WASHINGTON STATE HEALTH CARE AUTHORITY

Sleep Apnea Diagnosis and Treatment in Adults

Health Technology Assessment

FINAL EVIDENCE REPORT

Date: Wednesday, February 15th, 2012

Health Technology Assessment Program
676 Woodland Square Loop SE
P.O. Box 42712
Olympia, WA 98504-2712
http://www.hta.hca.wa.gov

Sleep Apnea Diagnosis and Treatment in Adults

February 2012

Center for Evidence-based Policy
Oregon Health & Science University
3455 SW US Veterans Hospital Road
Mailstop SN-4N, Portland, OR 97239-2941
Phone: 503.494.2182
Fax: 503.494.3807
http://www.ohsu.edu/ohsuedu/research/policycenter/med/index.cfm

Acknowledgements

This report was prepared by:

Ken Gleitsmann, MD, MPH
Heidi Kriz, MPH, RD
Aasta Thielke, MPH
Kendra Bunker
Katharine Ryan, MPH
Katie Lorish
Valerie King, MD, MPH

Center for Evidence-based Policy
Oregon Health and Science University

Suggested citation:

Gleitsmann, K., Kriz, H., Thielke, A., Bunker, K., Ryan, K., Lorish, K., & King, V. (2012). *Sleep apnea diagnosis and treatment in adults.* Portland, OR: Center for Evidence-based Policy, Oregon Health and Science University.

This technology assessment report is based on research conducted by a contracted technology assessment center, with updates as contracted by the Washington State Health Care Authority. This report is an independent assessment of the technology question(s) described based on accepted methodological principles. The findings and conclusions contained herein are those of the investigators and authors who are responsible for the content. These findings and conclusions may not necessarily represent the views of the HCA/Agency and thus, no statement in this report shall be construed as an official position or policy of the HCA/Agency.

The information in this assessment is intended to assist health care decision makers, clinicians, patients and policy makers in making sound evidence-based decisions that may improve the quality and cost-effectiveness of health care services. Information in this report is not a substitute for sound clinical judgment. Those making decisions regarding the provision of health care services should consider this report in a manner similar to any other medical reference, integrating the information with all other pertinent information to make decisions within the context of individual patient circumstances and resource availability.

Table of Contents

Executive Summary

Background

Sleep apnea refers to sleep-disordered breathing due to the recurrent collapse of pharyngeal tissues resulting in snoring, fitful sleep, and daytime somnolence. These episodes are characterized by either reduced airflow (hypopnea), or a complete obstruction (apnea), with a subsequent drop in oxygen saturation, interfering with gas exchange. Obstructive sleep apnea (OSA) is a cause of significant morbidity and mortality and is associated with hypertension, neuropsychological impairment, motor vehicle accidents, stroke, cardiovascular disease, diabetes, and decreased quality of life. The prevalence of OSA is 2 to 7% in the general adult population. Prevalence increases steadily with age, to approximately 20% among people older than age 60. Nationally, rates are also rising, likely due to the increasing frequency of obesity in the general population. Risk factors for OSA include male gender, age, obesity, airway characteristics, familial/genetic predisposition, smoking, and alcohol consumption. The majority of patients with OSA are asymptomatic, unaware of their sleep disordered breathing and associated health risks. As a result, most patients do not seek medical care and remain untreated.

The diagnosis as well as the treatment of OSA is complicated by the difficulty in defining the syndrome. There is controversy surrounding the parameters to be used in a clinical definition as well as which diagnostic method is most appropriate to detect OSA. The current standard for diagnosing OSA is polysomnography (PSG) administered in a sleep study facility. The frequency of obstructed breathing events (i.e., the apnea-hypopnea index (AHI)), combined with multiple other clinical features of obstruction (e.g., oxygen desaturation, air flow, choking episodes) are recorded during sleep. A diagnosis of OSA is generally made when AHI is greater than or equal to 15 or greater than 5 with noticeable daytime symptoms. Considerable costs and patient inconvenience are involved in a PSG study.

Portable PSG monitors, various questionnaires, and predictive models using anatomic and demographic variables have been developed to aid in screening candidates for referral for further diagnostic testing (e.g., sleep lab PSG). The results of these various diagnostic tests may differ in subgroups of patients based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke or airway characteristics. Patients with OSA are at higher risk of postoperative complications both in surgery for OSA and for unrelated surgeries. These risks can be cardiac, respiratory, or anesthesia related. Since OSA is commonly undiagnosed, it has been postulated that screening for OSA may optimize perioperative management.

There have been various modalities developed to treat OSA, most attempting to reduce the airway obstructive component. Continuous positive airway pressure (CPAP) is the first-line therapy for OSA and opens the airway with compressed air. However, the CPAP machinery required is poorly tolerated and compliance is a major concern. Various oral appliances, which attempt to splint open the airway, have been used as an alternative to CPAP. Surgical

procedures, including various surgeries on the oropharyngeal anatomy to alter airway mechanics, are performed to treat OSA. Bariatric surgery may be performed to reduce the volume of obstructive tissues. Other interventions that have been used to treat OSA include: weight loss regimens; smoking cessation; caffeine and alcohol avoidance; positional therapy; oropharyngeal physical therapy to strengthen the musculature and reduce obstruction; arrhythmia treatment for nocturnal bradycardia; complementary and alternative medicine (e.g., acupuncture), and a variety of pharmacologic agents.

Obstructive sleep apnea is a cause of significant morbidity and mortality, and is thus an important public health issue. In addition, the diagnosis and treatment of OSA have societal cost implications, making cost-effectiveness a concern for both of these aspects.

Methods

Key Questions #1 to #7

At the direction of the Washington Health Technology Assessment Program (HTA), the recent Agency for Healthcare Research and Quality (AHRQ) systematic review, *Comparative Effectiveness of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults*, was identified as the primary evidence source for Key Questions #1 to #7 (Balk [AHRQ] 2011). The Balk [AHRQ] (2011) review searched MEDLINE®, Cochrane Central Trials Registry®, and Cochrane Systematic Reviews® for literature published in English. The literature search dates were inclusive through September 2010.

For this WA HTA report, a subsequent search was conducted to identify published systematic reviews and individual studies (from June 2010 to Week 4, November 2011) using the same databases as Balk [AHRQ] (2011). The search strategies were parallel to those of Balk [AHRQ] (2011). An additional search using the Medicaid Evidence-based Decisions (MED) Project primary sources was completed to identify systematic reviews and technology assessments.

Balk [AHRQ] (2011) used a three-category grading system (A [good], B [fair/moderate], or C [poor]) to denote the methodological quality of each study. In addition, the strength of the body of evidence was graded for each analysis within each Key Question with modifications as described by Balk [AHRQ] (2011). The strength of the body of evidence was graded with one of the following descriptors: High, Moderate, Low, or Insufficient.

Key Question #8

At the direction of the Washington HTA program, Key Question #8 was added to address cost implications and cost effectiveness pertinent to OSA diagnosis and treatment. A search in the MEDLINE®, Cochrane Database of Systematic Reviews, and Cochrane Controlled Trials Registry sources was conducted to identify relevant articles published within the last 10 years (2001 to November 2011). An additional search was conducted using the MED Project primary sources to identify high quality systematic reviews and technology assessments.

The study selection criteria were modeled after that used by Balk [AHRQ] (2011). The methodological quality of the studies was assessed using a standard instrument developed and

adapted by the MED Project that are modifications of the BMJ (Drummond 1996), the Consensus on Health Economic Criteria list (Evers 2005), and the NICE economic evaluation checklist (NICE 2009). All studies were assessed by two independent and experienced raters. In cases where there was not agreement about the quality of the study, the disagreement was resolved by conference or the use of a third rater.

Guidelines

A search for relevant clinical practice guidelines was conducted using a list of predetermined high quality sources from the MED Project and additional relevant specialty organizations and associations. Guidelines included were limited to those published after 2005. The methodological quality of the guidelines was assessed using an instrument adapted from the Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration (AGREE Next Steps Consortium 2009). Each guideline was assigned a rating of good, fair, poor, based on the adherence to recommended methods and the potential for biases.

Policies

At the direction of the WA HTA program, select payer policies were searched and summarized. Aetna, Blue Cross Blue Shield, Group Health Cooperative, Federal (Medicare National and Local Coverage Decisions), and WA Medicaid policies were searched using the payers' websites.

Findings

For Key Questions #1 to #7 (Balk 2011), the search retrieved 15,816 citations, of which 249 articles, representing 234 studies, were included. There were 46 studies of diagnostic tests, 17 prediction studies, and 190 studies regarding treatment. As directed by the Washington HTA, a subsequent, updated search retrieved 92 citations. None of these subsequently published studies met inclusion criteria for Key Questions #1 to #7. For Key Question #8, the search retrieved 27 citations of which one systematic review and five individual studies met inclusion criteria.

The Balk [AHRQ] (2011) systematic review evaluated studies pertaining to the diagnosis and treatment of OSA in adults. Balk included surrogate or intermediate outcomes in addition to clinical outcomes. The Balk [AHRQ] (2011) systematic review (SR) was assigned a quality rating of good.

KQ#1. How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep?

> KQ#1a. How do the different tests compare in different subgroups of patients, based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke or airway characteristics?

Most experts consider laboratory-based PSG to be the reference standard for measuring Apnea-Hypopnea Index (AHI) in order to diagnose OSA. However, there are significant

challenges that can be raised in considering PSG to be the "gold standard". This would imply that this test is essentially error-free and therefore has the ability to prognosticate patients diagnosed with OSA from those without OSA. No current established threshold level for AHI exists that indicates the need for treatment (Tufts 2007). Furthermore, several facets raise uncertainty regarding PSG's place as the diagnostic "gold standard" (Balk [AHRQ] 2011):

- 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.

- Apnea-Hypopnea Index, 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.

- Apnea-Hypopnea Index as a predictor of clinical outcomes (Balk [AHRQ] 2011):

    - The strength of evidence is high that high baseline (AHI>30 events/hr or range) AHI is a strong and independent predictor of all-cause mortality over several years of follow-up (the association being strongest among people with severe OSA (AHI >30 events/hr).

        - Four studies found that AHI was a statistically significant independent predictor of death (follow up 2 to 14 years)

    - The association between baseline AHI and the following long-term clinical outcomes was analyzed by only one or two studies:

        - Cardiovascular (CV) disease

            - Studies reported mixed results regarding CV death, but AHI >30 was an independent predictor of nonfatal CV disease.

        - Stroke

            - A study suggested that the association between AHI and stroke may be confounded by obesity.

        - Hypertension

            - Studies had uncertain conclusions regarding the possible association between AHI and incident hypertension.

        - Non-insulin-dependent diabetes and other metabolic abnormalities

            - Studies reported mixed results that suggested an association between AHI and incident type 2 diabetes which, in one study was confounded by obesity.

        - Decreased quality of life

        - A single study found no significant association between AHI and future quality of life (SF-36 after 5 years).
    - Therefore, a link between reducing the AHI by OSA treatment and improved long-term clinical outcomes remains unproven.
- No current established threshold level for AHI exists that indicates the need for treatment.
- Given the uncertainty surrounding the clinical utility of the AHI, the measurement of this index is also subject to several sources of variability (Tufts 2007):
    - Airflow measurements are assessed by different instruments between laboratories and are subject to variation depending on the extent of mouth-breathing in the subject.
    - Oxygen saturation sampling is measured by different types of oximeters that may sample continuously or at various sampling rates.
    - Other probes are used that may be different between labs which measure respiratory movements and EEGs.
- Interpretation of the PSG results is another area of potential uncertainty:
    - Manual versus automated PSG scoring in the same lab may yield different results.
    - Intra- and inter-rater variability may not be completely negligible (Tufts 2007).
    - The definition of hypopnea varies, which results in different AHI measurements.
- Repeatability and reproducibility of PSG measurements are a concern:
    - Serial studies with the same patient in the same lab (repeatability) may result in differential classifications, especially in patients whose AHI scores are close to the OSA diagnostic cut-off point.
    - PSGs on the same patient in different labs (reproducibility) would be expected to have even more variation due to differing measurement apparatus.
        - In the population setting, this is of most importance as many patients will be seen across different sleep labs.

It should be clear from the points above that, while lab-based PSG indices provide the current reference standard, they alone are not a “gold standard” for diagnosing OSA. However, clinicians agree that from a pragmatic point of view, the PSG information is important in the

management of patients with disturbed sleep. Interestingly, no "strength of evidence" was assessed for this test, although it is the reference standard used throughout this report.

Diagnosing OSA, by detailing obstructive episodes, has been attempted using facility-based PSGs and various types of portable monitors, used in laboratory or home environments, which are categorized as follows:

- Type I: PSG in sleep facility
- Type II: Portable recording; same information as Type I (3 sleep arousal channels and minimum of 2 respiratory information channels)
- Type III: Portable recording; minimum of 2 respiratory channels (with no channels which differentiate waking and sleeping)
- Type IV: Portable monitors that fail Type III criteria

Several questionnaire designs and clinical prediction models have been used to assess sleep disordered breathing.

- Compared to the current diagnostic standard, the PSG, the strength of evidence that Type II, III, and IV monitors can accurately diagnosis OSA is **low to moderate** with wide bias in estimating the actual AHI.
- There is a **low** strength of evidence that the Berlin questionnaire is moderately accurate to screen for OSA. Other questionnaires could not be evaluated due to **insufficient** strength of evidence with the exception of one study suggesting the STOP-Bang instrument may accurately screen for OSA.
- There is a **low** strength of evidence supporting the usefulness of clinical prediction modeling in OSA diagnosis.

No subgroup analyses were performed due to insufficient evidence.

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

There was **insufficient** evidence for the utility of phased testing (i.e., using a screening test result to determine the next test to be performed in a series), as compared to PSG.

KQ#3. What is the effect of pre-operative screening for sleep apnea on surgical outcomes?

The utility of preoperative screening for OSA could not be determined due to **insufficient** strength of evidence.

KQ#4. In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index (AHI) or oxygen desaturation index (ODI) and other patient characteristics with long term clinical and functional outcomes?

Using AHI greater than 30 events per hour was found to be an independent predictor of all-cause mortality with a **high** strength of evidence. A higher AHI was also associated with incident diabetes based on a **low** strength of evidence. The association of diabetes and OSA may be confounded by obesity which may contribute to both conditions. There was **insufficient** evidence to determine an association of AHI with other clinical outcomes (e.g., cardiovascular mortality and hypertension).

KQ#5. What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults?

KQ#5a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors?

KQ#5b. Are any of these characteristics or factors predictive of treatment success?

- Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities.
- OSA severity or characteristics: Baseline questionnaire results, formal testing results (including hypoxemia levels), Baseline QoL; positional dependency, REM dependency
- Other: specific symptoms

A **moderate** strength of evidence was found for the effectiveness of treatment of OSA with CPAP. However, there was **insufficient** evidence to determine which patients CPAP might benefit the most.

- The reviewed studies report sufficient evidence supporting large improvements in sleep measures with CPAP compared with control (e.g., reducing AHI, improving ESS, reducing arousal index, and raising the minimum oxygen saturation).
- Weak evidence demonstrated no consistent benefit in improving quality of life, neurocognitive measures or other intermediate outcomes.
- Despite no or weak evidence for an effect of CPAP on clinical outcomes, given the large magnitude of effect on the intermediate outcomes of AHI and ESS, the strength of evidence that CPAP is an effective treatment to alleviate sleep apnea signs and symptoms was rated moderate.

However, the link between AHI reduction and long term clinical outcomes is not directly proven.

There was **insufficient** evidence regarding the comparison of various different CPAP devices, delivery methods, and regimens.

Mandibular advancement devices (MAD) had **moderate** strength of evidence supporting their use as an effective treatment for OSA. However, as with CPAP, there was **insufficient** evidence to indicate which patients might benefit from their use.

There was **moderate** evidence that the use of CPAP is superior to MAD with regard to improved sleep study measures, though no clinical outcomes were studied.

There was **insufficient** evidence to compare the different oral devices, other than MAD.

Surgical interventions for the treatment of OSA had **insufficient** evidence with which to evaluate their relative efficacy. When each modality was compared to CPAP, both surgical interventions and MAD had **insufficient** evidence to determine their relative merits. Of the other treatments for OSA that were considered, only intensive weight loss programs were an effective treatment in obese patients with OSA with a **low** strength of evidence. The remainder of the other management modalities (e.g., atrial overdrive pacing, medications, palatal implants, oropharyngeal exercises, tongue-retaining devices with positional alarms either in isolation or in combination, bariatric surgery, acupuncture, and auricular plaster) had **insufficient** evidence to determine the effects of using them for treatment of OSA.

KQ#6. In OSA patients prescribed non-surgical treatments, what are the associations of pre-treatment patient-level characteristics with treatment compliance?

Compliance in OSA patients, prescribed nonsurgical treatments: had **moderate** strength of evidence that compliance was greater with CPAP use with more severe OSA and **insufficient** evidence regarding potential predictors of MAD compliance.

KQ#7. What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?

The strength of evidence is **low** for indentifying any specific intervention which may improve CPAP compliance. No intervention type (e.g., education, telemonitoring) was more promising than others.

KQ#8. What is the evidence of cost implications and cost-effectiveness of sleep apnea diagnosis and treatment?

Cost Implications; Social, economic, and healthcare utilization consequences of OSA
In a study analyzing the social, economic, and healthcare utilization consequences of OSA, the direct and indirect costs for patients compared to controls were €5257 vs €1396 ($p<0.0001$). In another study, elderly and middle-aged patients with OSA consumed approximately two times as much in the way of healthcare resources as their paired controls.

Cost Effectiveness
In assessing sleep study alternatives to laboratory PSGs, the costs for additional QALYs incurred by full-night PSG and split-night PSG over home studies, and even by full-night PSG over split-

night PSG compared favorably with cost-utility estimates for a variety of widely accepted healthcare interventions.

Note: Consideration of this Key Question in regards to treating OSA should be undertaken with the caveat that the evidence about the long term improvements in clinical outcomes of these treatments is indirect. Therefore, the cost-effectiveness of an as yet unproven treatment that lacks high quality evidence of effectiveness is somewhat speculative.

Economic evaluations of various OSA treatment options, specifically comparing CPAP and Oral Appliances (OA), were presented using economic models. OA and CPAP are both highly cost-effective treatments for OSAH when compared to no treatment, with CPAP being the best option. These results corroborate the current recommendations on the use of CPAP as the primary treatment for moderate/severe OSAH, with OA the preferred treatment in patients unwilling or unable to use CPAP.

Guidelines

Guidelines addressed OSA diagnosis, pre-operative screening, and treatment. Three guidelines make recommendations for use of PSG in OSA diagnosis (AASM 2009; EFNS 2007; UTSN 2006). One guideline (AASM 2007a) recommends use of unattended portable monitors for OSA diagnosis, and one guideline recommends against autotitrated CPAP (APAP) for diagnostic use (AASM 2007b). Four guidelines recommend CPAP for treatment of OSA (AASM 2006b; AASM 2008; EFNS 2007; NICE 2008). A recommendation against soft palate implants was issued by one high-quality guideline (NICE 2007). Implants were recommended, in certain circumstances, by a fair-quality guideline (AASM 2010). Other surgical and non-surgical treatments were addressed by single guidelines. Recommendations for interventions such as multi-level, stepwise surgery and radiofrequency ablation were recommended by single guidelines. Maxilo-mandibular advancement was recommended by two guidelines (AASM 2009; AASM 2010). Tracheostomy was recommended, in certain situations, by three guidelines (AASM 2006b; AASM 2009; AASM 2010). Bariatric surgery as an adjunct to weight loss for obese patients was recommended by two guidelines (AASM 2006a; AASM 2009). Modafinil was recommended by one guideline (AASM 2006a) for treatment of residual daytime sleepiness despite successful CPAP treatment. This guideline recommended against various pharmaceuticals, such as serotonergic uptake inhibitors (SSRIs), protriptyline, methylxanthine derivatives and estrogen therapy for treatment of OSA. Pre-operative screening for OSA and CPAP initiation in certain cases were recommended by two guidelines (ASA 2006; ASPS 2009).

Policy Considerations

Federal, state, and private payer policies are fairly consistent in their coverage of the diagnosis and treatment of OSA. To diagnose OSA, Medicare and Aetna require a Type I PSG in a facility, or a Type II, III, or IV sleep test (with three or more channels) in a facility or at home. In contrast, Washington Medicaid requires a PSG at a Washington Health and Recovery Services Administration (HRSA)-approved sleep center for diagnosis. There were small variations across

payers in the noticeable symptoms that would lead to a positive diagnosis. Most payers cover CPAP as the first line treatment option, often followed by the use of a custom-fitted oral device, and varying forms of surgical treatment. The Medicare National Coverage Determination covers CPAP and specifies the sleep tests covered for diagnostic purposes. A number of Local Coverage Determinations, that include Washington State, indicate coverage for CPAP and custom-fitted oral devices. When those treatments fail and additional indications are present, three surgical options are covered. Coverage is inconsistent for lifestyle counseling as an initial part of treatment.

Overall Summary

Obstructive sleep apnea is a public health problem with a significant burden of morbidity and mortality. Accurately diagnosing and effectively treating OSA can improve symptoms of sleep disordered breathing and its consequences.

Diagnosing OSA, by detailing the obstructive episodes, has been done using facility-based PSG; portable monitors in a laboratory or home environments; several questionnaire designs; and, with clinical prediction modeling. Compared to the current diagnostic standard, the PSG, portable monitors, the Berlin questionnaire, and clinical modeling show low to moderate strength of evidence to support their use.

Using AHI greater than 30 events per hour was found to be an independent predictor of all-cause mortality, with a high strength of evidence. Otherwise AHI has not been correlated with clinical outcomes such as cardiovascular mortality and hypertension.

Moderate strength of evidence was found for the treatment of OSA with CPAP, though there was insufficient evidence to determine which patients CPAP might benefit the most, and for the various CPAP devices, delivery methods, and regimens regarding comparisons. There was moderate strength of evidence that CPAP compliance was greater in those patients with more severe OSA. The strength of evidence was low for indentifying any specific intervention which may improve CPAP compliance. No intervention type (e.g., education, telemonitoring), was more promising than any of the others.

Mandibular advancement devices (MAD) had moderate strength of evidence supporting their use as an effective treatment for OSA. There was moderate strength evidence, that the use of CPAP is superior to MAD, with regard to improved sleep study measures, though no clinical outcomes were studied.

Surgical interventions for the treatment of OSA had insufficient evidence with which to evaluate their relative efficacy. When each modality was compared to CPAP, both surgical interventions, and MAD had insufficient evidence to determine their relative merits.

