interventions to improve compliance with device (positive airway pressure, oral appliances, positional therapy) use on clinical and intermediate outcomes?

The strength of evidence is low that some specific adjunct interventions may improve CPAP compliance, but studies are heterogeneous and no general type of intervention (e.g., education, telemonitoring) was more promising than others. The 18 trials (two quality A, eight quality B, and eight quality C) had inconsistent effects across a wide variety of interventions. Studies generally had small sample sizes with less than 1 year of followup. Compared with usual care, several interventions were shown to significantly increase hours of CPAP use per night in some studies. These included intensive support or literature (designed for patient education), cognitive behavioral therapy (given to patients and their partners), telemonitoring, and a habit-promoting audio-based intervention. However, the majority of studies did not find a significant difference in CPAP compliance between patients who received interventions to promote compliance with device use and those who received usual care. No study of nurse-led care (which was not focused primarily on compliance) showed an effect on compliance rates.

Discussion

The findings of the systematic review have been summarized in Table A. Interventions (either diagnostic tests or treatments) that are not discussed lack studies meeting eligibility criteria. Interventions were not excluded from analysis unless explicitly stated as such.

Diagnosis

In theory, obstructive sleep apnea (OSA) is relatively simple to diagnose. However, PSG, the standard diagnostic test, is inconvenient, resource-intensive, and may not be representative of a typical night's sleep (particularly the first night the test is given). Furthermore, there are variations across laboratories in

the definitions of OSA (using different thresholds of AHI, from 5 to 15 events/hr) and in the way that the PSG results are read and interpreted. Moreover, AHI, which is used as the single metric to define OSA, can vary from night to night and does not take into account symptoms, comorbidities, or response to treatment.

Two approaches have been taken to reduce the resources involved in diagnosing OSA, including tests (questionnaires and clinical prediction rules) to screen for OSA and portable monitors to be used instead of sleep-laboratory PSG. Five questionnaires and 10 validated clinical prediction rules have been compared with PSG. However, very few of the screening tests have been evaluated by more than one set of researchers, and few have been directly compared with each other. Thus, *the strength of evidence is low* that the Berlin questionnaire is accurate in its ability to screen for OSA; the commonly used STOP and STOP-Bang questionnaires have not been adequately tested. For such tests to be of clinical value, apart from having very high sensitivity and specificity, they should be easy to administer and require only information from symptoms and signs easily obtainable during a physical examination. The evaluated clinical prediction models were all internally validated, but definitive conclusions on the external validity (i.e., generalizability) of these predictive rules in independent populations cannot be drawn from the available literature. *The strength of evidence is low* that some clinical prediction rules may be useful in the prediction of a diagnosis of OSA. No study examined the potential clinical utility of applying the questionnaires or prediction rules to clinical practice.

Numerous portable monitors (evaluated in 99 studies) have been developed for use in nonlaboratory settings; these use fewer "channels" (specific physiologic measures) than typical 16-channel PSG. The more recent studies do not substantially change the conclusions from the Tufts Evidence-based Practice Center's 2007 Technology Assessment on *Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome*[c]. Although most of the tested portable monitors fairly accurately predict OSA, it is unclear whether any of these monitors can replace laboratory-based PSG. The evidence suggests that the measured AHI from portable monitors is variable compared with PSG-derived AHI, but the source of this variability is unclear. So far, no studies have evaluated the predictive ability for clinical outcomes or response to treatment by portable monitors. Furthermore, no available studies have evaluated the impact of patient triage via screening tests and/or portable monitors.

The value of preoperative screening for OSA remains poorly defined. The only study that directly addressed this question was a retrospective study of patients undergoing bariatric surgery. It showed better perioperative outcomes from routine PSG. There are also no adequate studies that compared phased testing (simple tests followed by more intensive tests in selected patients) with full evaluation (by PSG).

Apnea-Hypopnea Index as a Predictor of Clinical Outcomes

The strength of evidence is high that high baseline (>30 events/hr or range) AHI is a strong and independent predictor of all-cause mortality over several years of followup, with the association being strongest among people with severe OSA (AHI >30 events/hr). However, the strength of evidence for the association between baseline AHI and other long-term clinical outcomes is generally insufficient, and thus the association between reductions in AHI by OSA treatment and improvements in long-term outcomes remains theoretical.