The diagnosis and treatment of OSA have societal cost implications, making cost-effectiveness a concern in both of these aspects. Full-night PSGs are more cost-effective than split-night PSG, followed by home sleep studies. However, when an intermediate or high probability of moderate to severe OSAHS exists, home sleep studies are an equally effective, lower cost alternative to lab PSGs. Social, economic, and healthcare utilization consequences of OSA are significantly higher than in those subjects without OSA. Elderly and middle-aged patients with OSA consumed approximately two times as much in the way of healthcare resources as their paired controls. Various OSA treatment options, specifically comparing CPAP and OAs, were presented using economic models. OA and CPAP are both highly cost-effective treatments for OSAH when compared to no treatment, with CPAP being the best option.

Guidelines addressed OSA diagnosis, pre-operative screening, and treatment.

Federal, state and private payer policies are fairly consistent in their coverage of the diagnosis and treatment of OSA. Requirements for devices used to test for OSA are common across payers. Washington Medicaid requires a PSG at a HRSA-approved sleep center for diagnosis. Most payers cover CPAP as the first line treatment option, often followed by the use of a custom-fitted oral device, and varying forms of surgical treatment. Medicare National Coverage Decisions provide coverage for CPAP and specify which sleep tests are covered for diagnostic purposes. A number of Local Coverage Decisions that include Washington indicate coverage for CPAP, custom-fitted oral devices.

Limitations of the evidence

- Almost all studies were conducted in academic or research centers raising concerns as to the generalizability and applicability of the findings to the general population and nonacademic settings.
- There was a lack of trials which evaluated long-term clinical outcomes (i.e., as compared to sleep study parameter outcomes).
- Several Key Questions could not be addressed as very few trials reported subgroup analyses.
- Study follow-up durations tended to be short and dropout rates were often high.
- Many studies with incomplete reporting and inadequate analyses required the authors of this systematic review to estimate pertinent results.
- Publication bias was a concern as many trials of devices were funded by industry.
- The studies identified pertaining to economic evaluations were not comprehensive in covering all of the areas of interest to this report topic (i.e., not all diagnostic tools, interventions, or outcomes were studied).

- The economic studies focused on moderate to high OSA severity of disease as little evidence exists which assesses mild cases.

- Some studies did not report the discount rate used in cost analyses.

Background

Clinical Overview

Sleep apnea refers to sleep-disordered breathing and may be characterized as central, obstructive, or mixed in origin. Central sleep apnea may occur in neurologic conditions (e.g., Parkinson's and Alzheimer's disease), with brain stem damage from cerebrovascular accidents (e.g., stroke) or encephalitis, and with congestive heart failure. Mixed sleep apnea has both central and obstructive features (Balk [AHRQ] 2011). This report focuses solely on adult obstructive sleep apnea (OSA), which occurs when the upper airway becomes recurrently restricted during sleep. Here, the pharyngeal tissues episodically collapse, resulting in either reduced airflow (hypopnea), or a complete obstruction (apnea). These episodes lead to a drop in oxygen saturation, a disruption of gas exchange, disrupt Rapid Eye Movement (REM) sleep, and cause snoring, fitful sleep, and daytime somnolence (Punjabi 2008).

While prevalence estimates vary, OSA in adults appears to steadily increase with age. Occurring in 2 to 7% of the general adult population (Jennum & Riha 2009; Punjabi 2008), OSA prevalence increases to 10% at age 40, and to approximately 20% above 60 years of age (Balk [AHRQ] 2011). These rates are rising, likely associated with the increasing frequency of obesity. Risk factors for OSA include male gender, age, obesity, airway characteristics, familial/genetic predisposition, smoking, and alcohol consumption (Punjabi 2008). The majority of patients with OSA are asymptomatic, unaware of their sleep disordered breathing and the associated health risks. As a result, most patients do not seek medical care, and remain untreated (Balk [AHRQ] 2011).

Obstructive sleep apnea is a cause of significant morbidity and mortality, and is thus an important public health issue. It is an independent risk factor for hypertension may lead to neuropsychological impairment with deficits in attention, concentration, dexterity, memory, and has been shown to be a contributing factor in motor vehicle accidents (Jennum & Riha 2009). OSA has been associated with a variety of adverse clinical outcomes, including cardiovascular disease, stroke, transient ischemic attacks, and diabetes, in addition to the negative effects on daily quality of life (Balk [AHRQ] 2011; Jennum & Riha 2009). In addition, the diagnosis and treatment of OSA have societal cost implications, making cost-effectiveness a concern in both of these aspects.

Diagnosis

The diagnosis as well as the treatment of OSA is complicated by the difficulty in defining the syndrome. There is controversy surrounding the requisite ventilatory parameters to be used in

a clinical definition, as well as which diagnostic method is most appropriate with which to detect OSA. In the face of this debate, there are generally accepted standards. Obstructive sleep apnea is typically diagnosed, and its severity assessed, by counting the hourly obstructive breathing events (i.e., the apnea-hypopnea index (AHI), combined with other clinical features of obstruction (i.e., oxygen desaturation, choking, etc.) (Balk [AHRQ] 2011). The American Academy of Sleep Medicine defines OSA as, either greater than 15 obstructive events per hour of sleep, or five obstructive events per hour with the presence of daytime drowsiness, loud snoring, and nighttime awakenings due to respiratory occlusion (Epstein 2009).

Currently, the diagnostic standard for OSA is multi-channel polysomnography (PSG) in a sleep study facility attended by a technologist, and interpreted by a sleep physician after completion. Overnight measurements of rapid eye movement (REM), respiratory movement, oximetry, electroencephalography, electromyography, nasal and oral air flow are recorded. Considerable costs and patient inconvenience are involved in a PSG study. Portable monitors with fewer channels of measurement, either at home or in sleep facility have been developed. These portable tools are classified by The American Sleep Disorders Association into four categories, according to which channels of information they record and evaluate (Ferber 1994). The details of these monitor categories can be found in Appendix A. Briefly they are:

- Type I: PSG in sleep facility
- Type II: Portable recording; same information as Type I (3 sleep arousal channels and minimum of 2 respiratory information channels)
- Type III: Portable recording; minimum of 2 respiratory channels (with no channels which differentiate waking and sleeping)
- Type IV: Portable monitors that fail Type III criteria

Several types of questionnaires have been developed to aid in the diagnosis of OSA, by providing screening for candidates in an attempt to identify patients who should be referred for further testing. Predictive models to diagnose OSA have been created, using variables such as, anatomic and demographic features and symptoms. These tools also are used to determine which patients should be referred for a facility-based PSG (Balk [AHRQ] 2011).

Patients with OSA are at higher risk of postoperative complications, both in surgery for OSA, and for unrelated surgeries (e.g., joint replacement). These risks can be cardiac, respiratory, or anesthesia related. Since OSA is commonly undiagnosed, it has been postulated that screening all patients for OSA, may optimize perioperative management, and the tools for such screening have been studied (Balk [ARHQ] 2011).

Issues regarding the "Gold Standard" of Polysomnography (PSG)

Most experts consider laboratory-based PSG to be the reference standard for measuring Apnea-Hypopnea Index (AHI) in order to diagnose OSA. However, there are significant challenges that can be raised in considering PSG to be the "gold standard". This would imply

that this test is essentially error-free and therefore has the ability to prognosticate patients diagnosed with OSA from those without OSA. No current established threshold level for AHI exists that indicates the need for treatment (Tufts 2007). Furthermore, several facets raise uncertainty regarding PSG's place as the diagnostic "gold standard" (Balk [AHRQ] 2011):

- PSG is inconvenient, resource-intensive, and may not be representative of a typical night's sleep (particularly the first night the test is given).
- 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.
- Apnea-Hypopnea Index, 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.
- Apnea-Hypopnea Index as a predictor of clinical outcomes (Balk [AHRQ] 2011):
    - *The strength of evidence is high* that high baseline (AHI>30 events/hr or range) AHI is a strong and independent predictor of all-cause mortality over several years of follow-up.
    - The association being strongest among people with severe OSA (AHI >30 events/hr).
    - However, the strength of evidence is not reported for the association between baseline AHI and the following long-term clinical outcomes:
        - Cardiovascular disease
          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. A single quality A study evaluated nonfatal cardiovascular disease and similarly found that only AHI >30 events/hr was an independent predictor.
        - Stroke
          A single quality B study suggested that the association between AHI and stroke may be confounded by obesity.
        - Hypertension
          There were 2 studies (1 quality A, 1 quality B) that came to uncertain conclusions regarding the possible association between AHI and incident hypertension.
        - Non-insulin-dependent diabetes and other metabolic abnormalities

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.

- Decreased quality of life
  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.

- Thus the association between reductions in AHI by OSA treatment and improvements in long-term outcomes remains theoretical.

- No current established threshold level for AHI exists that indicates the need for treatment.

- Given the uncertainty surrounding the clinical utility of the AHI, the measurement of this index is also subject to several sources of variability (Tufts 2007):

  - Airflow measurements are assessed by different instruments between laboratories and are subject to variation depending on the extent of mouth-breathing in the subject.

  - Oxygen saturation sampling is measured by different types of oximeters that may sample continuously or at various sampling rates.

  - Other probes are used that may be different between labs which measure respiratory movements and EEGs.

- Interpretation of the PSG results is another area of potential uncertainty:

  - Manual versus automated PSG scoring in the same lab may yield different results.

  - Intra- and inter-rater variability may not be completely negligible (Tufts 2007).

  - The definition of hypopnea varies which results in different AHI measurements.

- Repeatability and reproducibility or PSG measurements are a concern:
    - Serial studies with the same patient in the same lab (repeatability) may result in differential classifications, especially in patients whose AHI scores are close to the OSA diagnostic cut-off point.
    - PSGs on the same patient in different labs (reproducibility) would be expected to have even more variation due to differing measurement apparatus.
        - In the population setting, this is of most importance as many patients will be seen across different sleep labs.

It should be clear from the above bullets that, while lab-based PSG indices provide the current reference standard, they alone are not a "gold standard" for diagnosing OSA. However, clinicians agree that from a pragmatic point of view, the PSG information is important in the management of patients with disturbed sleep.

Treatment

There have been various modalities developed to treat OSA. Many attempt to reduce the airway obstruction component. These include:

Continuous positive airway pressure (CPAP): This is the first-line therapy for OSA which delivers compressed air and opens the airway mechanically. It has been shown to be effective in improving sleep patterns and quality of life, by decreasing daytime somnolence (Balk [AHRQ] 2011). The nasal mask and machinery required is poorly tolerated by many patients and issues of compliance are of major concern (Veasey 2006).

Mandibular/Oral devices: As an alternative to CPAP, these devices, fitted by a dentist, attempt to advance the mandible or otherwise provide a splint to open the airway. The efficacy of these appliances, according to the American Academy for Sleep Medicine as used to treat mild to moderate OSA, should be tested in facility-based or home sleep study.

Surgery: When there is a clear anatomical airway obstruction or prior treatment has failed, surgery may be an appropriate treatment. Specific procedures depend on airway anatomy and the specific cause of obstruction. Uvulopalatopharyngoplasty (UPPP) is the most common type of surgery for OSA though many other surgical approaches have been used to open the airway. Maxillary-mandibular advancement and other oral devices are also commonly used (Balk [AHRQ] 2011). Bariatric surgery is also available, when appropriate, to treat the obesity contribution in OSA.

Lifestyle: Since obesity is the principle cause of OSA for many patients, a weight loss regimen can be an effective treatment (Balk [AHRQ] 2011). Other behavioral interventions include smoking cessation, avoiding caffeinated drinks and alcohol, positional therapy, and physical therapy of the oropharynx to strengthen the musculature and reduce obstruction (Balk [AHRQ] 2011).

Other treatment modalities may include arrhythmia treatment for nocturnal bradycardia, complementary, and alternative medicine. Pharmacologic agents have been used to either stimulate ventilation or as REM sleep suppressants (i.e., for patients whose respiratory episodes occur during REM sleep). Other agents such as opiates and nicotine have been studied (Balk [AHRQ] 2011).

Policy context and cost information

The rising prevalence of OSA is increasing the impetus of the payers' to critically evaluate the diagnostic and treatment options available in order to best inform policy. Federal, state and private payer policies are generally consistent in their coverage of the diagnosis and treatment of OSA. To diagnose OSA, Medicare and Aetna require a Type I PSG in a facility, or a Type II, III, or IV sleep test (with three or more channels) in a facility or at home. In contrast, Washington Medicaid requires a PSG at a Washington Health and Recovery Services Administration (HRSA)-approved sleep center for diagnosis. Diagnostic criteria were similar across payers, with small variations in the noticeable symptoms required in establishing the diagnosis. Most payers cover CPAP as the first line treatment option, often followed by the use of a custom-fitted oral device. Although all private payers cover some form of surgical treatment, Medicare National Coverage Determinations (NCDs) only provide coverage for CPAP. Lifestyle counseling is covered inconsistently as an initial component of treatment.

Undiagnosed OSA is associated with a roughly two-fold increase in health care utilization and costs, in the years preceding the diagnosis, due largely to the number of attendant comorbidies. Non-medical economic costs are also a consideration, and include the societal costs of motor vehicle accidents, and lost productivity (Wittmann & Rodenstein 2004). Conversely, the patient-incurred physician costs after successful OSA treatment was shown to decrease by 33%, along with significant reductions in hospital utilization (American Academy of Sleep Medicine 2000).

The 2011 Medicare Durable Medical Equipment, Prosthetics, Orthotics, and Supplies fee schedule for CPAP machines ranged from $85 to $101. This includes the machine without any of the services or potential repair costs (Centers for Medicare and Medicaid (CMS) 2011). The cost of polysomnography is approximately $1000 per study.

Washington State Data

State Agency Data on Sleep Apnea is presented below. Data presented includes all patient/member services performed in 2006-2010 with a Sleep Apnea Diagnosis.

Data Limitations:

- Sleep Apnea diagnoses are defined as specific codes (See Related Medical Services-Figure 9– Sleep Apnea diagnosis codes are bolded and categorized as "Study Diagnosis" in the Additional Info column)
- Complete data are presented for 2006-2009. 2010 data were added to Figure 1a and 1b only.
- Data includes adults and children, except figure 1b where Medicaid adult-only data are presented.

The PEB and Medicaid Fee for Service average annual populations are listed below:

Agency	2006	2007	2008	2009	2010
PEB	159,569	172,009	204,804	210,501	213,487
Medicaid	*	379,000	393,000	417,000	424,000

*Medicaid 2006 fee for service population average for 2006 was not available

Analysis notes:

PEB data contains claims where more than one insurer pays part of the claim. When PEB averages are presented, claims where PEB is not the primary payer are excluded from the calculation. PEB is not the primary payer for approximately 20% of the claims for sleep apnea in this analysis.

In the following figures, four or five year member or patient counts are not the total of members counted in each year, but are a separate count of unique members over whole period. Members may have services in more than one year, but will count only once in the overall column. This is noted on each figure where it applies.

Abbreviations:

L&I: Labor and Industries
PEB: Public Employees Benefits
SA: Sleep Apnea
CPAP: Continuous Positive Airway Pressure
SS: Sleep Study

Figure 1a: All Agency Sleep Apnea Summary, 2006-2010

PEB Sleep Apnea (SA) Summary	2006	2007	2008	2009	2010	5 Year Overall*
PEB Total Members with SA	**4,846**	**5,799**	**7,855**	**9,175**	**9987**	**17,739**
PEB Total Pd for SA	**$5,000,563**	**$6,184,073**	**$8,508,200**	**$10,051,005**	**$9,676,265**	**$39,420,106**
PEB Overall Avg paid per member	**$1,032**	**$1,035**	**$1,083**	**$1,095**	**$969**	**$2,222**
PEB Sole Payer Avg per member**	**$1,234**	**$1,264**	**$1,279**	**$1,284**	**$1,131**	**$2,375**
PEB Overall Max Paid per member	**$17,831**	**$34,877**	**$50,806**	**$32,703**	**$36,408**	**$52,265**
PEB CPAP	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
CPAP Total Paid	$2,223,475	$2,897,697	$3,571,833	$4,376,266	$4,476,446	$17,545,717
CPAP Mbrs	3564	4448	6034	7216	7763	12526
CPAP Avg/Mbr**	$192	$201	$193	$184	$202	$480
PEB Sleep Studies (SS)	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
SS Total Paid	$2,253,705	$2,651,858	$3,902,879	$4,548,764	$4,043,749	$17,400,955
SS Mbrs	1718	1917	2521	2771	2782	10783
SS Avg/Mbr**	$1,638	$1,709	$1,909	$2,060	$1,842	$1,997
Medicaid SA Summary	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
Medicaid Total Patients with SA	**2632**	**3367**	**3924**	**4492**	**3118**	**11391**
Medicaid Total Paid for SA	**$4,060,031**	**$4,785,950**	**$5,068,910**	**$5,633,478**	$4,788,230	**$24,336,599**
Medicaid Overall Average paid per patient	**$1,543**	**$1,421**	**$1,292**	**$1,254**	**$1,536**	**$2,136**
Medicaid Overall Max Paid per patient	**$132,658*****	**$128,172*****	**$18,134**	**$45,147**	**$87,540*****	**$141,359**
Medicaid CPAP	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
CPAP Total Paid	$1,816,055	$2,497,497	$2,924,207	$2,956,335	$1,939,546	$12,133,641
CPAP Patients	1430	2039	2523	2995	1572	6505
CPAP Avg/Patient	$1,270	$1,225	$1,159	$987	$1,234	$1,865
Medicaid Sleep Studies (SS)	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
SS Total Paid	$1,774,087	$1,884,934	$1,901,878	$2,138,232	$2,291,504	$9,990,635
SS Patients	2260	2374	2500	2656	2964	10879
SS Avg/Patient	$785	$794	$761	$805	$773	$918

Figure 1a continued next page

Figure 1a All Agency Sleep Apnea Summary, 2006-2010, continued

L&I Sleep Apnea (SA) Summary	2006	2007	2008	2009	2010	5 Year Overall*
L&I Total Pts with SA	**21**	**33**	**47**	**45**	**30**	**126**
L&I Total Pd for SA	**$38,369**	**$56,909**	**$70,362**	**$75,865**	**$76,965**	**$318,470**
L&I Overall Avg paid per claimant	**$1,827**	**$1,725**	**$1,497**	**$1,686**	**$1,816‡**	**$2,126‡**
L&I Overall Max Paid per claimant	**$4,876**	**$5,173**	**$4,855**	**$5,095**	**$4,581‡**	**$9,833‡**
L&I CPAP	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
CPAP Total Paid	$8,643	$18,883	$21,741	$24,900	$23,245	$97,412
CPAP Claimants	4	9	15	16	6	31
CPAP Avg/Claimant	$2,161	$2,098	$1,449	$1,556	$3,874	$3,142
L&I Sleep Studies (SS)	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
SS Total Paid	$25,572	$32,433	$41,407	$42,474	$36,333	$178.218
SS Claimants	18	20	31	30	26	119
SS Avg/Claimant	$1,421	$1,622	$1,336	$1,416	$1,397	$1,498

Notes for Figure 1a (preceding page):

*5 year overall figures consider all years' data together, counting members once, and giving a 5 year average and maximum paid.

**Averages for PEB were calculated using only primary payer data (data where PEB paid as though it was the only insurance carrier), excluding about 20% of claims where benefits were coordinated with another payer who paid the primary claim amount.

*** Unusually high maximum figures for Medicaid for 2006, 2007, and 2010 are for inpatient surgical procedures.

‡Excludes one outlier more than 3 standard deviations from the mean.

Note that the CPAP and Sleep Study subcategories for each agency are not a breakdown of the full Agency expenditure, but just the two main payment categories.

Note that Figure 1 includes the population under 18 (2.4% of PEB data, 14.6% of Medicaid data) See Figure 5a and 5b

Figure 1b: Medicaid Sleep Apnea Summary, Adults

Medicaid SA Summary, Adults only	2006	2007	2008	2009	2010	5 Year Overall*
Medicaid Total Adult SA Patients	**2255**	**2913**	**3422**	**3936**	**2666**	**9816**
Medicaid Total Paid for SA adults	**$3,466,396**	**$4,002,044**	**$4,421,247**	**$4,832,493**	**$3,973,264**	**$20,695,444**
Medicaid Adult Avg paid per Adult Patient	**$1,537**	**$1,374**	**$1,292**	**$1,228**	**$1,490**	**$2,108**
Medicaid Adult Max Paid per Adult	**$132,658*** **	**$15,954**	**$18,134**	**$36,652**	**$8,733**	**$141,359**
Medicaid CPAP, Adults only	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
CPAP Total Paid	$1,553,899	$2,097,107	$2,427,546	$2,523,042	$1,714,774	$10,316,369
CPAP Patients	1264	1821	2295	2748	1433	6041
CPAP Avg/Patient	$1,229	$1,152	$1,058	$918	$1,197	$1,708
Medicaid Sleep Studies (SS), Adults only	**2006**	**2007**	**2008**	**2009**	**2010**	**5 Year Overall***
SS Total Paid	$1,521,757	$1,608,185	$1,604,771	$1,813,711	$1,983,044	$6,548,424
SS Patients	1790	1874	1971	2164	2532	9355
SS Avg/Patient	$850	$858	$814	$838	$783	$700

*5 year overall figures consider all years' data together, counting members once, and giving a 5 year average and maximum paid.

*** Unusually high maximum in 2006 was one inpatient surgical procedure.

Note that the CPAP and Sleep Study subcategories are not a breakdown of the full Agency expenditure, but just the two main payment categories.

Note that due to the higher proportion of children as a component of the complete Medicaid data presented in other tables (14.6%), Figure 1b excludes patients 18 and under.

Figure 2a: PEB Sleep Apnea Diagnosis Overview, 2006-2009

Diagnosis Code and Description		Member Counts				4 Yrs Overall *	Payments				4 Years Overall
		2006	2007	2008	2009		2006	2007	2008	2009	
327.2	ORGANIC SLEEP APNEA NOS	23	57	69	96	160	$5,088	$24,875	$31,847	$39,344	$101,154
327.21	PRIM CENTRAL SLEEP APNEA	6	30	70	110	95	$1,762	$5,711	$22,519	$45,996	$75,988
327.23	OBSTRUCTIVE SLEEP APNEA	2365	3545	6451	9268	10827	$1,814,524	$3,141,676	$5,248,849	$7,675,533	$17,880,582
327.27	CNTRL SLEEP APNEA OTH DIS	3	10	22	25	34	$949	$9,530	$6,006	$11,751	$28,236
327.29	ORGANIC SLEEP APNEA NEC	3	5	6	10	12	$393	$1,942	$1,749	$2,220	$6,304
780.51	INSOMN W SLEEP APNEA NOS	131	108	142	164	451	$24,652	$25,303	$50,777	$59,338	$160,070
780.53	HYPERSOM W SLP APNEA	4940	4711	5375	3951	8333	$2,381,824	$2,162,136	$2,187,397	$1,373,625	$8,104,982
780.57	SLEEP APNEA NOS	1545	1698	1944	1538	4181	$771,371	$812,900	$959,056	$843,198	$3,386,525
Grand Total		**9016**	**10164**	**14079**	**15162**	**14613***	**$5,000,563**	**$6,184,073**	**$8,508,200**	**$10,051,005**	**$29,743,841**

*Member Counts in the 4 Years Overall column do not repeat members who were included in more than one annual count.

Figure 2b: Medicaid Sleep Apnea Diagnosis Overview, 2006-2009

Diagnosis Code and Description		Patient Counts				4 Yrs Over all*	Payments				4 Years Overall
		2006	2007	2008	2009		2006	2007	2008	2009	
327.2	ORGANIC SLEEP APNEA NOS	24	21	24	15	80	$7,536	$5,142	$8,336	$7,467	$28,481
327.21	PRIM CENTRAL SLEEP APNEA	0	0	0	0	0	$0	$0	$0	$0	$0
327.23	OBSTRUCTIVE SLEEP APNEA	1457	1975	2861	3875	7060	$1,703,555	$2,362,053	$2,903,881	$4,311,762	$11,281,251
327.27	CNTRL SLEEP APNEA OTH DIS	8	12	27	21	54	$8,676	$15,736	$33,953	$26,060	$84,425
327.29	ORGANIC SLEEP APNEA NEC	5	25	2	4	13	$1,411	$70	$45	$310	$1,836
780.51	INSOMN W SLEEP APNEA NOS	87	73	94	101	331	$22,362	$18,484	$31,527	$27,110	$99,482
780.53	HYPERSOM W SLP APNEA	2004	2064	1944	1465	5071	$1,785,397	$1,704,812	$1,479,683	$784,991	$5,754,884
780.57	SLEEP APNEA NOS	833	1001	1066	942	3055	$531,094	$679,653	$611,486	$475,778	$2,298,011
Grand Total		**4418**	**5148**	**6018**	**6423**	**8734***	**$4,060,031**	**$4,785,950**	**$5,068,910**	**$5,633,478**	**$19,548,369**

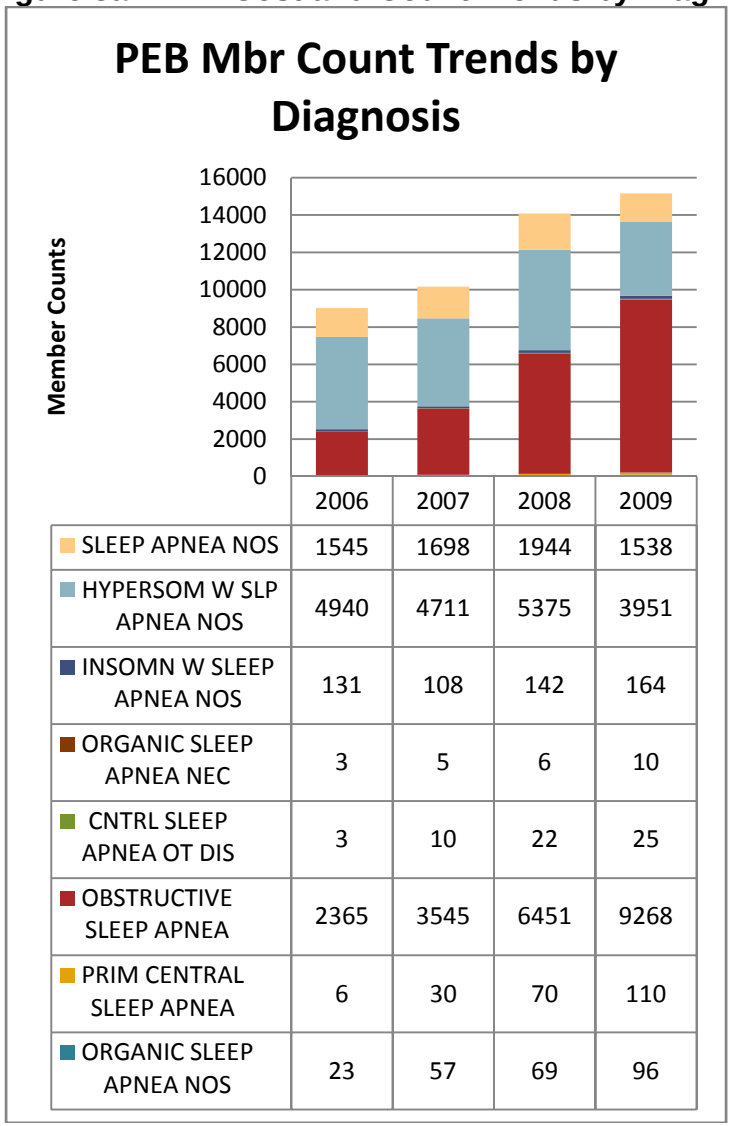
*Patient Counts in the 4 Years Overall column do not repeat members who were included in more than one annual count.

Figure 2c: L&I Sleep Apnea Diagnosis Overview, 2006-2009

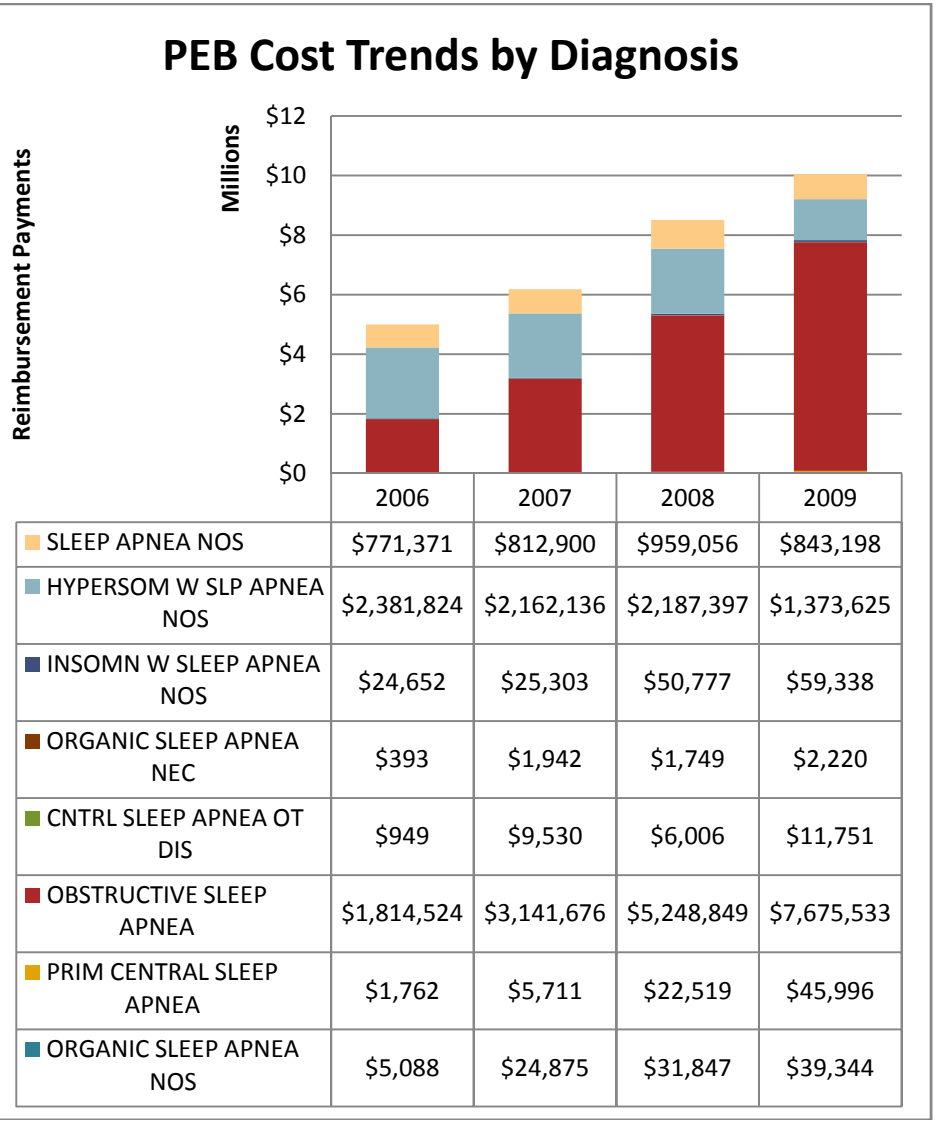
Diagnosis Code and Description		Claimant Counts				4 Years Overall*	Payments				4 Years Overall
		2006	2007	2008	2009		2006	2007	2008	2009	
327.21	PRIM CENTRAL SLEEP APNEA		1			**1**		$909			**$909**
327.23	OBSTRUCTIVE SLEEP APNEA	17	18	32	35	**77**	$21,079	$28,229	$46,372	$63,040	**$158,720**
327.27	CNTRL SLEEP APNEA OTH DIS	1				**1**	$76				**$76**
780.51	INSOMN W SLEEP APNEA NOS	1	1		1	**3**	$242	$449		$156	**$847**
780.53	HYPERSOM W SLP APNEA	11	16	19	11	**39**	$14,248	$20,823	$13,770	$9,657	**$58,497**
780.57	SLEEP APNEA NOS	4	7	9	7	**23**	$2,724	$6,499	$10,221	$3,013	**$22,457**
Grand Total		**34**	**43**	**60**	**54**	**96***	**$38,369**	**$56,909**	**$70,362**	**$75,865**	**$241,505**

*Claimant Counts in the 4 Years Overall column do not repeat members who were included in more than one annual count.

Figure 3a: PEB Cost and Count Trends by Diagnosis 2006-2009



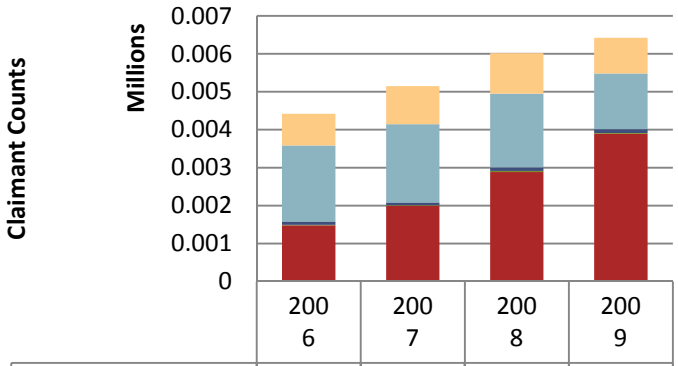
	2006	2007	2008	2009
SLEEP APNEA NOS	1545	1698	1944	1538
HYPERSOM W SLP APNEA NOS	4940	4711	5375	3951
INSOMN W SLEEP APNEA NOS	131	108	142	164
ORGANIC SLEEP APNEA NEC	3	5	6	10
CNTRL SLEEP APNEA OT DIS	3	10	22	25
OBSTRUCTIVE SLEEP APNEA	2365	3545	6451	9268
PRIM CENTRAL SLEEP APNEA	6	30	70	110
ORGANIC SLEEP APNEA NOS	23	57	69	96



	2006	2007	2008	2009
SLEEP APNEA NOS	$771,371	$812,900	$959,056	$843,198
HYPERSOM W SLP APNEA NOS	$2,381,824	$2,162,136	$2,187,397	$1,373,625
INSOMN W SLEEP APNEA NOS	$24,652	$25,303	$50,777	$59,338
ORGANIC SLEEP APNEA NEC	$393	$1,942	$1,749	$2,220
CNTRL SLEEP APNEA OT DIS	$949	$9,530	$6,006	$11,751
OBSTRUCTIVE SLEEP APNEA	$1,814,524	$3,141,676	$5,248,849	$7,675,533
PRIM CENTRAL SLEEP APNEA	$1,762	$5,711	$22,519	$45,996
ORGANIC SLEEP APNEA NOS	$5,088	$24,875	$31,847	$39,344

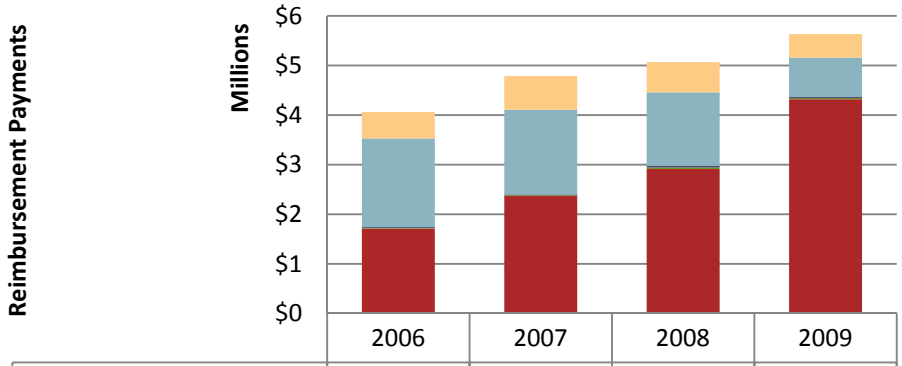
Figure 3b: Medicaid Cost and Count Trends by Diagnosis 2006-2009

Medicaid Sleep Apnea Count Trends by Diagnosis



	2006	2007	2008	2009
SLEEP APNEA NOS	833	1001	1066	942
HYPERSOM W SLP APNEA NOS	2004	2064	1944	1465
INSOMN W SLEEP APNEA NOS	87	73	94	101
ORGANIC SLEEP APNEA NEC	5	2	2	4
CNTRL SLEEP APNEA OT DIS	8	12	27	21
OBSTRUCTIVE SLEEP APNEA	1457	1975	2861	3875
PRIM CENTRAL SLEEP APNEA	0	0	0	0
ORGANIC SLEEP APNEA NOS	24	21	24	15

Medicaid Sleep Apnea Cost Trends by Diagnosis



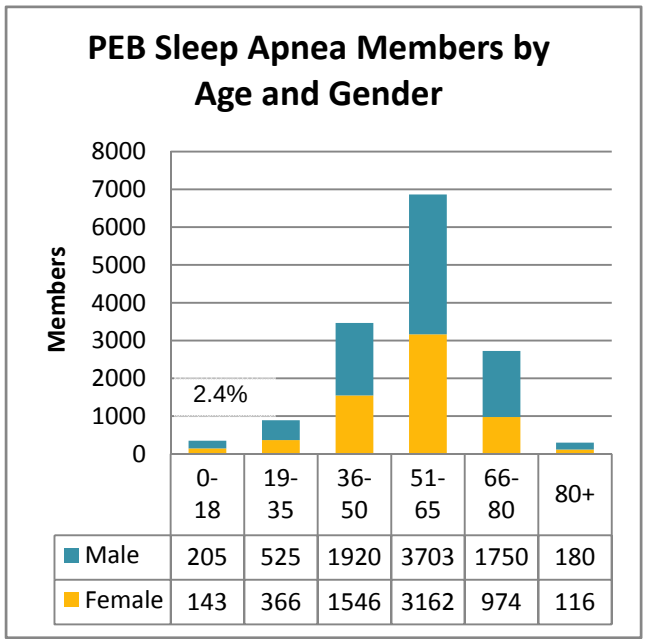
	2006	2007	2008	2009
SLEEP APNEA NOS	$531,094	$679,653	$611,486	$475,778
HYPERSOM W SLP APNEA NOS	$1,785,397	$1,704,812	$1,479,683	$784,991
INSOMN W SLEEP APNEA NOS	$22,362	$18,484	$31,527	$27,110
ORGANIC SLEEP APNEA NEC	$1,411	$70	$45	$310
CNTRL SLEEP APNEA OT DIS	$8,676	$15,736	$33,953	$26,060
OBSTRUCTIVE SLEEP APNEA	$1,703,555	$2,362,053	$2,903,881	$4,311,762
PRIM CENTRAL SLEEP APNEA	$0	$0	$0	$0
ORGANIC SLEEP APNEA NOS	$7,536	$5,142	$8,336	$7,467

Note: An equivalent table to 3a and 3b is not presented for L&I. Though L&I SA payments are rising year to year, patient counts do not show trends that are evident in a bar chart presentation. Data for L&I is shown in Figure 2c.**Figure 6. Sleep Apnea Member Counts by Year Compared to Expected Member Counts by Population Growth**

Figure 4: All Agency 4 year Payment Category Totals

Cost Category	Four year total Paid			Percent of 4 year total		
	PEB	Medicaid	L&I	PEB	Medicaid	L&I
Total Paid	$29,743,841	$19,548,369	$241,505			
Sleep Studies	$13,357,206	$7,699,131	$141,885	44.9%	39.4%	58.8%
Office Visits	$2,401,010	$1,110,995	$25,453	8.1%	5.7%	10.5%
Treatment	$13,985,625	$10,738,243	$74,167	47.0%	54.9%	30.7%
Treatment Subcategories						
CPAP	$13,069,271	$10,194,095	$74,167	43.9%	52.1%	30.7%
Surgery	$381,434	$534,390	$0	1.3%	2.7%	0.0%
Orthotics	$9,979	$0	$0	0.0%	0.0%	0.0%

Figure 5a, 5b: PEB/Medicaid SA Diagnosed Member Age and Gender



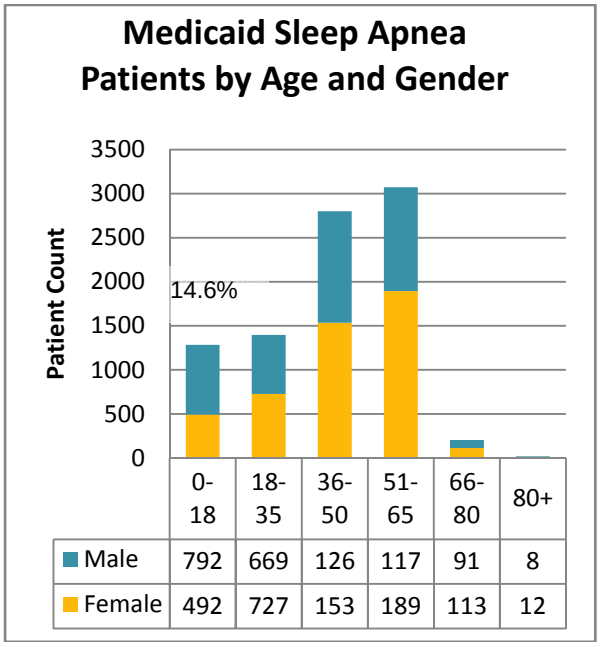


Figure 5c: L&I Sleep Apnea Diagnosed Claimant Age and Gender

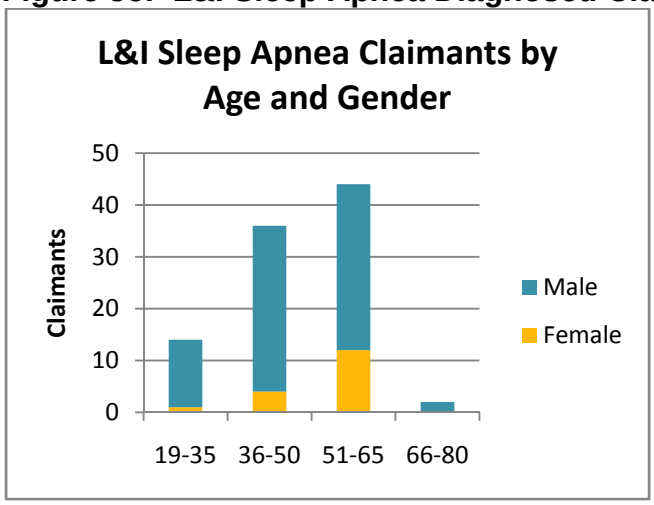


Figure 6: All Agency Sleep Study Repeats, 2006-2009

Sleep Study Count by Date	PEB Member Count	Medicaid Member Count	L&I Member Count
1	4944	5148	59
2	2989	2616	34
3	316	388	0
More than 3	105	173	0