Treatment

The strength of evidence is moderate that fixed CPAP is an effective treatment to minimize AHI and improve sleepiness symptoms, as supported by more than 40 trials of patients treated with CPAP or no treatment. However, no trial reported long-term clinical outcomes, and compliance with CPAP treatment is poor. Because patients frequently do not tolerate CPAP, many alternative treatments have been proposed. First, several alternative CPAP machines have been designed to vary the pressure during the patient's inspiratory cycle or to titrate the pressure to a minimum necessary level. Other modifications include different masks, nasal pads, and added humidification. The large majority of relevant trials have compared autotitrating CPAP (autoCPAP) with fixed CPAP and the strength of evidence of no clinical differences between them is *moderate*. *The strength of evidence is insufficient* for other device comparisons and, overall, the evidence does not support the use of one device for all patients, since such decisions should be individualized.

The second alternative to CPAP therapeutic option is the use of oral devices, which have been designed with the goal of splinting open the oropharynx to prevent obstruction. The most commonly tested are the mandibular advancement devices (MAD), for which the strength of evidence for their efficacy in sleep outcomes is *moderate*. Based on direct and indirect comparisons, CPAP appeared to be more effective than MAD. However, given the issues with noncompliance with CPAP, the decision as to whether to use CPAP or MAD will likely depend on patient preference.

The third major alternative to OSA treatment includes surgical interventions to alleviate airway obstruction. Given the very few randomized trials and the differences in the populations that choose to undergo surgery versus conservative treatment, the strength of evidence is insufficient to determine the relative value of surgery to no treatment, to CPAP, to MAD, or to alternative types of surgery. Additional interventions were also evaluated in randomized trials, (including weight loss programs, atrial overdrive pacing, eight different drugs, and other interventions) but in general *the strength of evidence is insufficient* to determine the effects of these potential treatments.

For all the treatment comparisons, it is important to identify which subgroups of patients may benefit most from specific treatments. Unfortunately, the trials are nearly silent on this issue. Very few trials reported subgroup analyses based on baseline characteristics, and for most comparisons there were too few studies or the interventions examined were too heterogeneous to analyze potential differences. Such analyses were feasible for the comparison between CPAP and control, where subgroup meta-analyses based on definitions of OSA (different minimum AHI thresholds) failed to demonstrate any difference in effectiveness of CPAP in reducing AHI or ESS. Though statistical heterogeneity existed across the trials, this was primarily attributed to study design factors that have no clinical implications. Despite statistical heterogeneity, and based on the consistency of findings that support CPAP as effective to minimize AHI in all patients with OSA, it is reasonable to conclude that the relative effectiveness in different populations is a moot point. The one exception to this may be patients with mild OSA (with AHI <15 events/hr), since people with low AHI cannot have as large an improvement in their AHI as people with severe OSA. Notably, across interventions there is little evidence supporting the hypothesis that any OSA treatment improves quality of life or neurocognitive function.

The strength of evidence is insufficient regarding the effect of interventions to improve CPAP compliance. The studies were very heterogeneous in their interventions and each evaluated different interventions. Higher baseline AHI and increased sleepiness as measured by the Epworth Sleepiness Scale are both predictors of improved compliance with CPAP (high strength and moderate strength of evidence, respectively). The unsurprising interpretation of this finding is that patients with more severe symptoms are more likely to accept the discomfort or inconvenience of using CPAP overnight.

Limitations

The most important limitations in the evidence were the lack of trials that evaluated long-term clinical outcomes, the sparseness of evidence to address several Key Questions, and the fact that no study of diagnostic tests or treatments attempted to assess how results may vary in different subgroups of patients. In general, the intervention trials were of quality B or C, with few quality A studies. Followup durations tended to be very short, and study dropout rates were frequently very high. Other frequent methodological problems with studies included incomplete reporting and/or inadequate analyses, which required estimations of pertinent results by the authors of this systematic review. The heavy reliance on industry support for trials of devices may lead to the concern of publication bias. However, this concern may be reduced since most of our conclusions were that the strength of evidence is either low or inadequate for interventions. Furthermore, the effects of CPAP and MAD on sleep measures are sufficiently large that conclusions about the effectiveness of these devices would be unlikely to change with the addition of unpublished trials.

Implications for Future Research

General Recommendation

- The recurrent problem of high dropout rates as evidenced in the literature we reviewed bears further investigation and is crucial for the conduct of future

trials. It is important to understand whether this a problem peculiar to this field, whether patients' symptoms interfere with their desire to fulfill their obligations as research participants, whether patients are not well informed about the serious consequences of sleep apnea and therefore are less motivated to comply with followup, or whether the treatments are so onerous that patients are refusing to continue with them.

Diagnostic Tests

- The most clinically useful evaluation of prediction rules and questionnaires (to screen for or diagnose OSA) would be trials to evaluate whether use of the tests improves clinical outcomes. Individual patient-data meta-analysis of measurements with portable monitors would provide insights on the diagnostic information contributed by different neurophysiologic signals. Future studies of the accuracy or bias of diagnostic tests should focus more on head-to-head comparisons of portable monitors, questionnaires, and prediction rules, to determine the optimal tool for use in a primary care setting to maximize initial evaluation of OSA and triage high-risk patients for prompt PSG. Direct comparisons among existing alternatives to PSG are more important than the current focus on developing new diagnostic tests.
- Trials are needed comparing potential phased testing strategies with direct PSG or addressing the value of preoperative screening for OSA. Studies of appropriate tests for patients, based on the type or severity of their symptoms, would be useful.

Treatments

- Only 3 of the 190 studies of treatments reported clinical outcomes; comparative studies focusing on long-term followup and clinical outcomes are needed.
- Fixed CPAP is clearly an effective treatment for OSA, and no further trials are needed to assess its efficacy, with the exception of trials assessing long-term clinical outcomes. All other interventions should either be:
  - directly compared with fixed CPAP, among patients naïve to CPAP, or
  - compared with no treatment or alternative treatment among patients who have failed to comply with CPAP treatment.
- Treatment effect heterogeneity should be investigated.
- The benefit from different degrees of mandibular advancement has to be determined.
- Head-to-head comparisons are needed of alternative treatments for patients who do not tolerate CPAP.
- Rigorously conducted head-to-head comparisons of surgical interventions versus CPAP are needed to overcome limitations of existing observational evidence.
- More studies are needed on the various additional interventions (including weight loss, drugs, and specific exercises), and their incremental benefit to accepted treatments for OSA should be examined.
- Interventions to improve compliance to CPAP and MAD should be tested in direct comparisons.