Figure 7a. PEB Duration of Treatment for 2006 Sleep Study Patients

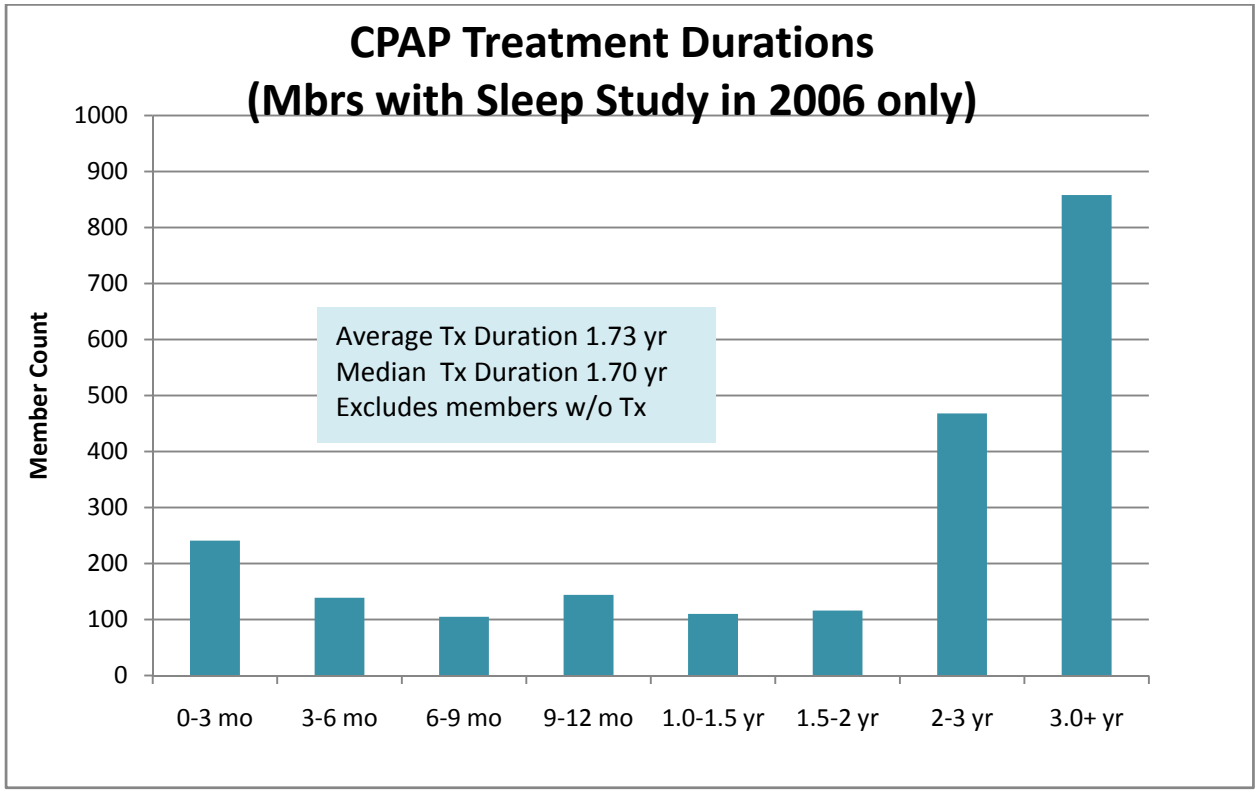
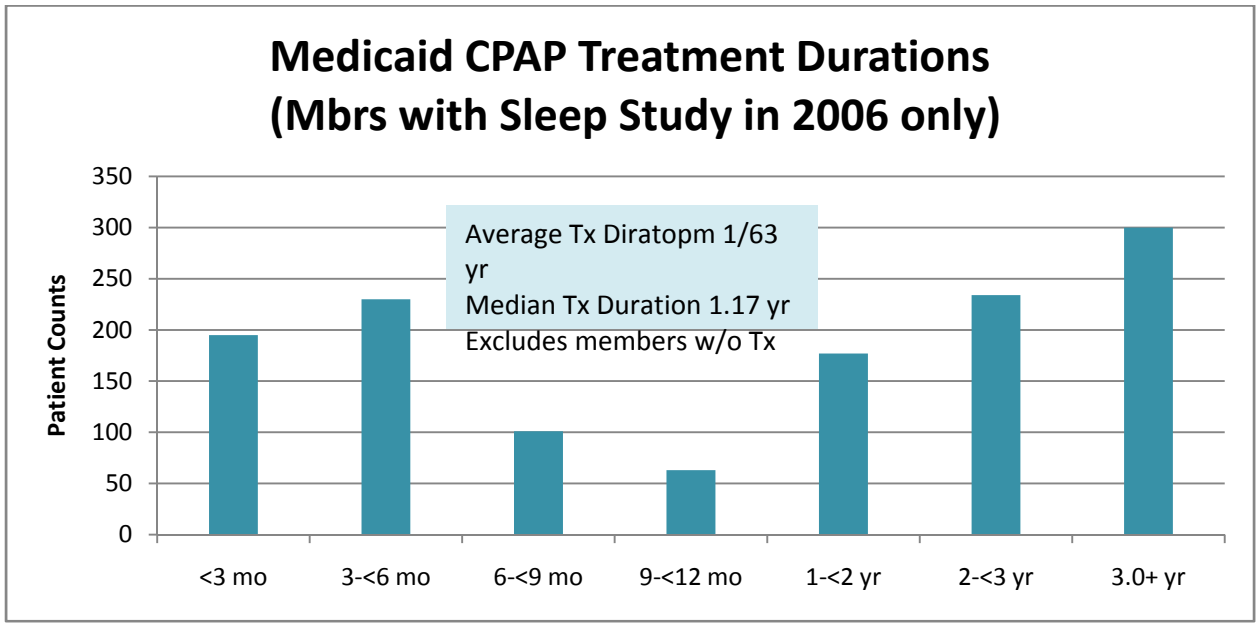
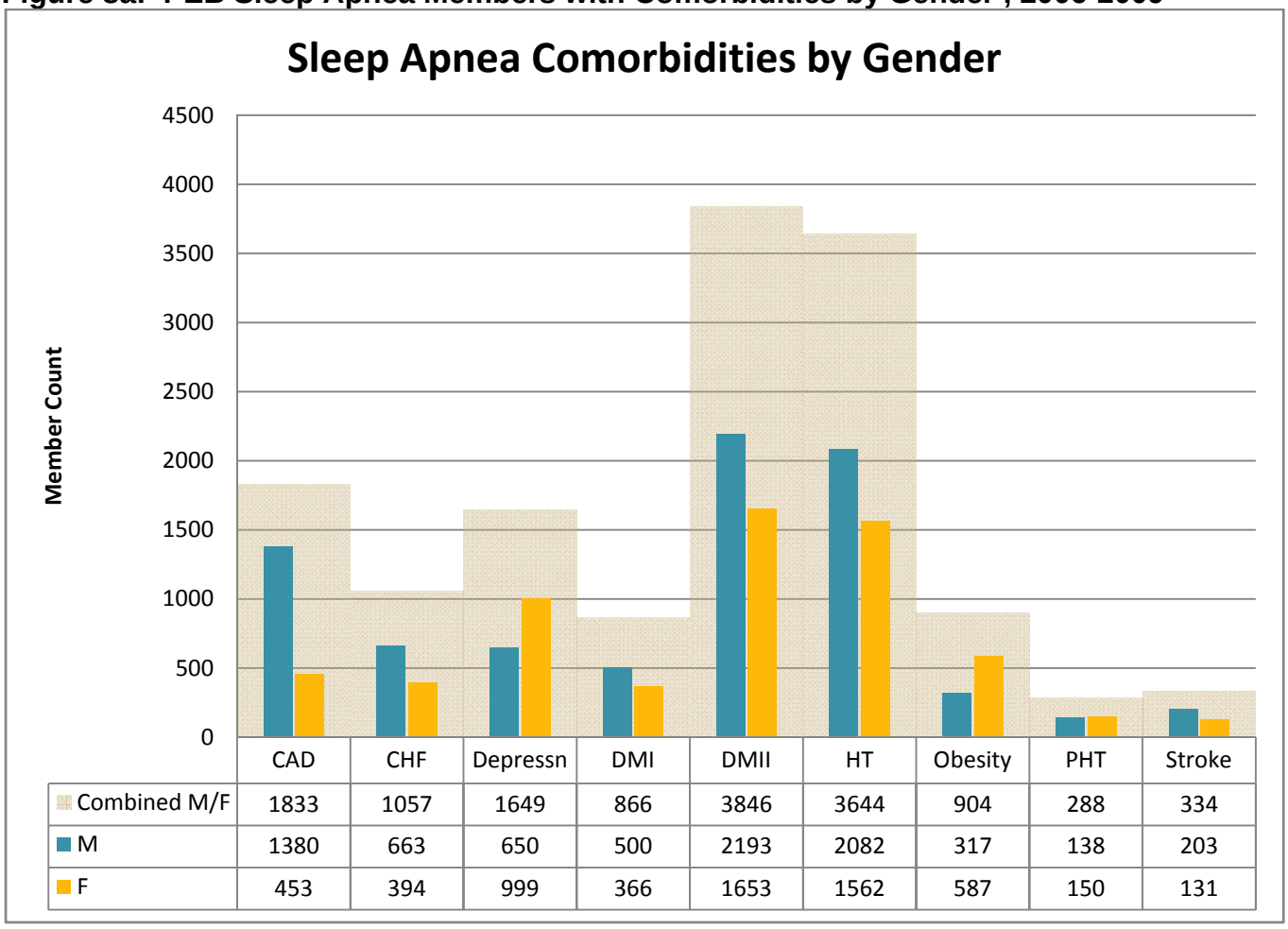


Figure 7b. Medicaid Duration of Treatment for 2006 Sleep Study Patients



Note: CPAP Treatment durations are not displayed for L&I .

Figure 8a: PEB Sleep Apnea Members with Comorbidities by Gender , 2006-2009

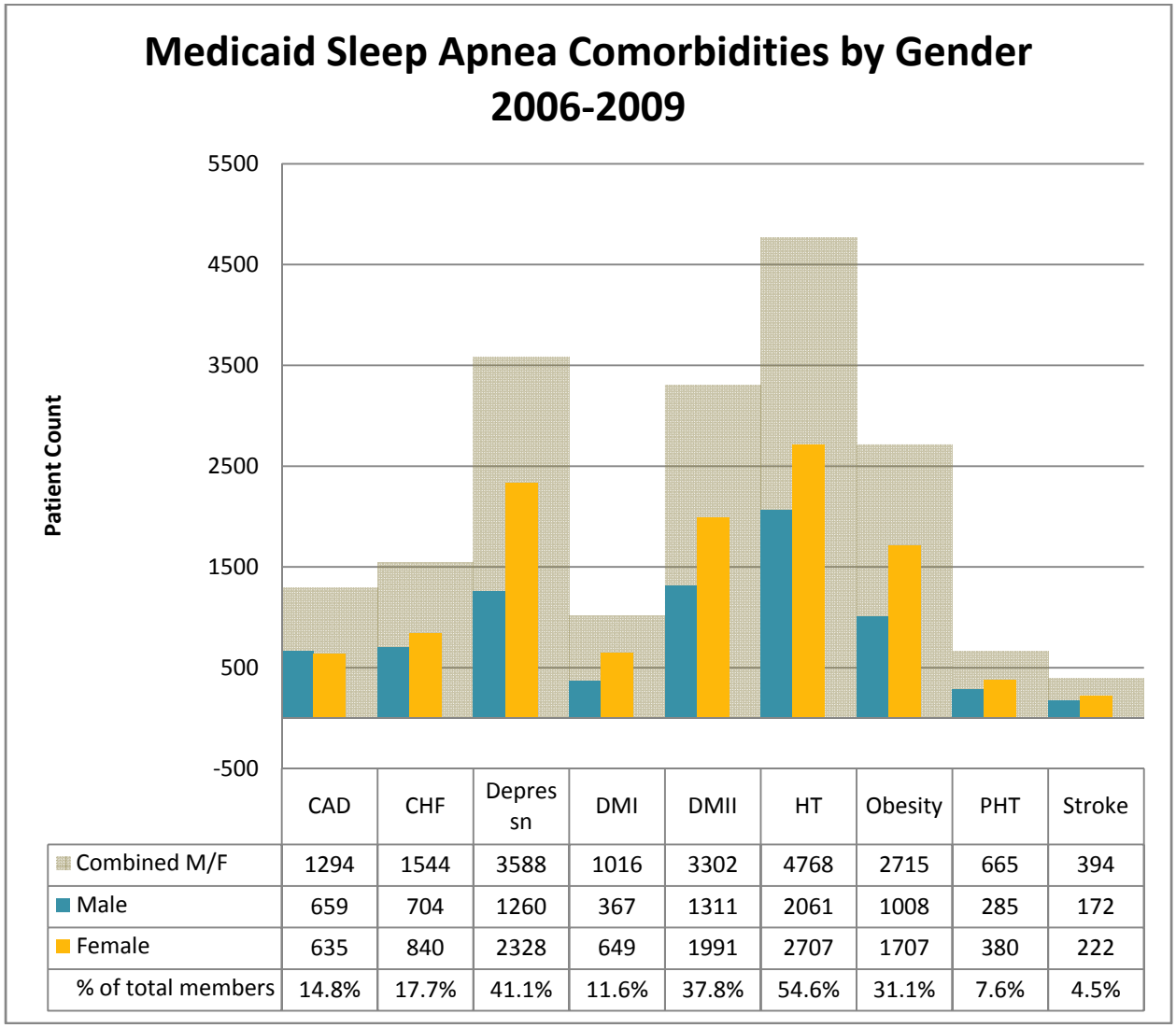


	CAD	CHF	Depressn	DMI	DMII	HT	Obesity	PHT	Stroke
Combined M/F	1833	1057	1649	866	3846	3644	904	288	334
M	1380	663	650	500	2193	2082	317	138	203
F	453	394	999	366	1653	1562	587	150	131

Starting with members who were reported with a Sleep Apnea diagnosis during 2006-2009, we counted members with any of the above conditions during the time period

CAD = Coronary Artery Disease
CHF = Congestive Heart Failure
DMI/DMII = Diabetes Mellitus Type 1 and 2
HT = Hypertension
PHT = Pulmonary Hypertension

Figure 8b: Medicaid Sleep Apnea Members with Comorbidities by Gender , 2006-2009



	CAD	CHF	Depressn	DMI	DMII	HT	Obesity	PHT	Stroke
Combined M/F	1294	1544	3588	1016	3302	4768	2715	665	394
Male	659	704	1260	367	1311	2061	1008	285	172
Female	635	840	2328	649	1991	2707	1707	380	222
% of total members	14.8%	17.7%	41.1%	11.6%	37.8%	54.6%	31.1%	7.6%	4.5%

Figure 9. Related Medical Codes

Related Medical Codes		
Codes	**Short Description**	**Additional Info**
ICD9 Diagnosis		
327.20	Organic sleep apnea, unspecified	Study Diagnosis
327.21	Primary central sleep apnea	Study Diagnosis
327.23	Obstructive sleep apnea (adult) (pediatric)	Study Diagnosis
327.27	Central sleep apnea in conditions classified elsewhere	Study Diagnosis
327.29	Other organic sleep apnea	Study Diagnosis
478.29	Nasopharyngeal obstruction	
770.81	Primary apnea of newborn	Excluded-outside scope
770.82	Other apnea of newborn	Excluded-outside scope
780.5	Sleep disturbance, unspecified	
780.51	Insomnia with sleep apnea, unspecified	Study Diagnosis
780.53	Hypersomnia with sleep apnea, unspecified	Study Diagnosis
780.54	Hypersomnia, unspecified	
780.57	Unspecified sleep apnea	Study Diagnosis
786.00	Unspecified respiratory abnormality	
786.01	Hyperventilation	
786.02	Orthopnea	
786.03	Apnea	Excludes sleep apnea (Code Manager 2010)
786.04	Cheyne-Stokes respiration	
786.05	Shortness of breath	
786.06	Tachypnea	
786.07	Wheezing	
786.09	Other dyspnea and respiratory abnormalities	
786.1	Stridor	
786.2	Cough	
786.30	Hemoptysis, unspecified	
786.31	Acute idiopathic pulmonary hemorrhage in infants [AIPHI]	Outside scope
786.39	Other hemoptysis	
786.4	Abnormal sputum	
786.50	Chest pain, unspecified	

Related Medical Codes		
Codes	**Short Description**	**Additional Info**
786.59	Chest pain, other	
786.6	Swelling, mass, or lump in chest	
786.7	Abnormal chest sounds	
786.8	Hiccough	
786.9	Other symptoms involving respiratory system and chest	
Comorbidities		
401.9	Hypertension	
416.x	Pulmonary hypertension	Comorbid Condition
428.x	Heart failure	Comorbid Condition
434.x	Stroke	Comorbid Condition
414.x	Coronary Artery Disease	Comorbid Condition
311.x	Depression	Comorbid Condition
250.x	Diabetes	Comorbid Condition
278.00	Obesity	Comorbid Condition
Treatments (CPT)		
21685	Hyoid myotomy and suspension	Include in treatment selection
41512	Tongue base suspension, permanent suture technique	Include in treatment selection
41530	Submucosal ablation of the tongue base, radiofrequency, 1 or more sites, per session	Include in treatment selection
ICD9 Procedures		
93.9	CPAP	Include in treatment selection
27.69,	Uvulopalatopharyngoplasty (UPPP)	Include in treatment selection
28.2	Tonsillectomy	Include in treatment selection
28.6	Adenoidectomy	Include in treatment selection
28.3	Tonsillectomy/adenoidectomy	Include in treatment selection
31.29	Tracheostomy	Include in treatment selection
21.31	Nasal surgery (remove polyps)	Include in treatment selection
21.88	Nasal surgery (repair deviated septum)	Include in treatment selection
CPAP Equip (HCPCS)		
A7030	Full face mask used with positive airway pressure device, each	Include in treatment selection
A7031	Face mask interface, replacement for full face mask, each	Include in treatment selection
A7032	Cushion for use on nasal mask interface, replacement only, each	Include in treatment selection

Related Medical Codes		
Codes	**Short Description**	**Additional Info**
A7033	Pillow for use on nasal cannula type interface, replacement only, pair	Include in treatment selection
A7034	Nasal interface (mask or cannula type) used with positive airway pressure device, with or without head strap	Include in treatment selection
A7036	Chinstrap used with positive airway pressure device	Include in treatment selection
A7037	Tubing used with positive airway pressure device	Include in treatment selection
A7038	Filter, disposable, used with positive airway pressure device	Include in treatment selection
A7039	Filter, nondisposable, used with positive airway pressure device	Include in treatment selection
A7035	Headgear used with positive airway pressure device	Include in treatment selection
A7524	Tracheostoma stent/stud/button, each	Include in treatment selection
G0398	Home sleep study test (HST) with type II portable monitor, unattended; minimum of 7 channels: EEG, EOG, EMG, ECG/heart rate, airflow, respiratory effort and oxygen saturation	Include in treatment selection
G0399	Home sleep test (HST) with type III portable monitor, unattended; minimum of 4 channels: 2 respiratory movement/airflow, 1 ECG/heart rate and 1 oxygen saturation	Include in treatment selection
G0400	Home sleep test (HST) with type IV portable monitor, unattended; minimum of 3 channels	Include in treatment selection
E0485	Oral device/appliance used to reduce upper airway collapsibility, adjustable or nonadjustable, prefabricated, includes fitting and adjustment	Include in treatment selection
E0486	Oral device/appliance used to reduce upper airway collapsibility, adjustable or nonadjustable, custom fabricated, includes fitting and adjustment	Include in treatment selection
E0470	Respiratory assist device, bi-level pressure capability, without backup rate feature, used with noninvasive interface, e.g., nasal or facial mask (intermittent assist device with continuous positive airway pressure device)	Include in treatment selection
E0471	Respiratory assist device, bi-level pressure capability, with back-up rate feature, used with noninvasive interface, e.g., nasal or facial mask (intermittent assist device with continuous positive airway pressure device)	Include in treatment selection
E0472	Respiratory assist device, bi-level pressure capability, with backup rate feature, used with invasive interface, e.g., tracheostomy tube (intermittent assist device with continuous positive airway pressure device)	Include in treatment selection
E0601	Continuous airway pressure (CPAP) device	Include in treatment selection
A4604	Tubing with integrated heating element for use with positive airway pressure device	Include in treatment selection

Related Medical Codes		
Codes	**Short Description**	**Additional Info**
Sleep Studies		
95806	Sleep study, unattended, simultaneous recording of, heart rate, oxygen saturation, respiratory airflow, and respiratory effort (e.g., thoracoabdominal movement)	Include in sleep study selection
95807	Sleep study, simultaneous recording of ventilation, respiratory effort, ECG or heart rate, and oxygen saturation, attended by a technologist	Include in sleep study selection
95800	Sleep study, unattended, simultaneous recording; heart rate, oxygen saturation, respiratory analysis (e.g., by airflow or peripheral arterial tone), and sleep time	Include in sleep study selection
95801	Sleep study, unattended, simultaneous recording; minimum of heart rate, oxygen saturation, and respiratory analysis (e.g., by airflow or peripheral arterial tone)	Include in sleep study selection
95805	Multiple sleep latency or maintenance of wakefulness testing, recording, analysis and interpretation of physiological measurements of sleep during multiple trials to assess sleepiness	Include in sleep study selection
95808	Polysomnography; sleep staging with 1-3 additional parameters of sleep, attended by a technologist	Include in sleep study selection
95810	Polysomnography; sleep staging with 4 or more additional parameters of sleep, attended by a technologist	Include in sleep study selection
95811	Polysomnography; sleep staging with 4 or more additional parameters of sleep, with initiation of continuous positive airway pressure therapy or bilevel ventilation, attended by a technologist	Include in sleep study selection
95803	Actigraphy testing, recording, analysis, interpretation, and report (minimum of 72 hours to 14 consecutive days of recording)	Include in sleep study selection
95822	Electroencephalogram (EEG); recording in coma or sleep only	Include in sleep study
95827	Electroencephalogram (EEG); all night recording	Include in sleep study
95819	Electroencephalogram (EEG); including recording awake and asleep	Include in sleep study selection
95806	Sleep study, unattended, simultaneous recording of, heart rate, oxygen saturation, respiratory airflow, and respiratory effort (e.g., thoracoabdominal movement)	Include in sleep study selection

Methods

Key Questions #1 to #7

Search strategy

At the direction of the Washington HTA program, the recent Agency for Healthcare Research and Quality (AHRQ) systematic review, *Comparative Effectiveness of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults*, was identified as the primary evidence source for Key Questions #1 to #7 (Balk [AHRQ] 2011). The Balk [AHRQ] (2011) review searched MEDLINE®, Cochrane Central Trials Registry®, and Cochrane Systematic Reviews® for literature published in English. The literature search dates were inclusive through September 2010.

For this WA HTA report, a subsequent search was conducted to identify published systematic reviews and individual studies (from June 2010 to Week 4, November 2011) in MEDLINE®, Cochrane Database of Systematic Reviews (through November 2011), and Cochrane Central Controlled Trials Register (4th Quarter). The search strategies were parallel to those of Balk [AHRQ] (2011) and are provided in Appendix B. A list of excluded studies with reasons for exclusion is provided in Appendix C. An additional search using the Medicaid Evidence-based Decisions (MED) project primary sources was completed to identify systematic reviews and technology assessments. The primary sources searched included: Cochrane Library (Wiley Interscience), UK National Institute for Health and Clinical Excellence (NICE), Blue Cross/Blue Shield Health Technology Assessment (HTA) program, Veterans Administration TA program, BMJ Clinical Evidence, the Canadian Agency for Drugs and Technologies in Health (CADTH), and the Agency for Health Research and Quality (AHRQ).

Study Selection

The study selection criteria used by Balk [AHRQ] (2011) is excerpted from the Balk [AHRQ] review and provided below. The same criteria were used for the subsequently published studies included in this WA HTA report.

> **Population and Condition of Interest** [Balk [AHRQ] (2011), p. 12-13]
>
> We included studies conducted only in adults (>16 years). By consensus with the Technical Expert Panel, we excluded studies in which more than 20 percent of the participants had neuromuscular disease, Down syndrome, Prader-Willi syndrome, major congenital skeletal abnormalities, narcolepsy, narcotic addiction, Alzheimer's disease, epilepsy, or who had experienced a disabling stroke. This threshold (20 percent) was chosen arbitrarily to avoid excluding potentially relevant small studies that included some patients with conditions not of interest to the current report. This turned out to be a moot point since no eligible studies explicitly included patients with any of these conditions.
>
> **Diagnostic testing (Key Questions #1 & #2)**. We included studies of adults with symptoms, findings, history, and comorbidities that indicated an increased risk of sleep apnea. Studies conducted in only symptomatic or healthy general-population participants, as well as those in patients with known sleep apnea, were excluded.
>
> **Preoperative screening (Key Question #3).** We included studies of all preoperative patients, irrespective of the surgery to be performed, as long as they were scheduled to receive general anesthesia. We

excluded studies in which all patients were known to have sleep apnea. There were no other restrictions based on patient symptoms or existing diagnoses.