Predictors of Clinical Outcomes and Compliance

- The question of whether OSA severity is associated with long-term outcomes (beyond all-cause mortality) may be informed by patient-level meta-analyses of available large cohorts.
- Predictive models of compliance and response to treatment are needed.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 1: Diagnosis Portable monitors vs. PSG	Low (Type II monitors); Moderate (Types III & IV monitors)	• No recent studies have compared Type II portable monitors to PSG[d]. A prior systematic review concluded that "based on [3 quality B studies], Type II monitors [used at home] may identify AHI suggestive of OSA with high positive likelihood ratios and low negative likelihood ratios," though "substantial differences in the [measurement of] AHI may be encountered between Type II monitors and facility-based PSG." • There were 29 studies that compared Type III portable monitors[e] with PSG. 7 of these are new since a previous report. 18 Type III monitors have been evaluated. • There were 70 studies that compared Type IV portable monitors[f] to PSG. 24 of these are new since a previous report. 23 Type IV monitors have been evaluated. • Overall, 15 studies were graded quality A, 45 quality B, and 39 quality C. The studies were applicable to the general population of patients being referred to specialized sleep centers or hospitals for evaluation of suspected sleep apnea. It is unclear if the studies are applicable to patients with comorbidities or who may have central sleep apnea. Most of the studies were conducted either in the sleep laboratory setting or at home. • Studies measured either concordance (comparisons of estimates of AHI), test sensitivity and specificity (to diagnose OSA as defined by PSG), or both. • Type III monitors had a wide range of mean biases (difference in AHI estimate from PSG), from -10 to +24 events/hr, with wide limits of agreements within studies. • Type IV monitors had a wide range of mean biases, from -17 to +12 events/hr, with wide limits of agreements within studies. • To diagnose OSA defined as a PSG AHI ≥5 events/hr, Type III monitors had sensitivities of 83–97% and specificities of 48–100%. Type III monitors commonly less accurately diagnosed OSA with AHI ≥15 events/hr, with sensitivities 64–100% and specificities 41–100%. • Evaluation of positive and negative likelihood ratios, and available ROC curves, suggest that Type III monitors are generally accurate in diagnosing OSA (as measured by PSG), with high positive likelihood ratios, low negative likelihood ratios, and high AUC. • To diagnose OSA, Type IV monitors had a very wide range of sensitivities and specificities. • Across studies (by indirect comparison), the range of sensitivities and specificities of both Type III and Type IV monitors largely overlapped, thus not demonstrating greater accuracy with either type of monitor.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 1: Diagnosis Portable monitors vs. PSG (continued)	Low (Type II monitors); Moderate (Types III & IV monitors) (continued)	• **Conclusion**: The strength of evidence is low that Type II monitors are accurate to diagnose OSA (as defined by PSG), but have a wide and variable bias in estimating the actual AHI. • **Conclusion**: The strength of evidence is moderate that Type III and IV monitors are generally accurate to diagnose OSA (as defined by PSG), but have a wide and variable bias in estimating the actual AHI. The evidence is insufficient to adequately compare specific monitors to each other.
Key Question 1: Diagnosis Questionnaires vs. PSG	Low/Insufficient	• There were 6 studies that compared 6 questionnaires with PSG diagnosis of OSA. Overall, these studies are applicable to patients visiting preoperative clinics, sleep laboratories, and primary care centers for evaluation of sleep apnea. • There were 1 quality A and 3 quality C studies that evaluated the Berlin Questionnaire (based on snoring, tiredness, and blood pressure), with OSA defined as AHI ≥5 events/hr; sensitivity ranged from 69–93%, specificity ranged from 56–95%. With an AHI ≥15 events/hr definition, sensitivity was somewhat lower and specificity was similar. To predict severe OSA (AHI ≥30 events/hr), sensitivity and specificity were generally lower. • Each of the following 4 questionnaires was evaluated in a single study (1 quality B, 2 quality C): STOP, STOP-Bang, ASA checklist, Hawaii Sleep Questionnaire, which all had relatively low specificity for OSA (AHI thresholds of 5, 10, or 30 events/hr), ranging from 37–67%. STOP, ESS, and the Hawaii Questionnaire had sensitivities <80%. STOP-Bang had high sensitivity to predict diagnosis of OSA, particularly those with AHI ≥15 or ≥30 events/hr (93 and 100%, respectively). The American Society of Anesthesiologists Checklist had a sensitivity of 87% to predict severe OSA, but lower sensitivity to predict those with lower AHI. In 1 quality A study, ESS had a low sensitivity (49%) and higher specificity (80%) to predict OSA with AHI ≥5. • **Conclusion**: The strength of evidence is low that the Berlin Questionnaire is moderately accurate (sensitivity and specificity generally <90%) to screen for OSA. The strength of evidence is insufficient to evaluate other questionnaires, but 1 study found that STOP-Bang may have high enough sensitivity to accurately screen for OSA.
Key Question 1: Diagnosis Clinical prediction, rules vs. PSG	Low	• There were 7 studies that compared 10 validated clinical prediction rules with PSG (3 quality A, 3 quality B, 1 quality C). Only 1 model has been externally validated (by independent researchers); thus the applicability of the studies to the general population is unclear. Of the models, 8 include variables obtainable through routine clinical history and examination.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 1: Diagnosis Clinical prediction, rules vs. PSG (continued)	Low (continued)	• A single morphometric model and a model that included pulmonary function test data had near perfect discrimination (AUC=0.996) or sensitivity (100%), but neither was independently validated. The other clinical prediction rules had variable accuracy for predicting OSA (AHI ≥5, 10, or 15 events/hr) or severe OSA (AHI ≥30 events/hr). • **Conclusion**: The strength of evidence is low that some clinical prediction rules may be useful in the prediction of a diagnosis of OSA.
Key Question 2: Diagnosis Phased testing	Insufficient	• A single quality C study partially addressed the value of phased testing, but had substantial verification bias due to implementation of the phased testing. • **Conclusion**: The strength of evidence is insufficient to determine the utility of phased testing.
Key Question 3: Diagnosis Preoperative screening	Insufficient	• There were 2 quality C studies that assessed the effect of preoperative screening for OSA on surgical outcomes, though only 1 of these was designed to address the question. • The retrospective study that compared mandatory prebariatric-surgery PSG with PSG performed based on clinical parameters (performed during different time periods) did not find significant differences in outcomes. The other study found only that those patients who volunteered for preoperative PSG were more likely to suffer cardiopulmonary perioperative complications than patients who refused PSG. • **Conclusion**: The strength of evidence is insufficient to determine the utility of preoperative sleep apnea screening.