Predictors of long-term outcomes (Key Question #4). We included studies of adults, regardless of health status, who had a baseline sleep study performed for any reason.

Treatment of OSA (Key Question #5) and treatment compliance (Key Questions #6 & #7). We included studies of adults with a confirmed diagnosis of OSA, whether associated with symptoms or not, and with formal sleep study testing demonstrating an apnea-hypopnea index (AHI) ≥5 events/hr. We excluded studies with 20 percent of study subjects without OSA, unless a subgroup analysis of OSA patients was reported. This restriction included patients with central sleep apnea or snoring without OSA.

Interventions, Predictors, and Comparators of Interest [Balk [AHRQ] (2011), p. 13-15]

Diagnostic testing (Key Question #1). We evaluated two types of comparisons: portable monitoring devices (used at home or setting other than a sleep laboratory) versus facility-based polysomnography (PSG); and questionnaires or prediction models versus PSG or portable monitors. Generally, portable devices (and PSG) are categorized by the number and type of channels measured. Each channel separately monitors and measures indicators of the physiological status of organs. Combinations of these channels are used in different types of devices for the diagnosis of sleep apnea. For example, a sleep-facility-based PSG includes at least the following channels: electroencephalography, electrooculography, electromyography, heart rate or electrocardiography (ECG), airflow, breathing/respiratory effort, and arterial oxygen saturation. Some portable devices have four monitored channels with at least two channels measuring respiratory movement, or one measuring respiratory movement combined with a channel measuring airflow, in addition to heart rate or ECG, and oxygen saturation. Other portable devices measure one, two, or three physiological indicators.

We followed the construct of our 2007 technology assessment on PSG. With the TEP, we came to agreement that PSG is an accurate measure of AHI and other (obstructive and nonobstructive) apnea measures, but is not a definitive test for OSA (syndrome) since the definition of the syndrome includes clinical judgment and arbitrary thresholds. We excluded studies with verification bias in which not everyone had PSG as the comparator.

We included all portable devices with any combination of two or more channels and those that measured the following single channels: pulse transit time, peripheral arterial tone, and pulse oximetry. We excluded studies on devices that used other single channel tests, specifically those that measured only heart rate, heart rate variability, or actigraphy alone. For the first analysis (portable versus PSG) we included only studies that performed an overnight PSG.

For the second analysis (questionnaires, etc. versus standard testing), we included studies that evaluated screening and other questionnaires, scales that included clinical criteria (e.g., signs, symptoms, history, and comorbidities), and other clinical decision making tools. These tests could be compared to either overnight PSG or portable testing. We excluded studies that assessed only single patient characteristics or risk factors. We also excluded tests that were not validated in a group of participants separate from the sample used to develop the test. Accepted studies either validated their models in a separate subgroup of study participants or had their models evaluated in subsequent studies.

Phased testing (Key Question #2). We included any study that directly compared phased testing (a series of tests performed dependent on the results of initial tests) with full testing (overnight PSG) alone.

Preoperative screening (Key Question #3). We included studies that assessed any test or predictor of sleep apnea. Predictors of long-term outcomes (Key Question #4). We included studies that assessed AHI (or similar sleep study measures) together with other potential predictors of long-term outcomes.

Treatment of OSA (Key Question #5) and treatment compliance (Key Questions #6 & #7). We included studies that assessed almost any proposed intervention or combination of interventions to treat (or manage) OSA or to improve compliance with OSA treatment (listed below). However, for nonsurgical interventions, the patients must have used the intervention at home (or equivalent). Thus studies in which the patients received the intervention only in the sleep laboratory (primarily studies of positive airway pressure devices) were excluded. The included interventions, alone or in combination, were:

- Positive airway pressure devices (continuous positive airway pressure [CPAP], bilevel positive airway pressure, autotitrating continuous positive airway pressure, other similar devices, and device modifications designed to improve comfort or compliance)
- Oral appliances and dental devices (mandibular advancement devices, tongue-retaining devices, and other similar devices)
- Devices designed to alter sleep positions (positional therapy)
- Weight loss interventions (where the goal was improvement of OSA)
- Physical therapy, training, or strengthening of the airway
- Surgical implants in the oropharynx
- Any surgery to the airway designed to reduce airway obstruction
- Medications of current interest for possible treatment of OSA
- Based on decisions of the TEP, we excluded drugs that treat sleepiness, sleep quality, or bruxism, but not OSA, drugs used only in highly selected patients with OSA (e.g., those with Alzheimer's disease). The excluded drugs include: armodafinil, bromocriptine, donepezil, eszopiclone, and modafinil.
- Miscellaneous interventions (including, but not limited to, drugs, complementary and alternative medicine, and atrial overdrive pacing).

In studies relevant to Key Question #6, patients must have received a nonsurgical treatment (a treatment with which they would need to comply). In studies relevant to Key Question #7, patients must have received either CPAP (or a variation), an oral or dental device, or a positional therapy device, in addition to an intervention whose purpose was to improve the compliance with the device.

Outcomes of Interest [Balk [AHRQ] (2011), p. 15-16]

Diagnostic testing (Key Questions #1 & #2). We included all studies reporting concordance or agreement among tests, predictive value (sensitivity, specificity) for diagnosis, change in clinical management, and clinical outcomes.

Preoperative screening (Key Question #3). We included studies reporting all intraoperative events, surgical recovery events, surgical recovery time, postsurgical events, length of intensive care or hospital stay, and intubation or extubation failures.

Predictors of long-term outcomes (Key Question #4). We included analyses of long-term clinical outcomes of interest, including all-cause mortality, cardiovascular death, nonfatal cardiovascular disease, incident hypertension, quality of life measures, incident stroke, and incident type 2 diabetes mellitus.

Treatment of OSA (Key Question #5). We included all studies reporting the following apnea related outcomes of interest (see below for descriptions of selected OSA-related outcomes):

- Sleep/wakefulness clinical outcomes
    - Quality of life outcomes, both disease specific (e.g., Functional Outcomes of Sleep Questionnaire [FOSQ], Calgary questionnaire) as well as general (e.g., Short Form survey instrument 36 [SF-36]).
    - Sleepiness / somnolence measures, including validated subjective (e.g., Epworth Sleepiness Scale) and objective measures (e.g., Multiple Sleep Latency Test, Maintenance of Wakefulness Test).
    - Neurocognitive tests, as reported by studies
    - Accidents ascribed to somnolence (e.g., motor vehicle, home accidents)
    - Productivity outcomes (e.g., work days lost)

- Objective clinical outcomes
    - Mortality
    - Cardiovascular events, including categorical changes in hypertension diagnosis or stage
    - Non-insulin-dependent diabetes (diagnosis, resolution, start or end treatment)
    - Depression events (diagnosis, recurrence, etc.).

- Intermediate or surrogate outcomes

    - Sleep study measures (from a minimum of 6 hour sleep studies)
        - Apnea-hypopnea index (AHI, continuous or categorical). If AHI not reported, we captured respiratory disturbance index or oxygen desaturation index
        - Arousal index
        - Time in deeper sleep stages (stages 3-4 and rapid eye movement sleep)
        - Sleep efficiency (percent of time spent asleep)
        - Minimum (nadir) oxygen saturation

    - Comorbidities surrogate outcomes
        - Hemoglobin A1c
        - Blood pressure (systolic, diastolic, and mean arterial pressures)

- Compliance (adherence), either categorically (whether adhering or not) or quantitatively (time using device)

- Adverse events, complications, and harms

Description of OSA-related outcomes

- Epworth Sleepiness Scale (ESS): A self-administered questionnaire which asks the patients the chances of their dozing in eight situations often encountered in daily life. Each item is rated on a 4-point scale, with a total score that can range from 0 to 24. It measures—sleep propensity as it asks about actual dozing, not—subjective sleepiness. Based on a study of normal subjects, the reference range is defined as ≤10. Domain experts consider a 1 point change in ESS to be clinically significant.

- Multiple sleep latency test (MSLT): A measurement of how quickly a subject falls asleep (when asked to) lying down in a quiet, darkened room. Sleep onset is monitored by electrodes and other wires. Though a reference range is not used in clinical practice, based on several studies of normal volunteers, a plausible reference range is 3.2 to 20 minutes.

- Maintenance of wakefulness test (MWT): A measurement of how long a subject can stay awake (when asked to) sitting in bed, resting against pillows, in a quiet, dimly lit room. Sleep onset is monitored by electrodes and other wires. Using a 20 minute protocol, a plausible reference range is approximately 12 to 20 minutes (staying awake).

- Apnea-hypopnea index (AHI): The number of episodes of apnea (complete airflow cessation) plus the number of hypopneas (reduced airflow) per hour of monitored sleep. Only PSG and portable monitors that measure airflow directly measure AHI. As noted above, the American Academy of Sleep Medicine uses a threshold of 15 events/hr (with or without OSA symptoms) or 5 events/hr with OSA symptoms to define OSA. Portable monitors that do not measure airflow may measure an oxygen desaturation index (ODI), the frequency of predefined oxygen desaturations (usually decreases of 3 or 4 percent). A related measure is the respiratory disturbance index (RDI), the frequency of respiratory events that disrupt sleep (in addition to apneas and hypopneas).

- Arousal index: The frequency per hour of arousals from sleep measured by electroencephalography as sudden shifts in brain wave activity.

- Slow wave sleep (stage 3 or 4 sleep): The percentage of time while asleep that the subject is in stage 3 or 4 sleep, measured by electroencephalography.

- Sleep efficiency: The percentage of time that a subject is asleep while in bed.

- Minimum oxygen saturation: The minimum oxygen saturation measured during sleep.

Treatment compliance (Key Questions #6 & #7). We included studies reporting adherence or compliance outcomes that were measured categorically as well as continuously (time spent using device per each time period).

Study Designs [Balk [AHRQ] (2011), p. 17]

We included only English-language, published, peer-reviewed articles. We did not include abstracts, conference proceedings, or other unpublished —grey literature. Sample size thresholds were chosen based primarily on practical consideration of available resources and time balanced with the likely amount of available literature.

Diagnostic testing and screening (Key Questions #1-#3). We included all prospective crosssectional or longitudinal studies of any followup duration. At least 10 study participants had to be analyzed with each test of interest. For studies pertaining to Key Question #1, we did not reevaluate studies included in the 2007 Technology Assessment of Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome, also written by the Tufts EPC. The findings of relevant studies from the previous report are summarized briefly in the appropriate sections of the Results section. These studies were also included in relevant figures; however, they are not presented in the summary tables of the present review.

Predictors of long-term outcomes (Key Question #4). We included longitudinal studies enrolling ≥500 participants with a followup ≥1 year. Included studies had to report a multivariable analysis.

Treatment of Sleep Apnea (Key Question #5) and treatment compliance (Key Question #7). We included longitudinal studies that analyzed ≥10 patients per intervention. Nonsurgical studies were restricted to randomized controlled trials (RCTs). We also included retrospective and nonrandomized prospective studies that compared surgery (including bariatric surgery) to other modes of intervention. Furthermore, we included prospective or retrospective noncomparative cohort studies of surgical interventions. However, these studies were restricted to those with at least 100 patients who received a given type of surgery. From these surgical cohort studies we evaluated only adverse events (complications). For Key

Question #5, studies of any duration were accepted as long as the interventions were used in the home setting (or equivalent). Studies for Key Question #7 were restricted to those with ≥2 weeks followup.

Treatment compliance (Key Question #6). We included longitudinal studies that analyzed ≥100 patients who were followed for ≥1 month. For analyses of compliance with CPAP, we included only prospective studies that reported multivariable analyses. We included any analysis of compliance with other devices.

Quality assessment

Balk [AHRQ] (2011) used a three-category grading system (A, B, or C) to denote the methodological quality of each study as described in the *AHRQ Methods Guide for Comparative Effectiveness Reviews* (AHRQ 2008). The grading system is defined below.

- **A (good).** Quality A studies have the least bias, and their results are considered valid. They generally possess the following: a clear description of the population, setting, interventions, and comparison groups; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting errors; clear reporting of dropouts and a dropout rate less than 20 percent dropout; and no obvious bias. For treatment studies, only RCTs may receive a grade of A.

- **B (fair/moderate).** Quality B studies are susceptible to some bias, but not sufficiently to invalidate results. They do not meet all the criteria in category A due to some deficiencies, but none likely to introduce major bias. Quality B studies may be missing information, making it difficult to assess limitations and potential problems.

- **C (poor).** Quality C studies have been adjudged to carry a substantial risk of bias that may invalidate the reported findings. These studies have serious errors in design, analysis, or reporting and contain discrepancies in reporting or have large amounts of missing information.

The strength of the body of evidence was graded for each analysis within each Key Question as per the AHRQ methods guide, with modifications as described by Balk [AHRQ] (2011). The same grading system was also used for studies which updated the Key Questions. The strength of the body of evidence was graded with one of the following: High, Moderate, Low, and Insufficient. Ratings were assigned based on the level of confidence that the evidence reflected the true effect for the major comparisons of interest (Balk [AHRQ] 2011, p. 22). Ratings were defined as follows:

- **High**. There is high confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect. No important scientific disagreement exists across studies. At least two quality A studies are required for this rating. In addition, there must be evidence regarding objective clinical outcomes.

- **Moderate**. There is moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate. Little disagreement exists across studies. Moderately rated bodies of evidence contain fewer than two quality A studies or such studies lack long-term outcomes of relevant populations. Upon reviewing the evidence, we decided that when there was no or weak evidence for clinical outcomes but sufficient evidence (see further below on this page) of a large clinical and highly statistically significant effect on the relatively important sleep study and sleepiness measures (i.e., AHI, arousal

index, minimum oxygen saturation, ESS, and FOSQ), we would rate the overall strength of evidence as moderate, despite the weak evidence on clinical outcomes.

- **Low**. There is low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate. Underlying studies may report conflicting results. Low rated bodies of evidence could contain either quality B or C studies.

- **Insufficient**. Evidence is either unavailable or does not permit a conclusion. There are sparse or no data. In general, when only one study has been published, the evidence was considered insufficient, unless the study was particularly large, robust, and of good quality.

When there were disagreements on effect estimates across different outcomes within the same comparison or when a large amount of evidence existed for only an important surrogate outcome (e.g., AHI), we also rated the strength of evidence for particular outcomes within a comparison (Balk 2011, p. 22). Similar rating categories and criteria were used; however, the descriptors were altered to delineate between rating the comparison and rating the individual outcomes within a comparison. These descriptors are modifications of the standard AHRQ approach:

- **Sufficient.** There is sufficient assurance that the findings of the literature are valid with respect to the outcome of interest within a comparison. No important scientific disagreement exists across studies. Further research is unlikely to change our confidence in the estimate of effect for this outcome.

- **Fair**. There is fair assurance that the findings of the literature are valid with respect to the outcome of interest within a comparison. Little disagreement exists across studies. Further research may change our confidence in the estimate of effect and may change the estimate for this outcome.

- **Weak**. There is weak assurance that the findings of the literature are valid with respect to the outcome of interest within a comparison. Underlying studies may report conflicting results. Further research is likely to change our confidence in the estimate of effect and may change the estimate for this outcome.

- **Limited or no evidence**. Evidence is either unavailable or does not permit estimation of an effect due to lacking or sparse data for the outcome of interest within a comparison

Methods - Key Question #8

At the direction of the Washington HTA program, Key Question #8 was added to address cost and cost effectiveness.

Search Strategy

A search in the MEDLINE®, Cochrane Database of Systematic Reviews, and Cochrane Controlled Trials Registry sources noted above for Key Questions #1 to #7, was conducted to identify relevant articles published within the last 10 years (2001 to November 2011). The search strategy for this Key Question is provided in Appendix D. A list of excluded studies with reasons for exclusion is provided in Appendix E. An additional search was conducted using the MED primary sources to identify high quality systematic reviews and technology assessments that addressed cost and economic evaluations. The primary sources searched include: Cochrane Library (Wiley Interscience), UK National Institute for Health and Clinical Excellence (NICE), Blue Cross/Blue Shield Health Technology Assessment (HTA) program, Veterans Administration TA program, BMJ Clinical Evidence, the Canadian Agency for Drugs and Technologies in Health

(CADTH), Agency for Health Research and Quality (AHRQ), Tufts Cost Effectiveness Analysis Registry, and the Trip Database.

Study Selection

The study selection criteria were modeled after that used by Balk [AHRQ] (2011) as described for Key Questions #1 to #7 above. For example, the same study populations, outcomes, interventions, and study designs were used to establish study inclusion and exclusion parameters for this Key Question.

Quality Assessment

The methodological quality of the studies was assessed using a standard instrument developed and adapted by the MED Project that are modifications of the BMJ (Drummond 1996), the Consensus on Health Economic Criteria list (Evers 2005), and the NICE economic evaluation checklist (NICE 2009). In brief, good quality economic evaluations include a well described research question with economic importance and detailed methods to estimate the effectiveness and costs of the intervention. A sensitivity analysis is provided for all important variables and the choice and values of variables are justified. Good quality economic evaluations also have low potential for bias from conflicts of interest and funding sources. Fair quality economic evaluations have incomplete information about methods to estimate the effectiveness and costs of the intervention. The sensitivity analysis may not consider one or more important variables, and the choice and values of variables are not completely justified. All of these factors might mask important study limitations. Poor quality economic evaluations have clear flaws that could introduce significant bias. These could include significant conflict of interest, lack of sensitivity analysis, or lack of justification for choice of values and variables. All of the included economic studies for Key Question #8 were rated "good" with regard to the quality assessment for minimization of bias. All studies were assessed by two independent and experienced raters. In cases where there was not agreement about the quality of the study the disagreement was resolved by conference or the use of a third rater. The economic evaluation checklist is provided in Appendix M.

Guidelines

Search Strategy

A search for relevant clinical practice guidelines (CPGs) was conducted, using the following sources: the National Guidelines Clearinghouse database, the Institute for Clinical Systems Improvement (ICSI), the Scottish Intercollegiate Guidelines Network (SIGN), the National Institute for Health and Clinical Excellence (NICE), the Veterans Administration/Department of Defense (VA/DOD) guidelines, US Preventive Services Task Force (USPSTF), Australian National Health and Medical Research Council, New Zealand Guidelines Group, Center for Disease Control and Prevention (CDC), and American Academy of Sleep Medicine. Included guidelines were limited to those published after 2005.

Quality Assessment

The methodological quality of the guidelines was assessed using an instrument (Appendix M) adapted from the Appraisal of Guidelines Research and Evaluation (AGREE) Collaboration (AGREE Next Steps Consortium 2009). The guidelines were rated by two individuals. A third rater was used to obtain consensus if there were disagreements. Each guideline was assigned a rating of good, fair, poor, based on its adherence to recommended methods and potential for biases. A guideline rated as good quality fulfilled all or most of the criteria. A fair quality guideline fulfilled some of the criteria and those criteria not fulfilled were thought unlikely to alter the recommendations. If no or few of the criteria were met, the guideline was rated as poor quality.

Policies

At the direction of the WA HTA program, select payer policies were searched and summarized. Aetna, Blue Cross Blue Shield, Grouphealth, Federal (Medicare National and Local Coverage Decisions), and WA Medicaid policies, were searched using the respective payers websites.

Findings

For Key Questions #1 to #7 (Balk 2011), the search retrieved 15,816 citations, of which 249 articles representing 234 studies were included. There were 46 studies of diagnostic tests, 17 prediction studies, and 190 studies regarding treatment. The subsequent search retrieved 92 citations. No subsequently published studies were identified that met inclusion criteria for Key Questions #1 to #7. For Key Question #8, the search retrieved 27 citations. One systematic review and five individual studies met inclusion criteria.

Balk [AHRQ]. (2011). *Comparative Effectiveness of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults*. Rockville, MD: Agency for Healthcare Research and Quality.

At the direction of the Washington HTA program, the Balk [AHRQ] (2011) systematic review was identified as the primary source of evidence for this report. The Balk [AHRQ] (2011) systematic review evaluated studies pertaining to the diagnosis and treatment of OSA in adults and included surrogate or intermediate outcomes. The Key Questions and findings of the report are summarized below. The full report is available at: Comparative Effectiveness Review of the Diagnosis and Treatment of Obstructive Sleep Apnea in Adults. The Balk [AHRQ] (2011) systematic review was assigned a quality rating of good.

Key Questions #1 to #4 cite reference to a previous 2007 *Technology Assessment of Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome* conducted by the Tufts Evidence-based Practice Center (Trikalinos 2007). The highlights of this Technology Assessment of adults follow:

- A systematic literature review of 95 studies was done; eligible studies:
    - Assessed the ability of sleep studies at baseline to predict response to CPAP

- Compared portable monitors with lab PSG
- Assessed the safety of sleep studies

- Baseline sleep indices (i.e. AHI, O_2 Sat, arousals) only modestly associated with CPAP in **patients with high (pre-test) probability for OSA**
- Did not assess distal clinical outcomes (e.g. mortality, MIs, strokes)
- Based on limited data, Type II and Type III monitors may identify AHI values that are suggestive of OSA
- Lab PSGs are not interchangeable with portable monitors, especially in home settings
- Rates of unsatisfactory studies and data corruption are higher for portable monitors in the home and the lab setting compared to the lab PSG
- The rate and severity of adverse events in sleep studies is low, whether in lab PSGs or portable monitors in any setting

Key Questions

KQ#1. How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep?

KQ#1a. How do the different tests compare in different subgroups of patients, based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke, or airway characteristics?

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

KQ#3. What is the effect of pre-operative screening for sleep apnea on surgical outcomes?

KQ#4. In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index (AHI) or oxygen desaturation index (ODI), and other patient characteristics with long term clinical and functional outcomes?

KQ#5. What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults?

KQ#5a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors?

KQ#5b. Are any of these characteristics or factors predictive of treatment success?

- Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities.

- OSA severity or characteristics: Baseline questionnaire (etc.) results, formal testing results (including hypoxemia levels), Baseline QoL; positional dependency, REM dependency

- Other: specific symptoms

KQ#6. In OSA patients prescribed non-surgical treatments, what are the associations of pre-treatment patient-level characteristics with treatment compliance?

KQ#7. What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?

KQ#1. How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep?