Key Question 4: Predictors AHI as a predictor of long-term clinical outcomes	Variable (High for all-cause mortality; Low for diabetes; Insufficient for other long-term clinical outcomes)	• There were 11 studies (of 8 large cohorts) that performed multivariable analyses of AHI as an independent predictor of long-term clinical outcomes. • There were 4 studies (3 quality A, 1 quality B) that evaluated all-cause mortality. All found that AHI was a statistically significant independent predictor of death during 2–14 years of followup. The association was strongest among people with an AHI >30 events/hr. There was 1 study, however, that found an interaction with sex and age such that AHI was associated with death only in men ≤70 years of age. The evidence on mortality is applicable to the general population, with and without OSA, and also more specifically to men with OSA symptoms or evidence of OSA. • There were 2 quality A studies that evaluated cardiovascular mortality. There was 1 study that found that only AHI >30 events/hr predicted cardiovascular death; the other study found no association.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)		
Key Question	**Strength of Evidence**	**Summary/Conclusions/Comments**
Key Question 4: Predictors AHI as a predictor of long-term clinical outcomes (continued)	Variable (High for all-cause mortality; Low for diabetes; Insufficient for other long-term clinical outcomes) (continued)	• A single quality A study evaluated nonfatal cardiovascular disease and similarly found that only AHI >30 events/hr was an independent predictor. • A single quality B study suggested that the association between AHI and stroke may be confounded by obesity. • There were 2 studies (1 quality A, 1 quality B) that came to uncertain conclusions regarding the possible association between AHI and incident hypertension. • There were 2 studies (1 quality A, 1 quality B) that suggested an association between AHI and incident type 2 diabetes, though 1 study found that the association was confounded by obesity. • A single quality A study found no significant association between AHI and future quality of life (SF-36 after 5 years). This conclusion appears to be applicable for both the general population and specifically for patients diagnosed with sleep disordered breathing. • **Conclusion**: The strength of evidence is high that an AHI >30 events/hr is an independent predictor of all-cause mortality; although one study found that this was true only in men under age 70. The strength of evidence is low that a higher AHI is associated with incident diabetes, though possibly confounded with obesity. The strength of evidence is insufficient to determine the association between AHI and other clinical outcomes.
Key Question 5: Treatment OSA treatments CPAP vs. control with OSA	Moderate	• There were 43 trials that compared CPAP devices with either no treatment or sham CPAP. All but 2 evaluated fixed CPAP. Of the 43 trials, 4 were rated quality A, 22 quality B, and 17 quality C. Overall, the studies are applicable to a broad range of patients with OSA. • Only 1 study evaluated a clinical outcome, namely heart failure symptomatology, and found no significant effect after 3 months. • By meta-analysis, CPAP results in a statistically significant large reduction in AHI (-20 events/hr compared with no treatment and -46 events/hr compared with sham CPAP). All studies found statistically significant effects, though there was statistical heterogeneity across studies that could not be fully explained. There were no clear, consistent relationships across studies between definition of OSA (by minimum threshold AHI) or other clinical features and effect size. • By meta-analysis, CPAP results in a statistically and clinically significant improvement in sleepiness as measured by ESS (-2.6 compared with no treatment and -2.7 compared with sham CPAP). The studies were statistically significant and most, but not all, found significant improvements in ESS. No factors clearly explained the heterogeneity. • CPAP also generally resulted in improvements in other sleep study measures, but had inconsistent effects on other sleepiness tests, quality of life tests, neurocognitive tests, and blood pressure.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 5: Treatment OSA treatments CPAP vs. control with OSA (continued)	Moderate (continued)	• All adverse events related to CPAP treatment were potentially transient and could be alleviated with either stopping treatment or with ancillary interventions. Generally about 5-15% of patients in trials had specific adverse events they considered to be a major problem while using CPAP. These included claustrophobia, oral or nasal dryness, epistaxis, irritation, pain, and excess salivation. No adverse event with potentially long-term consequences was reported. • **Conclusion**: Despite no evidence or weak evidence on clinical outcomes, given the large magnitude of effect on the important intermediate outcomes AHI, ESS, and other sleep study measures, the strength of evidence is moderate that CPAP is an effective treatment for OSA. However, the strength of evidence is insufficient to determine which patients might benefit most from treatment.
Key Question 5: Treatment OSA treatments Different CPAP devices vs. each other	Variable (Moderate for autoCPAP vs. CPAP; Low for C Flex vs. CPAP; Insufficient for others)	• No study evaluated clinical outcomes. • There were 21 trials that compared autoCPAP with fixed CPAP. Of these, 1 trial was rated quality A; 10 trials each were rated quality B or C. These studies are applicable mainly to patients with AHI more than 15 events/hr and BMI more than 30 kg/m^2. By meta-analysis there was statistically significant, but clinically nonsignificant better improvement in ESS (-0.5), minimum oxygen saturation (1%), and compliance (11 minutes) with autoCPAP than fixed CPAP, and no statistically significant differences in AHI or arousal index. • There were 4 trials comparing C-Flex™ to fixed CPAP. No statistically significant differences were found for compliance, sleep study measures, or other tested outcomes. • There were 14 trials comparing bilevel or flexible bilevel CPAP with fixed CPAP, humidification with no humidification (with fixed CPAP), or oral with nasal fixed CPAP. The studies had either inconsistent results, were sparse, or had imprecise results. • **Conclusion**: Despite no or weak evidence on clinical outcomes, overall, there is moderate strength of evidence that autoCPAP and fixed CPAP result in similar compliance and treatment effects for patients with OSA. • **Conclusion**: The strength of evidence is low of no substantial difference in compliance or other outcomes between C-Flex and CPAP. • **Conclusion**: The strength of evidence is insufficient regarding comparisons of different CPAP devices (or modifications).
Key Question 5: Treatment OSA treatments MAD vs. control	Moderate	• There were 10 trials comparing various MADs with either no treatment or with sham devices (without mandibular advancement). No studies were rated quality A, 8 quality B, 2 quality C. The studies are generally applicable to patients with AHI ≥15 events/hr, though less so to patients with comorbidities or excessive sleepiness. All studies excluded edentulous patients or those with periodontal diseases.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 5: Treatment OSA treatments MAD vs. control (continued)	Moderate (continued)	• No study evaluated clinical outcomes. • By meta-analysis, MAD results in a statistically significant reduction in AHI (-12 events/hr). All studies found statistically significant improvements in AHI, ranging from -6 to -25 events/hr, without statistical heterogeneity. • By meta-analysis, MAD results in a statistically and clinically significant improvement in sleepiness as measured by ESS (-1.4). Of 8 studies, 5 found statistically and clinically significant improvements in ESS, ranging from -1 to -4.5, without statistical heterogeneity. • MAD also generally resulted in improvements in other sleep study measures, but had inconsistent effects on or inadequate evidence for other outcomes of interest. • There was insufficient evidence to address whether study heterogeneity could be explained by different definitions of OSA or other clinical factors, particularly in light of the clinical heterogeneity across studies due to the difference in MADs. • In 2 studies about 5% of patients had tooth damage (or loosening). Substantial jaw pain was reported in about 2–4% of patients, but no study reported on the long-term consequences of any adverse events. • **Conclusion**: Despite no evidence or weak evidence on clinical outcomes, given the large magnitude of effect on the important intermediate outcomes AHI, ESS, and other sleep study measures, overall, the strength of evidence is moderate that MAD is an effective treatment for OSA in patients without comorbidities (including periodontal disease) or excessive sleepiness. However, the strength of evidence is insufficient to address which patients might benefit most from treatment.
Key Question 5: Treatment OSA treatments Oral devices vs. each other	Insufficient	• There were 5 trials comparing different oral devices; 3 compared different MADs; 2 compared different tongue devices. Of these 5 trials, 4 were rated quality B and 1 quality C. These studies are applicable mostly to patients with AHI of15 to 30 events/hr and BMI less than 30 kg/m^2. All studies were restricted to patients with a sufficient number of teeth to anchor the mandibular devices in place. • No study evaluated clinical outcomes. In general, the studies found no differences among devices in sleep study or other measures. Only 1 study (comparing 2 tongue-retaining devices) evaluated compliance and found no difference. • **Conclusion**: The strength of evidence is insufficient regarding comparisons of different oral devices.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)		
Key Question	**Strength of Evidence**	**Summary/Conclusions/Comments**
Key Question 5: Treatment OSA treatments CPAP vs. MAD	Moderate	• There were 10 trials comparing different MADs with CPAP. A single study of an extraoral device vs. autoCPAP was rated quality C; 9 studies of oral MAD vs. fixed CPAP were rated quality B. The studies are generally applicable to patients with AHI >5-10 events/hr. • No study evaluated clinical outcomes. • A single study compared compliance rates, finding that patients used MAD significantly more hours per night and nights per week than CPAP. • There were 2 studies that found that CPAP was significantly more likely to result in 50% reductions in AHI and achieved AHI <5 events/hr, but 1 study found no difference in achieving <10 events/hr. By meta-analysis, CPAP resulted in significantly greater reductions in AHI (-8 events/hr); 7 of 9 studies found statistically significant differences. By meta-analysis, CPAP results in a statistically significant greater improvement in AHI than MAD (-8 events/hr). • The studies had inconsistent findings regarding the relative effects of MAD and CPAP on ESS. • The studies generally found superior effects of CPAP over MAD for other sleep study measures, but no differences in quality of life or neurocognitive function. • A single study found no differences with either device in achieving an AHI of either <5 or <10 events/hr based on baseline severity of OSA (at an AHI threshold of 30 events/hr). • **Conclusion**: Despite no evidence or weak evidence on clinical outcomes, overall the strength of evidence is moderate that the use of CPAP is superior to MAD. However, the strength of evidence is insufficient to address which patients might benefit most from either treatment.
Key Question 5: Treatment OSA treatments Surgery vs. control	Insufficient	• There were 7 studies comparing 7 different surgical interventions to sham surgery, conservative therapy, or no treatment. Of these, 3 studies were rated quality A, 1 quality B, and 3 quality C. • No study evaluated clinical outcomes. • Of these 7 studies, 4 found statistically significant improvements in AHI, other sleep study measures, and/or sleepiness measures. The remaining studies found no differences in these outcomes or quality of life or neurocognitive function.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 5: Treatment OSA treatments Surgery vs. control (continued)	Insufficient (continued)	• Adverse events from surgery (also evaluated from large surgical cohort studies) were generally due to perioperative complications, including perioperative death in about 1.5% in two studies of UPPP—though most studies reported no deaths, hemorrhage, nerve palsies, emergency surgical treatments, cardiovascular events, respiratory failure, and rehospitalizations. Long-term adverse events included speech or voice changes, difficulties swallowing, airway stenosis, and others. In smaller studies, when these adverse events were reported they occurred in about 2–15% of patients. However the largest 2 studies (of 3,130 UPPP surgeries and 422 RFA surgeries) reported no long-term complications (not including perioperative death or cardiovascular complications). • **Conclusion**: Overall, the strength of evidence is insufficient to evaluate the relative efficacy of surgical interventions for the treatment of OSA.
Key Question 5: Treatment OSA treatments Surgery vs. CPAP	Insufficient	• Of 12 eligible studies comparing surgery with CPAP (1 quality A, 11 quality C), only 2 were RCTs. • There were 2 retrospective studies that evaluated mortality in UPPP vs. CPAP. Of these, 1 study found higher mortality over 6 years among patients using CPAP (HR = 1.31; 95% CI 1.03, 1.67) and 1 study found no difference in 5-year survival. • Both trials found no difference in outcomes either between RFA and CPAP after 2 months or between maxillomandibular advancement osteotomy and CPAP at after 12 months. The remaining studies were heterogeneous in their conclusions. • **Conclusion**: The strength of evidence is insufficient to determine the relative merits of surgical treatments versus CPAP.
Key Question 5: Treatment OSA treatments Surgery vs. MAD	Insufficient	• A single trial (quality B) compared UPPP and MAD treatment. • The trial did not evaluate clinical outcomes. The study found that significantly more patients using MAD achieved 50% reductions in AHI at 1 year and significantly lower AHI at 4 years. • **Conclusion**: The strength of evidence is insufficient to determine the relative merits of surgical treatments versus MAD.
Key Question 5: Treatment OSA treatments/ Other treatments	Variable (Low for weight loss vs. control; Insufficient for others)	• There were 3 trials (1 quality A, 2 quality B) comparing weight loss interventions with control interventions. The studies were heterogeneous in terms of baseline OSA severity, presence of comorbidities, and severity of obesity. The studies are generally applicable to people with BMI >30 kg/m^2. • No study evaluated clinical outcomes. • A single study found increased odds of achieving an AHI <5 events/hr after 1 year of a very low calorie diet compared with no treatment (OR=4.2, 95% CI 1.4, 12). All 3 trials found significant relative reductions in AHI with diet, from -4 to -23 events/hr. Other outcome data are inconsistent or sparse.