KQ#1a. How do the different tests compare in different subgroups of patients, based on: race, gender, body mass index (BMI), existing non-insulin dependent diabetes mellitus (NIDDM), existing cardiovascular disease (CVD), existing hypertension (HTN), clinical symptoms, previous stroke or airway characteristics?

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for this Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 25-42). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

> The American Sleep Disorders Association classified the different monitors that have been used in sleep studies into four categories, depending on which channels they record and evaluate.[34] Type I monitors are facility-based polysomnography (PSG). Type II monitors record the same information as Type I with fewer channels, and record signals that allow for the reliable identification of arousals from sleep (electroencephalography, electrooculography, electromyography, electrocardiography), and have at least two airflow channels or one airflow and one effort channel. Type III monitors contain at least two airflow channels or one airflow and one effort channel. 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. We evaluate Type III monitors separately from Type IV monitors.

To address this Key Question, we evaluated three types of comparisons: portable monitoring devices (Types II, III, and IV) versus PSG, questionnaires versus PSG or portable monitors, and clinical prediction models versus PSG or portable monitors.

We searched for prospective cross-sectional or longitudinal studies of any followup duration with at least 10 study participants analyzed with each test of interest. We did not reevaluate studies included in the 2007 *Technology Assessment of Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome* conducted by the Tufts Evidence-based Practice Center.[26] We briefly summarize the findings of the previous report. We do not present studies included in the 2007 Technology Assessment in our summary tables, but we include them in graphs, when applicable.

Comparison of Portable Devices and Polysomnography

Type II Monitors

The 2007 Technology Assessment identified three quality B studies that compared two different Type II monitors in the home setting to either the same monitor in the laboratory setting (two studies) or full laboratory PSG (one study). Difference versus average (mean bias) analyses of the apnea-hypopnea index (AHI) ranged from 0 to -2 events/hr. However, based on the 95 percent limits of agreement between portable and laboratory AHI measurements, discrepancies between the monitors and PSG were as wide as -36 to 36 events/hr. In one study, the difference between the two measurements was dependent on their average value; the portable monitor overestimated laboratory-based measurements for AHI<20 events/hr, but underestimated it in more severe cases. One study assessed the ability of a Type II monitor to predict an AHI>15 events/hr with laboratory-based PSG. Sensitivity was 81 percent, specificity 97 percent, and positive likelihood ratio >10.

No Type II monitors were identified in the update.

Type III Monitors

Findings of the 2007 Technology Assessment

The 2007 Technology Assessment included 22 studies that compared 13 different Type III monitors with facility-based PSG in various settings. In all studies, difference versus average analyses suggested that measurements of AHI with facility-based PSG and respiratory disturbance index (RDI) with portable monitors can differ substantially. The mean difference of AHI-RDI ranged from -10 to 24 events/hr. Based on the 95 percent limits of agreement between AHI and RDI measurements, discrepancies between the monitors and PSG varied from -39 to 54 events/hr. Such large discrepancies can affect clinical interpretation in some patients. For example, a discrepancy of 30 events/hr is important when the measurements are 4 and 34 events/hr by PSG and the device, respectively, but it may be irrelevant if the measurements are 40 and 70 events/hr. In most studies, the difference versus average analyses plots showed that the discordance between facility-based PSG and portable monitors increases as the AHI or RDI values get higher. None of the studies accounted for this in their analyses of concordance, and this makes the interpretation of the above findings difficult.

Analysis of sensitivity and specificity found that Type III monitors may have the ability to predict an elevated AHI (as determined by PSG) with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in laboratory-based PSG.

Description of Studies Published After the Completion of the 2007 Technology

Assessment (Balk [AHRQ] (2011) Appendix D Tables 1.1.1 & 1.1.2)

We identified seven studies[66-72] published after the completion of our previous Technology Assessment (Appendix D Table 1.1.1). Three studies were performed in the sleep laboratory setting,[68,70,71] with simultaneous recording of physiological parameters by both the device and the PSG machine, three

studies were performed both in the sleep laboratory as well as at home[66,67,69] and one study was performed in the home setting.[72] When studies were performed at home, the measurements taken by the device and the PSG machine are on different nights. The seven different Type III monitors that were included were Apnoescreen II respiratory polygraph, Stardust II, Apnea Risk Evaluation System (ARES™) Unicorder, Morpheus Hx (bedside computerized analysis system), Embletta portable diagnostic system, CID102L8 Type II device and SOMNOcheck® (SC), resulting in a total of 20 unique Type III monitors when pooled with the studies in the 2007 Technology Assessment (Appendix D Table 1.1.2). Twelve of the 20 monitors are assessed in only a single study, 7 are evaluated in 2 studies each, and one monitor is assessed in 3 studies. Therefore there is inadequate evidence to perform indirect comparisons of diagnostic efficacy between the monitors.

The number of analyzed participants in these studies ranged from 45 to 149. Three studies were graded quality A and four were graded quality B due to potential bias, the reasons for which varied across studies–incomplete reporting of population, unclear reporting of concordance results and unclear analytical strategy.

Participants were referral cases for the evaluation of suspected sleep apnea and were recruited from sleep centers or respiratory clinics. The population of subjects in the sleep laboratory setting was not different from the population of subjects assessed outside the sleep laboratory. In all studies, the majority of the participants were males. The mean ages of patients ranged from 45 to 63 years. Patients had mean Epworth Sleepiness Scale (ESS) scores (a standard measure of sleepiness symptoms) ranging from 8 to 12. At PSG, patients' mean AHI ranged from 15 to 39.9 events/hr. The data loss, or the proportion of participants who did not complete the study, ranged from 2 to 23 percent.

Concordance (Balk [AHRQ] (2011) Appendix D Table 1.1.2)

Six of the seven new studies provided enough information to perform analyses of the concordance between AHI readings from Type III monitors and PSG.[66,67,69-72] In the seventh study, the difference versus average analyses plots were not interpretable from the figure provided.[68] The Apnoescreen II, Stardust II, ARES, Morpheus Hx (bedside computerized analysis system), Embletta portable diagnostic system, CID102L8 Type II device and SOMNOcheck monitors were used in these studies.

The mean bias is the average difference between the AHI (or RDI or ODI) estimated with the portable device and the AHI measured by PSG. The mean difference of AHI-RDI ranged from -4 to 3 events/hr. Based on the 95 percent limits of agreement between AHI and RDI measurements, discrepancies between the monitors and PSG varied from -31 to 36 events/hr. Among studies that were conducted using the same monitor in both the laboratory (simultaneous recording of signals by device and PSG) and home setting (nonsimultaneous recording of signals by device and PSG), there was no major difference in the range of mean bias reported in both settings.

When we considered all studies, including the 22 studies from the 2007 Technology Assessment, the results pointed to the same direction. The mean difference of AHI-RDI ranged from -10 to 24 events/hr. Based on the 95 percent limits of agreement between AHI and RDI measurements, discrepancies between the monitors and PSG varied from -39 to 54 events/hr.

Sensitivity and Specificity (Balk [AHRQ] (2011) Tables 2a and 2b; Balk [AHRQ] (2011) Appendix D Table 1.1.3; Figure 3)

All seven studies assessed the sensitivity and specificity of portable monitor recordings to identify AHI suggestive of obstructive sleep apnea (OSA).[66-72] Two studies used a cutoff of AHI of 5 events/hr[68,69] and one study used a cutoff of AHI of 15 events/hr[70] in facility-based PSG to diagnose OSA. The other four studies did not report an AHI cutoff.[66,67,71,72] They reported the sensitivity and specificity for a cutoff range of 5 to 30 events/hr.

Garcia-Diaz 2007 reported sensitivity and specificity pairs for three cutoffs of RDI derived from the Type III monitor (10, 15, and 30 events/hr), recorded independently by two observers. The sensitivity for these three cutoffs ranged from 94.6 to 100 percent, and the specificity ranged from 88 to 100 percent. To 2009 used three different cutoffs for oxygen desaturation with the ARES Unicorder (drops of 4, 3, and 1 percent). A single cutoff for diagnosing sleep apnea (≥5 events/hr) was used for all desaturation levels. The best sensitivity was found with 1 percent oxygen desaturation (sensitivity 97 percent, specificity 63 percent).

Among studies that were conducted using the same monitor in both the laboratory (simultaneous recording of signals by device and PSG) and home setting (nonsimultaneous recording of signals by device and PSG), there was no major difference in the range of sensitivity and specificity reported in both settings. Across all 29 studies, including the 22 studies from the 2007 Technology Assessment, the range of sensitivity of Type III devices for predicting OSA with an AHI cutoff of 5 was 83 to 97 percent, and the range of specificity was 48 to 100 percent (Appendix D Table 1.1.3). When the AHI cutoff was increased to 15, the range of sensitivity was 64 to 100 percent and the range of specificity was 41 to 100 percent. Raising the AHI cutoff to 30, the range of sensitivity was 75 to 96 percent and the range of specificity was 79 to 97 percent.

Across all 29 studies, including the 22 studies from the 2007 Technology Assessment, the positive and negative likelihood ratios were calculated and plotted on graphs for each AHI cutoff of 5, 10, 15, 20, 30, and 40 events/hr. These graphs are presented as a matrix of plots in Figure 3, illustrating the diagnostic ability of Type III portable monitors to predict an elevated AHI, at various AHI cutoffs as determined by PSG. Each cutoff of AHI is depicted in a separate plot in receiver operating characteristics (ROC) space. Each circle represents one study, and sensitivity/specificity pairs from the same study (from different cutoffs or a different device setting) are connected with lines. Studies to the left of the near-vertical thin diagonal line have a positive likelihood ratio ≥10, and studies above the near-horizontal thin diagonal line have a negative likelihood ratio ≤0.1. A high positive likelihood ratio and a low negative likelihood ratio indicate that testing with a portable monitor can accurately predict an elevated AHI (as determined by PSG).

With an AHI cutoff of 5 events/hr, most of the studies have a positive likelihood ratio ≥10 and a negative likelihood ratio close to 0.1. At the AHI cutoff of 10 events/hr, most of the studies have a positive likelihood ratio of ≥10, with some studies having a positive likelihood ratio ≥10 and a negative likelihood ratio ≤0.1. This is also seen with a cutoff of 15 events/hr. There are fewer studies evaluating the cutoff of 20 and 30 events/hr, but the results indicate a trend towards better prediction of OSA. (Figure 3)

The ROC space plots indicate that Type III monitors generally accurately diagnose OSA (determined by full PSG), and also predict different severities of OSA (defined by having AHI above different thresholds) with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in PSG.

Table 2a. Range of sensitivity and specificity of Type III monitors (n=7)

AHI cutoff by PSG (events/hr)	Sensitivity (%)	Specificity (%)
5	83 – 97	48 – 100
15	64 – 100	41 – 100
30	75 – 96	79 – 97

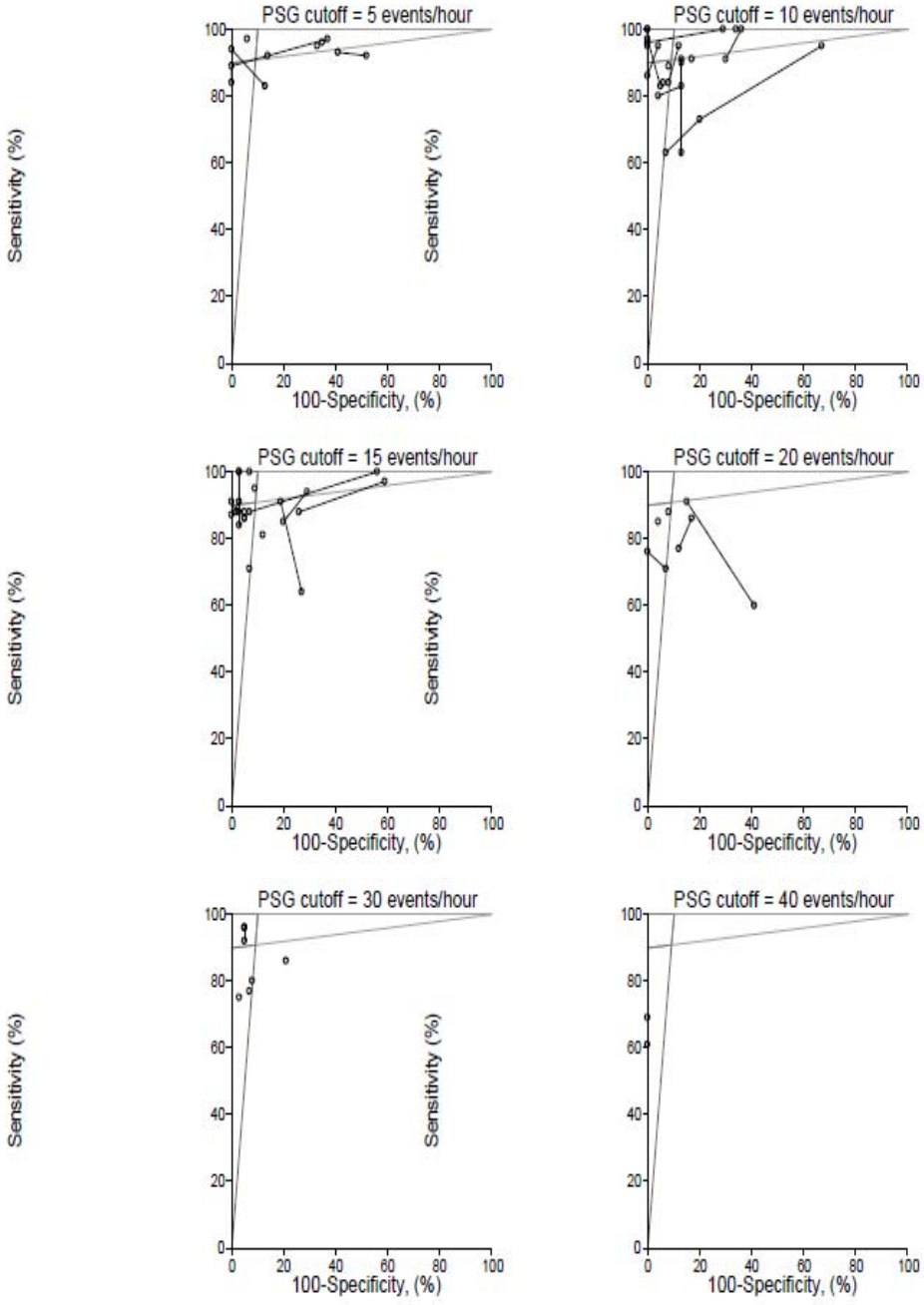
AHI = apnea-hypopnea index, nd = no data, PSG = polysomnography.

Table 2b. Range of sensitivity and specificity of Type IV monitors with ≥3, 2, and 1 channels (n=24)

AHI cutoff by PSG (events/hr)	≥3 channels (n=6) Sensitivity (%)	≥3 channels (n=6) Specificity (%)	2 channels (n=6) Sensitivity (%)	2 channels (n=6) Specificity (%)	1 channel (n=12) Sensitivity (%)	1 channel (n=12) Specificity (%)
5	85 – 100	67 – 100	92 – 98	50 – 100	85 – 96	50 – 100
15	75 – 92	50 – 100	67 – 91	78 – 96.4	43 – 100	42 – 100
30	88	100	Nd	nd	18 – 100	50 – 100

AHI = apnea-hypopnea index, nd = no data, PSG = polysomnography.

Figure 3. Diagnostic ability of Type III monitors to identify AHI cutoffs suggestive of diagnosis of OSA, and its severity, as per laboratory-based polysomnography



Sensitivity and specificity of Type III monitors in receiver operating characteristics space. Each circle represents one pair of sensitivity/specificity measurements for a given Type III monitor. Circles connected by lines represent the same monitors being tested at different thresholds within a study. The thin diagonal lines represent the thresholds for a positive likelihood ratio >10 (to the left of the near-vertical line) and negative likelihood ratio ≤0.1 (above the near-horizontal line). A high positive likelihood ratio and a low negative likelihood ratio indicate that the portable monitor has very good ability to predict the results of PSG. AHI = apnea-hypopnea index, PSG = polysomnography.

Type IV Monitors

Findings of the 2007 Technology Assessment

The 2007 Technology Assessment included 46 studies that compared 11 different Type IV monitors with facility-based PSG in various settings. In all studies, difference versus average analyses suggested that measurements of AHI with facility-based PSG and of RDI with portable monitors can differ greatly. The mean difference of AHI-RDI ranged from -17 to 12 events/hr. Based on the 95 percent limits of agreement between AHI and RDI measurements, discrepancies between the monitors and PSG varied from -49 to 61 events/hr.

Analysis of sensitivity and specificity found that studies of Type IV monitors that record at least three bioparameters showed high positive likelihood ratios and low negative likelihood ratios. Studies of Type IV monitors that record one or two bioparameters also had high positive likelihood ratios and low negative likelihood ratios for selected sensitivity and specificity pairs from ROC curve analyses.

Description of Studies Published After the Completion of the 2007 Technology Assessment (Balk [AHRQ] (2011) Appendix D Tables 1.1.2; 1.2.1-1.2.3)

We identified 24 new studies[73-96] that compared Type IV monitors with facility-based PSG in various settings. Their description and findings, stratified based on their number of channels, i.e., the number of different physiological parameters that were being measured, are presented in Appendix D Table 1.2.1 (≥3 channels), Appendix D Table 1.2.2 (2 channels), and Appendix D Table 1.2.3 (1 channel).

Fifteen studies were performed only in the sleep laboratory setting,[74,75,77-81,83-86,89,91,93,96] six were performed in both the sleep laboratory as well as the home setting,[73,87,88,92,94,95] two were performed in the home setting,[76,82] and one in a community setting.[90] The different Type IV monitors included were, ApneaLink™, ARES Unicorder, Apnomonitor, FlowWizard®, Holter Monitor, Oximetry devices, Embletta™ PDS (portable diagnostic system), ClearPath System Nx 301, Lifeshirt®, MESAM 4, RUSleeping™ RTS, SOMNIE, and WatchPAT™, resulting in a total of 23 unique monitors when pooled with the studies in the 2007 Technology Assessment (Appendix D Table 1.1.2). In one study, we reclassified a device from a Type III to a Type IV because of the particular channels used in the ARES Unicorder.[73] Six devices had more than three channels,[73,79,87,88,93,96] six had two channels,[74,80,81,85,89,95] and 12 had only a single channel.[75-78,82-84,86,90-92,94] Oximetry (either alone or in combination with snoring sound recording), ECG, or actigraphy was assessed in 22 studies. Among the remaining monitors, 14 of the 23 monitors were assessed in a single study, four (ARES, Holter ECG, Oxiflow, Sleep Strip) were assessed in two or three studies, and four (ApneaLink, Autoset, MESAM IV, WatchPAT 100) were assessed in five to eight studies. Given the heterogeneity of studies and monitors, we determined it was not appropriate to perform indirect comparisons of diagnostic efficacy between specific monitors.

The number of analyzed participants in these studies ranged from 14 to 366. Seven studies were graded quality A. Eleven studies were graded quality B due to potential bias, the reasons for which varied across studies − multiple sites with difference between sites, incomplete reporting of population, unclear reporting of results, and incomplete reporting of test blinding protocols. Six studies were graded quality C due to significant bias, with varying reasons across different studies − nonblinding of portable device tests results from PSG results, unclear reporting of results and population characteristics, and more than 50 percent dropout rate.

Participants in 19 studies were referral cases for the evaluation of suspected sleep apnea and were recruited from sleep centers or hospitals.[73,75-77,79-87,89,92-96] One study enrolled commercial motor vehicle drivers,[90] two studies recruited patients with heart failure,[74,88] one study recruited diabetic patients,[78] and one study was conducted in patients referred for uvulopalatopharyngoplasty.[91] In all studies, the proportion of male participants ranged from 32 to 100 percent. The mean ages of patients ranged from 37

to 61 years. Patients had mean ESS scores (a standard measure of sleepiness symptoms) ranging from 5.8 to 13.3. At PSG, patients' mean AHI ranged from 14 to 44 events/hr. The data loss, or the proportion of participants who did not complete the study ranged, from 0 to 78 percent. In one study among commercial truck drivers, the high rate of data loss was explained by reasons unrelated to the device performance, including termination of employment and previous history of PSG diagnosis.[90] Excluding this study, the range of data loss was 0 to 18 percent.

Concordance (Balk [AHRQ] (2011) Appendix D Tables 1.3.1-1.3.3)

Fifteen of the 24 studies provided enough information to perform analyses of the concordance between AHI readings from Type IV monitors and PSG.[73,75,77-79,81,82,85,86,88,89,92,94-96] In the other nine studies, Bland-Altman analyses were either not conducted or the Bland-Altman plots were not interpretable.

The mean difference of AHI-RDI ranged from -10 to 12 events/hr. Based on the 95 percent limits of agreement between AHI and RDI measurements, discrepancies between the monitors and PSG varied from -32 to 49 events/hr. Among studies that were conducted using the same monitor in both the laboratory (simultaneous recording of signals by device and PSG) and home setting (nonsimultaneous recording of signals by device and PSG), there was no major difference in the range of mean bias reported in both settings.

When we considered all studies, including the 46 studies from the 2007 Technology Assessment, the mean difference of AHI-RDI ranged from -17 to 12 events/hr. Based on the 95 percent limits of agreement between AHI and RDI measurements, discrepancies between the monitors and PSG varied from -49 to 61 events/hr., affecting clinical interpretation. As seen in the 2007 Technology Assessment, the difference versus average analyses plots showed that the discordance between facility-based PSG and portable monitors increases as the AHI or RDI values get higher. None of the studies accounted for this in their analyses of concordance, and this makes the interpretation of the above findings difficult.

Sensitivity and Specificity (Balk [AHRQ] (2011) Tables 2a and 2b; Balk [AHRQ] (2011) Appendix D Tables 1.1.3; 1.3.1-1.3.3, 1.1.3; Balk [AHRQ] (2011) Figure 4)

All of the studies reported the sensitivity and specificity of portable monitor recordings to identify AHI suggestive of OSA. They reported the sensitivity and specificity for a range of cutoffs from 5 to 30 events/hr.

Among the devices with three or more channels,[73,79,84,87,88,93,96] the range of sensitivity of these devices for predicting OSA with an AHI cutoff of 5 events/hr was 85 to 100 percent, and the range of specificity was 67 to 100 percent (Appendix D Table 1.3.1). When the AHI cutoff was increased to 15 events/hr, the range of sensitivity was 75 to 96 percent and the range of specificity was 50 to 100 percent. Raising the AHI cutoff to 30, one study reported a sensitivity of 88 percent and specificity of 100 percent.[79]

When evaluating devices with only two channels[74-76,81,85,89,91,95] the range of reported sensitivity of these devices for predicting OSA with an AHI cutoff of 5 events/hr was 91.8 to 97.7 percent, and the range of reported specificity was 50 to 100 percent. When the AHI cutoff was increased to 15 events/hr, the range of sensitivity was 67 to 90.6 percent and the range of specificity was 78 to 96.4 percent. (Appendix D Table 1.3.2)

In studies that assessed devices with only one channel the range of reported sensitivity of these devices for predicting OSA with an AHI cutoff of 5 events/hr was 85.4 to 96 percent and the range of reported specificity was 50 to 100 percent.[77,78,80,82,83,86,90-92,94] When the AHI cutoff was increased to 15 events/hr, the range of sensitivity was 42.5 to 100 percent and the range of specificity was 42 to 100 percent. Raising the AHI cutoff to 30 events/hr, the range of sensitivity was 18 to 100 percent and range of specificity was 50 to 100 percent (Appendix D Table 1.3.3).

Table 2 summarizes the range of sensitivity and specificity of Type IV devices with different channels.

Among studies that were conducted using the same monitor in both the laboratory (simultaneous recording of signals by device and PSG) and home setting (nonsimultaneous recording of signals by device and PSG), there was no major difference in the range of sensitivity and specificity reported in both settings.

Across all studies, including the 46 studies from the 2007 Technology Assessment, the range of sensitivity of Type IV devices for predicting OSA with an AHI cutoff of 5 was 85 to 100 percent, and the range of specificity was 50 to 100 percent. When the AHI cutoff was increased to 15, the range of sensitivity was 7 to 100 percent and the range of specificity was 15 to 100 percent.

There were 22 of 24 studies that had information that could be extracted for analysis.[73-85,87-92,94-96] Across all studies, including the 46 studies from the 2007 Technology Assessment, the positive and negative likelihood ratios were calculated and plotted on graphs for each AHI cutoff of 5, 10, 15, 20, 30 and 40 events/hr. These graphs are presented as a matrix in Figure 4, illustrating the diagnostic ability of Type IV portable monitors to predict an elevated AHI at different thresholds (as determined by PSG). With an AHI cutoff of 5 events/hr, most of the studies have a negative likelihood ratio close to 0.1. At the AHI cutoff of 10 events/hr, the studies are equally distributed in regions that indicate either a positive likelihood ratio $\geq$10 or a negative likelihood ratio $\leq$0.1. With a cutoff of 15 events/hr, the studies are spread out in regions that indicate a positive likelihood ratio $\geq$10 or a negative likelihood ratio $\leq$0.1, as well as the intersection of these regions. The studies that fall into the intersection region have the best ability to predict an elevated AHI. Similar trends are seen when cutoffs of 20 and 30 events/hr are used (Figure 4).

The ROC space plots indicate that Type IV monitors generally accurately predict an elevated AHI (as determined by PSG), though the positive likelihood ratios are lower, and negative likelihood ratios are higher, than is seen with Type III monitors.

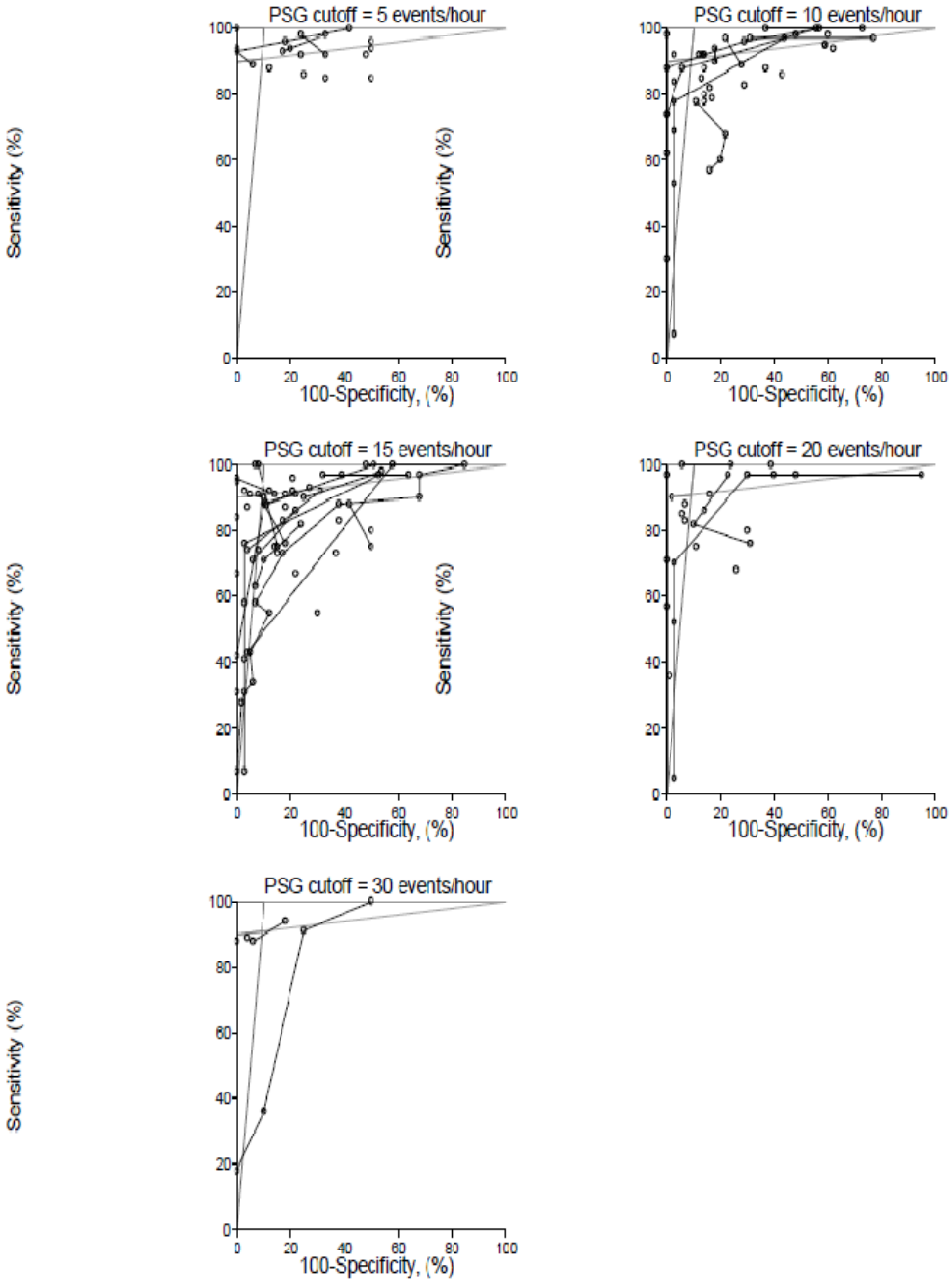
Summary

Analysis of difference versus average analyses plots suggest that substantial differences in the measured AHI may be encountered between both Type III and Type IV monitors, and PSG. Large differences compared with PSG cannot be excluded for all monitors. These studies on Type III and Type IV monitors are applicable to the general population referred to specialized sleep centers or hospitals for evaluation of suspected sleep apnea. Most of the studies are conducted either in the sleep laboratory setting or at home. Fifteen studies were graded quality A (six evaluating Type III monitors, nine assessing Type IV monitors), 45 studies were graded quality B (13 evaluating Type III monitors, 32 assessing Type IV monitors), and 39 studies were graded quality C (10 evaluating Type III monitors, 29 assessing Type IV monitors). No specific Type III monitor was evaluated by more than three studies. Among Type IV monitors, oximetry was evaluated by different monitors in 22 studies; no other monitor was evaluated by more than eight studies. No study directly compared different portable monitors to each other.

The strength of evidence is moderate that Type III and Type IV monitors may have the ability to accurately predict an elevated AHI (as determined by PSG) 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. The evidence is insufficient to adequately compare specific monitors to each other.

Based on a prior systematic review, 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. The prior review concluded that ―based on [three 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.‖

Figure 4. Diagnostic ability of Type IV monitors to identify AHI cutoffs suggestive of diagnosis of OSA, and its severity, as per laboratory-based polysomnography



Sensitivity and specificity of Type IV monitors in receiver operating characteristics space. Each circle represents one pair of sensitivity/specificity measurements for a given Type IV monitor. Circles connected by lines represent the same monitors being tested at different thresholds within a study. The thin diagonal lines represent the thresholds for a positive likelihood ratio >10 (to the left of the near-vertical line) and negative likelihood ratio ≤0.1 (above the near-horizontal line). A high positive likelihood ratio and a low negative likelihood ratio indicate that the portable monitor has very good ability to predict the results of PSG. AHI = apnea-hypopnea index, PSG = polysomnography.

Comparison of Questionnaires and Polysomnography

We identified six studies that compared sleep questionnaires with facility-based PSG in various settings (Balk [AHRQ] (2011) Appendix D Table 1.4.1). Three papers described studies performed in sleep laboratory settings,[36,97,98] one in a home setting,[99] and two in a hospital, but not in a sleep clinic or sleep laboratory.[100,101]

Two of the five studies were conducted in the same group of patients visiting a preoperative clinic;[36,97] one study was carried out among adult sleep disorder clinic patients;[98] one study was done in patients visiting their primary care physician;[99] one other study was conducted among patients attending a medical outpatient department in a tertiary care medical center;[100] and, one study was conducted among patients attending a hypertension clinic of a hospital.[101] The number of analyzed participants in these studies ranged from 53 to 211. The validated questionnaires that were administered in these studies included Berlin, STOP (Snoring, Tiredness during daytime, Observed apnea, and high blood Pressure), the STOP-Bang (STOP with body mass index [BMI], age, neck circumference, and sex variables), the American Society of Anesthesiologists (ASA) screening checklist for OSA in surgical patients, Hawaii Sleep Questionnaire, and the Epworth Sleepiness Scale. In all the studies, the cutoff of AHI in facility-based PSG that were considered suggestive of OSA was 5 events/hr.

One study was graded quality A as it had no issues in reporting of the study.[101] However, the study was not primarily designed to evaluate the two instruments (Berlin questionnaire and the Epworth Sleepiness Scale), and it assessed the association of various clinical factors with the risk for OSA. It was included because the sensitivity and specificity for the index tests were reported. One study was graded quality B due to inadequate reporting of the results of the PSG, and four were graded quality C either due to selection bias or a dropout rate higher than 40 percent. These studies are applicable to patients visiting preoperative clinics, sleep laboratories, and primary care centers for evaluation of sleep apnea.

Berlin Questionnaire (Balk [AHRQ] (2011) Appendix D Table 1.4.2)
Four studies assessed the sensitivity and specificity of the Berlin questionnaire in identifying AHI suggestive of OSA.[97,99-101] The Berlin questionnaire predicts the risk of OSA as high or low based on a score in three categories of questions related to snoring, tiredness, and blood pressure.

The number of subjects enrolled in the three studies ranged from 53 to 2,127, but the number of subjects analyzed ranged from 53 to 211. The subjects were either patients from preoperative clinics,[97] or from the population visiting their primary care physician,[99] or a department in a hospital.[100,101] The percentage of male subjects ranged from 42 to 80 percent, with the average age ranging from 46 to 55 years and average BMI ranging from 28 to 30 kg/m2. The mean baseline AHI ranged from 5 events/hr to 21 events/hr.

Chung 2008 reported sensitivity and specificity pairs for three cutoffs of the AHI index (5, 15, and 30 events/hr). With an AHI cutoff of 5 events/hr, sensitivity was 69 percent and specificity 56 percent. At the AHI cutoff of 15 events/hr, the sensitivity was higher (79 percent) the specificity was lower (51 percent). At an AHI cutoff of 30 events/hr, regarded as diagnostic of severe sleep apnea, the sensitivity was higher still (87 percent) and specificity lower (46 percent). The area under the receiver operating characteristics curve (AUC) for ability of the Berlin questionnaire to predict an AHI above 5, 15, and 30 events/hr ranged from 0.67 to 0.69. In Netzer 1999, with an AHI cutoff of 5 events/hr, the sensitivity of OSA prediction per the Berlin questionnaire was 86 percent and specificity was 77 percent. Changing the AHI cutoff to 15 events/hr decreased the sensitivity (54 percent) and increased the specificity (97 percent). At AHI cutoffs of 30 events/hr, the sensitivity further decreased (17 percent) and specificity remained the same (97 percent). In Sharma 2006, a cutoff of 5 events/hr resulted in a sensitivity of 86 percent and specificity of 95 percent. In Drager 2010, with an AHI cutoff of 5 events/hr, the sensitivity of OSA prediction per the Berlin Questionnaire was 93 percent and specificity was 59 percent. Figure 5 plots the sensitivity and

specificity in the receiver operating characteristics space, illustrating the diagnostic ability of the Questionnaire to identify AHI cutoffs suggestive of diagnosis of OSA.

In summary, using an AHI cutoff of 5 events/hr, sensitivity ranged was from 69 to 93 percent and specificity from 56 to 95 percent. Using an AHI cutoff of 15, the range of sensitivity was 54 percent to 79 percent, and specificity was 51 percent to 97 percent. For the definition of severe sleep apnea using a cutoff of 30, the range of reported sensitivity was 17 percent to 87 percent and specificity was 46 percent to 77 percent. The two studies were inconsistent as to whether the Berlin Questionnaire had a high positive likelihood ratio of ―diagnosing‖ OSA or a low negative likelihood ratio of rejecting the diagnosis of sleep apnea.

STOP Questionnaire (Balk [AHRQ] (2011) Appendix D Table 1.4.2)
Chung 2008 (Pubmed identifier 18431116) a quality C study, reported the sensitivity and specificity of the STOP Questionnaire to identify AHI suggestive of OSA.[36] The STOP questionnaire predicts the risk of OSA as high or low based on answers to questions related to snoring, tiredness, witnessed apneas, and blood pressure. With an AHI cutoff of 5 events/hr, the sensitivity was 66 percent and specificity was 60 percent. Changing the AHI cutoff to 15 events/hr increased the sensitivity (74 percent) and decreased the specificity (53 percent). At AHI cutoffs of 30 events/hr, sensitivity increased (80 percent) and specificity decreased (49 percent). The AUC for the ability of the STOP questionnaire to predict an AHI above 5, 15, and 30 events/hr ranged from 0.703 to 0.769.

STOP-Bang Questionnaire (Balk [AHRQ] (2011) Appendix D Table 1.4.2)
Chung 2008 (Pubmed identifier 18431116) a quality C study, assessed the sensitivity and specificity of the STOP-Bang questionnaire to identify AHI suggestive of OSA.[36] The STOPBang questionnaire predicts the risk of OSA as high or low based on answers to questions related to snoring, tiredness, witnessed apneas, and blood pressure (as in the STOP questionnaire) in combination with anthropometric data, namely BMI (whether >35 kg/m^2), age (>50 years), neck circumference (>40 centimeters), and sex. With an AHI cutoff of 5 events/hr, sensitivity was 84 percent and specificity was 56 percent. Changing the AHI cutoff to 15 events/hr sensitivity increased to 93 percent and specificity decreased to 43 percent. At AHI cutoffs of 30 events/hr, sensitivity further increased to 100 percent and specificity decreased to 37 percent. The AUC for ability of the STOP-Bang questionnaire to predict an AHI above 5, 15 and 30 events/hr ranged from 0.782 to 0.822.

American Society of Anesthesiologists Checklist (Balk [AHRQ] (2011) Appendix D Table 1.4.2)
Chung 2008 (Pubmed identifier 18431117) a quality C study assessed the sensitivity and specificity of the ASA screening checklist to identify AHI suggestive of OSA in surgical patients.[97] The ASA checklist predicts the risk of OSA as high or low based on results from three categories: predisposing physical characteristics (including BMI, neck circumference, craniofacial abnormalities, nasal obstruction, and tonsillar position), history of apparent airway obstruction during sleep, and reported or observed somnolence. With an AHI cutoff of 5 events/hr the sensitivity was 69 percent and specificity was 56 percent. An AHI cutoff of 15 increased the sensitivity to 79 percent and decreased specificity to 51 percent. Using an AHI cutoff of 30 events/hr increased sensitivity to 87 percent and decreased to specificity 46 percent. The AUC for the ability of the ASA Checklist to predict an AHI above 5, 15, and 30 events/hr ranged from 0.617 to 0.783.

Hawaii Sleep Questionnaire (Balk [AHRQ] (2011) Appendix D Table 1.4.2)
Kapuniai 1988 (quality B) assessed the sensitivity and specificity of the apnea score derived from the Hawaii Sleep Questionnaire to identify an AHI suggestive of OSA.[98] The questionnaire included queries about characteristics in sleep apnea patients including, stopping breathing during sleep, loud snoring, and waking from sleep gasping for or short of breath. Additional questions on sex, age, height, weight, sleep history, and history of tonsillectomy or adenoidectomy were also collected. The final model included self-reports of loud snoring, breathing cessation during sleep, and adenoidectomy in a regression model to calculate an Apnea Score. An apnea score $\geq$3 as per the model was considered high risk for sleep apnea.

Additionally, an apnea score ≥2 without details about adenoidectomy was used as a cutoff to indicate a high risk of sleep apnea. With an AHI cutoff of 5 events/hr, the sensitivity of OSA prediction per an apnea score of ≥3 was 59 percent and the specificity 69 percent. When the apnea score cutoff of ≥2 was used, sensitivity was 70 percent and specificity was 65 percent. Using an AHI cutoff of 10, the sensitivity was 78 percent and specificity was 67 percent.

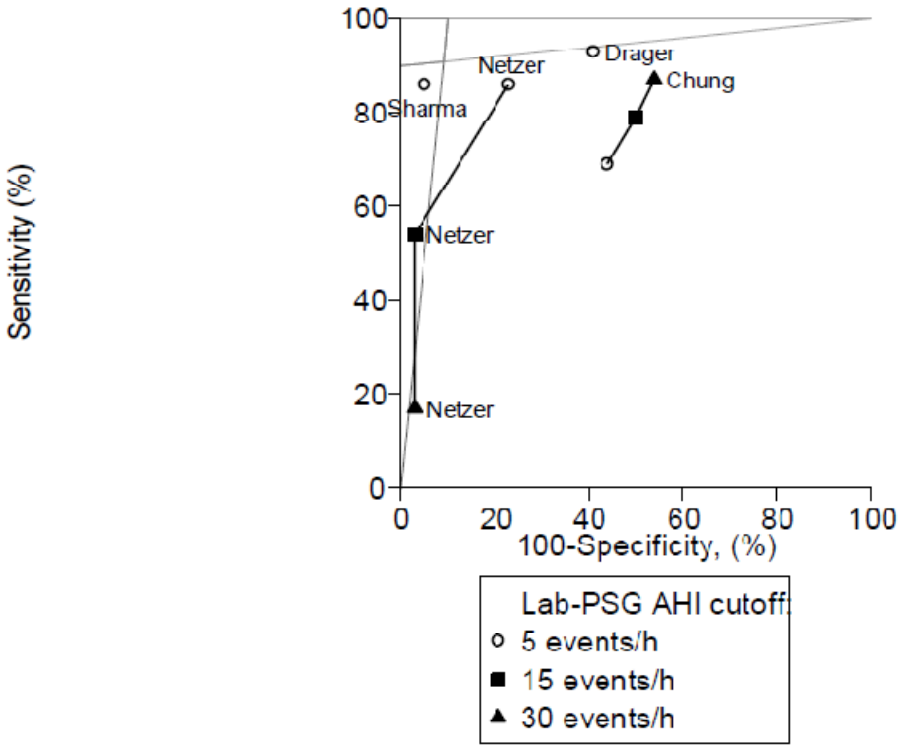
Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 1.4.2)

Drager 2010 (quality A) assessed the sensitivity and specificity of ESS to identify an AHI suggestive of OSA.[101] With an AHI cutoff of 5 events/hr, the sensitivity of OSA prediction per an ESS score >10 (defined as excessive daytime sleepiness) was 49 percent and the specificity 80 percent.

Summary

Overall, largely because of the likely selection biases in the quality C studies, the strength of evidence is low supporting the use of the Berlin questionnaire in screening for sleep apnea. Only one study each investigated the use of the STOP, STOP-Bang, ASA Checklist, Hawaii Sleep questionnaire, and ESS each. The strength of evidence is insufficient to draw definitive conclusions concerning these questionnaires.

Figure 5. Diagnostic ability of the Berlin questionnaire to identify AHI cutoffs suggestive of diagnosis of OSA and its severity as per laboratory-based polysomnography



AHI – apnea-hypopnea index, PSG – polysomnography.

Clinical Prediction Rules and Polysomnography

Overall Description of Studies Using Clinical Prediction Rules (Balk [AHRQ] (2011) Table 3; Balk [AHRQ] (2011) Appendix D Table 1.5.1)

We identified seven studies that compared clinical prediction rules with facility-based PSG in various settings (Appendix D Table 1.5.1).[102-108] All studies had either validated their models in a separate subgroup of study participants or had their models evaluated in subsequent studies. Thus, all examined clinical prediction rules are considered internally or externally validated. Six papers described studies performed in sleep laboratory settings[102-104,106-108] and one[105] in a hospital or nursing home setting.

The populations enrolled in these studies included patients referred for sleep-disordered breathing and suspected sleep apnea. The number of analyzed participants in these studies ranged from 101 to 425. The mean age of patients ranged from 47 to 79 years; the study by Onen 2008 limited enrollment to elderly individuals (≥70 years). With regard to overall methodologic quality, three studies were graded as quality A,[103,106,107] three quality B,[102,105,108] and one quality C.[104] The main methodological concerns in the quality C study were the high risk for selection bias and the high dropout rate (29 percent).

The definition of sleep apnea was based on AHI in five studies (≥5 events/hr in one study, ≥10 in one study, and ≥15 in three studies) and on RDI in two studies (≥5 events/hr). The 10 predictive models utilized questionnaire items and clinical variables in two studies,[102,103] morphometric parameters in one study,[104] standardized nurse observations during the sleep study in one study,[105] clinical variables and observations during the sleep study in two studies[106,107] and pulmonary functional data in one study[108] (Table 3).