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)

Key Question	Strength of Evidence	Summary/Conclusions/Comments
Key Question 5: Treatment OSA treatments/ Other treatments (continued)	Variable (Low for weight loss vs. control; Insufficient for others) (continued)	• A total of 19 studies evaluated 21 other interventions including atrial overdrive pacing, 8 different drugs, palatal implants, oropharyngeal exercises, a tongue-retaining device, a positional alarm, combination tongue-retaining device and positional alarm, bariatric surgery, nasal dilator strips, acupuncture, and auricular plaster. All of these interventions were evaluated by 1 or 2 studies only. No study evaluated clinical outcomes. • **Conclusion**: The strength of evidence is low to show that some intensive weight loss programs are effective treatment for OSA in obese patients. • **Conclusion**: The strength of evidence is insufficient to determine the effects of other potential treatments for OSA.
Key Question 6: Predictors Predictors of treatment compliance	Variable (see Conclusions)	• There were 5 large cohort studies that conducted multivariable analyses of potential predictors of compliance with CPAP treatment. Of these, 1 study was rated quality A, 1 quality B, and 3 quality C. In general, the studies are applicable to patients initiating CPAP whose AHI is greater than 30 events/hr. • Of these 5 cohort studies, 4 studies all found that higher baseline AHI was associated with greater compliance. Also, 2 of 3 studies found that higher baseline ESS was a predictor of greater compliance. And 2 of 3 studies found that age was not a predictor of compliance. Only 1 or 2 studies evaluated other potential predictors, with no consistent findings. • A single quality C cohort study evaluated potential predictors of compliance with newly initiated MAD. The study did not identify any statistically significant predictors. • **Conclusion**: The strength of evidence is moderate that more severe OSA as measured by higher AHI is associated with greater compliance with CPAP use. The strength of evidence is moderate that higher ESS is also associated with improved compliance. • **Conclusion**: The strength of evidence is insufficient regarding potential predictors of compliance with MAD.
Key Question 7: Treatment Treatments to improve compliance	Low	• There were 18 trials evaluating interventions to improve CPAP compliance. Of these, 2 were rated quality A, 8 quality B, and 8 quality C. These studies are mostly applicable to patients initiating CPAP with AHI >30 events/hr and BMI greater than 30 kg/m^2. No study evaluated interventions to improve compliance with other devices. • There were 9 studies evaluating extra support or education. These studies had inconsistent findings regarding the effect of the interventions on compliance. Only 3 of 7 studies found increased number of hours of CPAP use; only 1 of 3 studies found persistent improved compliance (and that was of compliance with followup visits).

Table A. Summary of findings of studies addressing Key Questions on obstructive sleep apnea (continued)		
Key Question	**Strength of Evidence**	**Summary/Conclusions/Comments**
Key Question 7: Treatment Treatments to improve compliance (continued)	Low (continued)	• There were 3 studies evaluating telemonitoring. No study found a statistically significant increase in CPAP usage (hours per night). • A single study evaluated the effect of cognitive behavioral therapy, and showed that the behavioral intervention significantly increased hours of CPAP use per night compared with usual care (difference = 2.8 hours; 95% CI 1.8, 3.9; P<0.0001). • There were 2 studies evaluating 2 other interventions: the hypnotic zolpidem and nasal pillows. No intervention was found to be effective to improve compliance. • There were 3 studies evaluating nursing care models. None improved compliance. • **Conclusion**: The strength of evidence is low that some specific adjunct interventions may improve CPAP compliance among overweight patients with more severe OSA who are initiating CPAP treatment. However, studies are heterogeneous and no general type of intervention (e.g., education) was more promising than others.

AHI = apnea-hypopnea index, AUC = area under the ROC curve, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure, ESS = Epworth Sleepiness Scale, HR = hazard ratio, MAD = mandibular advancement device, OSA = obstructive sleep apnea, PSG = polysomnography (sleep-laboratory based), RFA = radiofrequency ablation, ROC = receiver operating characteristics, SF-36 = Short Form Health Survey 36, UPPP = uvulopalatopharyngoplasty.

Notes

a. Please refer to the main report for references.

b. Criteria for selecting topics for systematic review include appropriateness, importance, lack of duplication, feasibility, and potential value. See http://www.effectivehealthcare.ahrq.gov/index.cfm/submit-a-suggestion-for-research/how-are-research-topics-chosen/.

c. Tufts-New England Medicine Center EPC. Home diagnosis of obstructive sleep apnea-hypopnea syndrome. Health Technology Assessment Database www.cms.gov/determinationprocess/downloads/id48TA.pdf. 2007;2010.

d. Type II monitors are portable devices that record all the same information as PSG (type I monitors).

e. Type III monitors are portable devices that contain at least two airflow channels or one airflow and one effort channel.

f. Type IV monitors comprise all other devices that fail to fulfill criteria for Type III monitors. They include monitors that record more than two physiological measures as well as single channel monitors.

Full Report

This executive summary is part of the following document: Balk EM, Moorthy D, Obadan NO, Patel K, Ip S, Chung M, Bannuru RR, Kitsios GD, Sen S, Iovin RC, Gaylor JM, D'Ambrosio C, Lau J. Diagnosis and Treatment of Obstructive Sleep Apnea in Adults. Comparative Effectiveness Review No. 32. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-100551). AHRQ Publication No. 11-EHC052-EF. Rockville, MD: Agency for Healthcare Research and Quality. July 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

For More Copies

For more copies of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults: Executive Summary No. 32 (AHRQ Publication No. 11-EHC052-1), please call the AHRQ clearinghouse at 1-800-358-9295 or email ahrqpubs@ahrq.gov.

AHRQ Pub. No. 11-EHC052-1
July 2011