Detailed Description of Clinical Prediction Rules (Balk [AHRQ] (2011) Appendix D Table 1.5.2)

Gurubhagavatula 2001 developed two clinical prediction rules based on a combination of a multivariable apnea prediction questionnaire score and oximetry results in 359 patients. The clinical prediction rules were developed for two separate objectives: first, to predict the diagnosis of OSA, defined as RDI ≥5 events/hr and, second, to predict the diagnosis of severe OSA, defined as RDI ≥30 events/hr, and thus select appropriate patients for split night studies. The multivariable apnea prediction questionnaire score rates apnea risk between zero and one, with zero representing low risk and one representing high risk. The authors separated the subjects into three groups based on predefined threshold scores. Those who had high scores were predicted to have OSA, those with low scores were predicted to be free of OSA, and those with intermediate scores underwent nocturnal pulse oximetry. Among these subjects, those with oxygen desaturation index (ODI) above predefined thresholds were predicted to have OSA. The optimal model parameters for each of the two clinical prediction rules were obtained by the bootstrapping technique.

The optimal model for prediction of OSA (RDI ≥5 events/hr) was determined to use the following parameters: lower score threshold = 0.14, upper score threshold = 0.58, and ODI threshold = 5.02 events/hr. This model displayed a sensitivity of 94.1 percent and a specificity of 66.7 percent.

The optimal model for the prediction of severe OSA (RDI ≥30 events/hr) was defined using the following parameters: lower score threshold = 0.38, upper score threshold = 0.9, and ODI threshold = 21 events/hr. This model displayed a sensitivity of 83.3 percent and a specificity of 94.7 percent.

Kushida 1997 developed a prediction rule based only on morphometric parameters. These parameters included the palatal height, the maxillary intermolar distance between the mesial surfaces of the crowns of the maxillary second molars, the mandibular intermolar distance between the mesial surfaces of the crowns of the mandibular second molars, the horizontal overlap of the crowns of the maxillary and

mandibular right central incisors, BMI, and neck circumference measured at the level of the cricothyroid membrane. By using a morphometric-calculated value of 70 as a threshold (range of calculated values 40-160), the model predicted the diagnosis of OSA (AHI ≥5 events/hr) with a sensitivity of 97.6 percent (95 percent CI 95.0, 98.9), a specificity of 100 percent (95 percent CI 92.0, 100), and an AUC of 0.996. The authors proposed the use of their model as a screening tool rather than a substitute for PSG.

Onen 2008 developed the Observation-based Nocturnal Sleep Inventory, a set of nurse observations performed in the patient's hospital room and made in five standardized hourly bedside visits over the course of one night. As designed, at each visit, approximately 5 minutes of listening and observation is required to detect three nocturnal conditions that characterize sleep-disordered breathing: interrupted breathing (apnea), gasping, or choking; snoring; and awakening. The authors examined three different combinations of thresholds of snoring episodes and apnea to predict diagnosis of OSA, defined as AHI ≥15 events/hr. The test accuracy of these sets of observations were: ≥2 snoring episodes or ≥1 apnea episode produced a sensitivity of 89.7 percent and a specificity of 81.4 percent; ≥3 snoring episodes or ≥1 apnea episode produced a sensitivity of 74 percent and a specificity of 93 percent; and ≥5 snoring episodes or ≥1 apnea episode produced a sensitivity of 56 percent and a specificity of 100 percent.

Rodsutti 2004 developed a clinical prediction rule based on three clinical variables (age, sex, and BMI) and two items from a self report questionnaire (reported snoring, and reported cessation of breathing during sleep). Each of these variables was stratified into two or more categories and scores were assigned to each category. The sum of the individual scores for the five variables was then calculated to obtain a summary score that could range from 0 to 7.3. The calculated sensitivities and specificities for the three categories of the summary score were: <2.5–sensitivity 0 percent, specificity 89 percent; 2.5-4.2–sensitivity 44 percent, specificity 85 percent; ≥4.2–sensitivity 76 percent, specificity 60 percent.

Crocker 1990 developed a statistical model to predict the probability of a patient having an AHI >15 events/hr, based on logistic regression of data from a 24-item questionnaire and clinical characteristics on 105 patients. The regression equation that was developed included witnessed apneas, hypertension, BMI, and age. The model displayed relatively high sensitivity (92 percent), but low specificity (51 percent). The same model was examined by Rowley 2000 in an independent set of patients.

Rowley 2000 tested the performance of Crocker's model to predict either the presence of OSA (defined as AHI ≥10 events/hr) or prioritize patients for a split-night protocol (defined as AHI ≥20 events/hr). In this dataset, the model displayed a sensitivity of 84 percent and a low specificity (39 percent) with a relatively low discrimination (AUC=0.669) for the prediction of OSA. For prioritizing patients for a split-night protocol (AHI ≥20 events/hr), the model had a sensitivity of 33 percent and a specificity of 90 percent with an AUC = 0.7.

In addition to the model developed by Crocker 1990, Rowley 2000 examined three other clinical prediction rules for the presence of OSA (defined as AHI ≥10 events/hr) or prioritizing patients for a split-night protocol (defined as AHI ≥20 events/hr). The models utilized different combinations of clinical, morphometric, and sleep observation variables. The second clinical prediction formula was based on snoring, BMI, age, and sex. This formula had a sensitivity of 96 percent with a specificity of 13 percent for the prediction of OSA, and a sensitivity of 34 percent and a specificity of 87 percent for prioritizing patients for a split-night protocol.

The third clinical prediction formula utilized snoring, gasping or choking, hypertension, and neck circumference. The performance characteristics of this prediction rule were: prediction of AHI ≥10 events/hr–sensitivity 76 percent, specificity 54 percent; prediction of AHI ≥20 events/hr–sensitivity 34 percent, specificity 89 percent.

Finally, the fourth clinical prediction formula using snoring, gasping, witnessed apneas, BMI, age, and sex predicted AHI ≥10 events/hr with a sensitivity of 87 percent and a specificity of 35 percent. With regards to the prediction of AHI ≥20 events/hr, the model had a high specificity (93 percent) with a low sensitivity (39 percent). The authors examined the predictive performance of these models in subgroups by sex, which was used as a variable in the second and the fourth clinical prediction formulas. In general, higher AUC values were attained in men (range 0.761-0.801) compared with women (range 0.611-0.648).

Zerah-Lancner 2000 developed a predictive index for OSA based on pulmonary function data obtained through spirometry, flow−volume curves, and arterial blood gas analysis. This model calculated probabilities of having a PSG positive for OSA based on specific respiratory conductance (derived from respiratory conductance and functional reserve capacity) and daytime arterial oxygen saturation. Using a threshold index of 0.5, the model predicted the presence of OSA (defined as AHI ≥15 events/hr) with 100 percent sensitivity and 84 percent specificity.

Summary

In summary, 10 different clinical prediction rules have been described in seven papers. The strength of evidence is low that some clinical prediction rules may be useful in the prediction of a diagnosis of OSA. Nine of the clinical prediction rules have been used for the prediction of diagnosis of OSA (using different criteria, AHI or RDI-based), while five of these models have been either specifically developed or also tested for the prediction of severe OSA (defined as AHI ≥20 or ≥30 events/hr), a diagnosis used for prioritizing patients for a split-night protocol. With the exception of the model by Zerah-Lancner 2000, which requires pulmonary function data, and the model by Onen 2008, which requires direct observation of patients' sleep, all other models are parsimonious, utilizing easily attainable variables through clinical interview and examination (including oximetry and morphometric measurements) and items collected from questionnaires. Only Rowley 2000 examined different prediction rules in the same patients. In this study, no predictive rule with desirable performance characteristics (both high sensitivity and specificity) was found for the prediction of OSA (range of sensitivities 76-96 percent, range of specificities 13-54 percent) or severe OSA (ranges of sensitivities 33-39 percent, range of specificities 87-93 percent). Of the remaining models, the morphometric model by Kushida 1997 gave near perfect discrimination (AUC= 0.996), and the pulmonary function data model by Zerah-Lancner 2000 had 100 percent sensitivity with 84 percent specificity. However, while all the models were internally validated, definitive conclusions on the applicability to the population at large of these predictive rules in independent populations cannot be drawn from the available literature. It should be further noted that no study examined the potential clinical utility of applying these prediction rules to clinical practice.

Table 3. Descriptions of clinical prediction rules

Study PMID	Clinical Prediction Rule	Description
Crocker, 1990[102] 2368960	Statistical model	Derived by logistic regression on data from a 24-item questionnaire and clinical features.
Gurubhagavatula, 2001[103] 11734444	Clinical prediction rule, derived	Combination of Multivariable Apnea Prediction (MAP) questionnaire score and oximetry results. MAP score predicts apnea risk using a score between 0 and 1, with 0 representing low risk and 1 representing high risk. Oximetry desaturation index (ODI) using a 3% drop (ODI3) as well as a 4% drop (ODI4) in oxygen saturation. Optimal model parameters obtained by the bootstrapping technique.
Kushida, 1997[104] 9341055	Morphometric model	Model: P + (Mx - Mn) + 3 X OJ + 3 X [Max (BMI -25, 0)] X (NC / BMI) P = palatal height (in millimeters), Mx is the maxillary intermolar distance (in millimeters) between the mesial surfaces of the crowns of the maxillary second molars. Mn is the mandibular intermolar distance (in millimeters) between the mesial surfaces of the crowns of the mandibular second molars, OJ is the overjet (in millimeters) or the horizontal overlap of the crowns of the maxillary and mandibular right central incisors, BMI is the body mass index (kg/m2; ideal BMI <25), Max (BMI -25, 0) refers to the larger of the two quantities: BMI - 25, or zero. If BMI is <= 25, then [Max (BMI - 25, 0)] is zero; if BMI >25, then BMI - 25 is inserted into the formula; NC is neck circumference (in centimeters) measured at the level of the cricothyroid membrane.
Onen, 2008[105] 18775037	Observation-based Nocturnal Sleep Inventory (ONSI)	Nurse observations made in five standardized hourly bedside visits over the course of one night.
Rodsutti, 2004[106] 15283004	Clinical prediction rule, derived	Sum of the individual scores for age, sex, snoring, stops breathing, and BMI; range = 0 -7.3.
Rowley, 2000[107] 11083602	Model #1	Clinical prediction model #1: Probability of predicting AHI ≥10 = $1/(1+e^{-(-13.5+0.06a+2.58b+0.23c+1.35d)})$ where a = age; b= 1 if witnessed apneas present, 0 if witnessed apneas absent; c = BMI; d = 1 if patient has hypertension, 0 if hypertension absent.
	Model #2	Clinical prediction model #2: Probability of predicting AHI ≥10 = $e^x/(1+e^x)$ where, x = -10.5132 + 0.9164*sex + 0.0470*age + 0.1869*BMI+1.932*snoring; where sex = 1 for male, 0 for female, snoring = 1 for present, 0 for absent.
	Model #3	Clinical prediction model #3: Probability of predicting AHI ≥10 = $(10^{[-2.132+0.069*NC+0.31*H+0.206*HS+0.224*PR]}+1)$ where NC=neck circumference, H=1 if hypertension, 0 if hypertension absent, HS=1 if habitual snorer, 0 if not, PR = 1 if reports nocturnal choking/gasping, 0 if no nocturnal choking/gasping.
	Model #4	Clinical prediction model #4: Probability of predicting AHI ≥10 = $e^x/(1+e^x)$ where, x = -8.160+1.299*Index1+0.163*BMI-0.025*BMI*Index1+0.032*age +1.278*sex where, sex=1 if male, 0 if female, index1 = the mean of nonmissing values for frequency of snorting/gasping, loud snoring, breathing stops/chokes.
Zerah-Lancner, 2000[108] 11112139	Based on Pulmonary function data	Probability (p) of having a polysomnography positive for sleep apnea: logit (p)= -136 sGrs + 2.5 (100 - SaO_2) + 4.2 where specific respiratory conductance (sGrs) (in $cmH_20^{-1} * s^{-1}$) = respiratory conductance (Grs) / functional reserve capacity (FRC) SaO_2 = daytime arterial oxygen saturation in %. The estimated value of p was derived from logit (p)= $\log_e(p/1-p)$, from 0 to 1 range.

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

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

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for this Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 43-44). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

> To address this question, our literature search included any study that directly compared phased testing (a series of tests performed dependent on the results of initial tests) with full testing (overnight polysomnography [PSG]) alone. We included all prospective cross-sectional or longitudinal studies of any followup duration. At least 10 study participants had to be analyzed with each test of interest to warrant inclusion. Only one study met our inclusion criteria.[109]
>
> Gurubhagavatula 2004 assessed the accuracy of phased testing with full testing among 1,329 respondents from a pool of 4,286 randomly selected commercial driver's license holders in Pennsylvania.[109] Those respondents with an existing diagnosis of obstructive sleep apnea (OSA) or obesity-hypoventilation syndrome, or using supplemental oxygen were excluded. The respondents were mostly male (94 percent) with a mean age of 44 years, and a mean body mass index (BMI) of 28.4 kg/m^2. The study suffered from verification bias as only the participants considered to be at high risk for OSA in early testing phases were followed up with PSG. The study received a quality rating of C.
>
> To assess the presence of sleep apnea, the study compared five case-identification strategies with PSG. Of the five strategies, one assessed a two-stage testing strategy that involved the calculation of a multivariable clinical prediction rule score (from a multivariable apnea prediction questionnaire) for all participants (Stage I). The prediction score ranged from zero (no risk) to one (maximal risk for OSA), and was calculated by combining a symptom score (symptoms included self-reported frequency of gasping or snorting, loud snoring, and the frequency of breathing stops, choking, or struggling for breath) with BMI, age, and sex. A score between 0.2 and 0.9 was defined as an intermediate risk score. Participants in this category received subsequent nocturnal pulse oximetry testing (Stage II) and those with ODI ≥5 events/hr underwent PSG. OSA was defined as an ODI ≥5 events/hr and severe OSA as ≥10 events/hr. Of the 1,329 respondents, 406 (31 percent) underwent oximetry and PSG testing. Of the 1,329 respondents, 551 subjects had a multivariable apnea prediction score above 0.436 (considered a high-risk stratum), and 247 subjects (45 percent) were enrolled from that group for oximetry and PSG testing. From the group with a prediction score below 0.436 (considered a low-risk stratum), 159 participants (20 percent) were randomly enrolled for oximetry and PSG testing. From the pooled sample of 406 subjects, OSA was diagnosed in 28 percent of the subjects. In the low risk stratum, 11 percent of the subjects had sleep apnea as compared to 52 percent of those in the high risk stratum.
>
> The proportion of patients with OSA among those who were classified as intermediate risk by the multivariable apnea prediction score (between 0.2 and 0.9) and had further oximetry was not reported. The proportion of OSA in patients who were considered either high risk (score >0.9) or low risk (score <0.2) were also not reported.
>
> **Summary**

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 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 because not all participants had PSG testing. The methodological problems with this study also limit the applicability to the general population of people with OSA.

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

KQ#3. What is the effect of pre-operative screening for sleep apnea on surgical outcomes?

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for this Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 44-45). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

To address this question, our literature search included any prospective, cross-sectional or longitudinal study of any followup duration that compared use of routing screening with no or limited screening and reported all intraoperative events, surgical recovery events, surgical recovery times, postsurgical events, length of intensive care or hospital stays, and intubation or extubation failures among patients with no previous OSA diagnosis undergoing surgical procedures.

Two studies met selection criteria (Balk [AHRQ] (2011) Appendix D Table 3.1).[97,110] Both studies were rated quality C as they had different selection criteria for enrolling subjects in the two comparative arms, indicating a substantial risk of selection bias.

Hallowell 2007, in a retrospective chart review of patients who had undergone bariatric surgery, compared 576 patients who had a PSG based on results from a clinical and physical examination (a positive, but undefined, Epworth Sleepiness Scale score, symptoms of loud snoring or daytime sleepiness, or clinical suspicion by the surgeon or pulmonologist) with 318 patients who underwent a mandatory PSG. The reported outcomes included intensive care unit (ICU) admission, respiratory-related ICU admission, duration of hospital stay, and mortality. The mean age of the patients (13 percent male) was 43 years and mean body mass index (BMI) of 51 kg/m^2. The followup period was restricted to the immediate postoperative interval.

Chung 2008 was a study designed to compare different screening tools with polysomnography (PSG) in a cohort of preoperative patients (and is discussed under Key Question #1). Only about half their enrolled patients consented to PSG. The study thus compared patients who did or did not have preoperative screening with polysomnography (PSG) for complication rates (respiratory, cardiac, or neurological complications), use of prolonged oxygen therapy, requirement of additional monitoring, intensive care unit (ICU) admissions, hospital stay after surgery, readmission, and emergency department visits. The study included 416 patients scheduled to undergo elective procedures in general surgery, gynecology, orthopedics, urology, plastic surgery, ophthalmology, or

neurosurgery. Subjects were 51 percent male with a mean age of 55 years and a mean BMI of 30.1 kg/m^2. The followup period was 30 days. Though included in this review, the value of this study to address this Key Question is dubious as there was a systematic difference between those patients who did and did not have PSG. It is highly likely that those who underwent testing were (or considered themselves to be) sicker and at higher risk of having sleep apnea.

Duration of Hospital Stay (Balk [AHRQ] (2011) Appendix D Table 3.2)
The duration of stay in the hospital was evaluated in both studies. In Hallowell 2007, among bariatric surgery patients, those who underwent mandatory testing with PSG were released on average 9.6 hr earlier than those who underwent PSG based on criteria from the physical and clinical examinations. No data were reported as to whether this difference was statistically significant. In Chung 2008, among patients who had elective general surgery procedures, those who volunteered for PSG had a nonsignificantly longer median hospital stay than those who refused PSG (difference of medians 15.5 hr)

Intensive Care Unit Admission (Balk [AHRQ] (2011) Appendix D Table 3.3)
Both studies evaluated ICU admission. In Hallowell 2007, among bariatric surgery patients, those who underwent mandatory PSG testing had a somewhat lower risk of being admitted to the ICU (relative risk [RR] = 0.62; 95 percent confidence interval [CI] = 0.32, 1.22), as compared with those who underwent selective PSG testing. In Chung 2008, among patients who had elective general surgery procedures, a greater percentage of patients who volunteered for PSG were admitted to the ICU than those who refused preoperative PSG (RR = 3.16; 95 percent CI 1.05, 9.52) [The RR's and 95 percent CI's were calculated from reported data].

Other Postoperative Outcomes (Balk [AHRQ] (2011) Appendix D Table 3.3)
In Hallowell 2007, among bariatric surgery patients, those who underwent mandatory PSG testing had a substantially, but nonsignificantly lower risk of respiratory complications leading to ICU admission (RR = 0.16; 95 percent CI 0.02, 1.27), as compared with those who underwent selective PSG testing. In Chung 2008, those who volunteered for PSG testing had significantly more total complications, and nonsignificantly more respiratory complications, cardiac complications, prolonged oxygen therapy, and additional monitoring, but nonsignificantly fewer emergency department visits within 30 days. There were no apparent differences in neurological complications, or hospital readmission within 30 days.

Summary
Two quality C prospective studies assessed the effect of preoperative screening for sleep apnea on surgical outcomes among patients with no prior OSA diagnosis. One study found that patients undergoing bariatric surgery who had mandatory PSG possibly had somewhat shorter hospital stays and, possibly, fewer respiratory-related ICU admissions than those patients who had (in a previous era) PSG based on clinical parameters. However, these differences were not statistically significant. 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 patients not willing to undergo the procedure. The methodological problems with the studies and their restricted eligibility criteria limit their applicability to the general population of people with OSA.

Overall, the strength of evidence is insufficient regarding postoperative outcomes with mandatory screening for sleep apnea.

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

KQ#4. In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index (AHI) or oxygen desaturation index (ODI) and other patient characteristics with long term clinical and functional outcomes?

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for this Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 45-49). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

> To address this question, our literature search was restricted to longitudinal studies of at least 500 participants who were assessed with formal sleep testing at baseline and followed for at least 1 year. Outcomes of interest included incident clinical events, quality of life, and psychological or neurocognitive measures. Analyses of interest were restricted to multivariable analyses of apnea-hypopnea index (AHI) (or similar sleep study measure) and demographic and clinical variables. We preferentially included analyses of baseline variables only.
>
> Eleven articles met eligibility criteria. Four evaluated predictors of all-cause mortality,[1,2,111,112] two cardiovascular death,[1,6] one each nonfatal cardiovascular events[6] and stroke,[113] two hypertension,[11,114] two type 2 diabetes mellitus,[115,116] and one quality of life.[117] Three articles each evaluated the Sleep Heart Health Study (SHHS)[2,114,117] and the Wisconsin Sleep Cohort Study.[1,11,115]
>
> **All-Cause Mortality (Balk [AHRQ] (2011) Appendix D Tables 4.1 & 4.2)**
> Four studies evaluated AHI as a predictor of all-cause mortality in multivariable analyses.[1,2,111,112] Among these studies, three enrolled participants primarily during the 1990s; the smallest study enrolled participants during the 1970s and 1980s (Lavie 1995). The two studies by Lavie (2005 & 1995) were restricted to adult men with sleep apnea symptoms or evidence of sleep apnea. The two other studies (SHHS [Punjabi 2009] and Wisconsin [Young 2008]) were large, prospective cohort studies of adults from the general population. Three of the four studies were rated quality A; the SHSS article was deemed to be quality B as a stratified analysis with cross-product terms was used instead of a full multivariable regression.
>
> All four studies found that higher baseline AHI was predictive of increased mortality over about 2 to 14 years of followup. Three of the studies evaluated categories of AHI. Each found that people with AHI >30 events/hr had a statistically significant risk of death compared with those with a low AHI (<5-10 events/hr); hazard ratios ranged from about 1.5-3.0. People in these studies with an AHI of between approximately 5 to 10 and 30 events/hr had a nonsignificantly increased risk of death. The oldest study (Lavie 1995) evaluated AHI as a continuous variable and found a significant linear association (OR = 1.012 per unit of AHI).

The SHHS analysis (Punjabi, 2009) found an interaction between AHI and both age and sex such that the association between AHI and death was seen only in men up to age 70 years. In older men (>70 yr) and in women, no significant association was found. Both SHHS and Lavie 1995 reported no substantial changes in the associations between AHI and death with the iterative addition of other predictors.

Summary

Four studies (three quality A, one quality B) found that AHI was a statistically significant independent predictor of death with long-term followup (2-14 years). The association was strongest among people with an AHI >30 events/hr. The SHHS study, however, found an interaction with sex and age such that AHI was associated with death only in men ≤70 years old.

Cardiovascular Mortality (Balk [AHRQ] (2011) Appendix D Tables 4.3 & 4.4)

Two studies evaluated AHI as a predictor of cardiovascular mortality in multivariable analyses.[1,6] Both enrolled participants primarily in the 1990s. Marin 2005 was restricted to otherwise healthy men with sleep disordered breathing. The Wisconsin Sleep Cohort Study included adults from the general population. Both studies were rated quality A.

Marin 2005 found a statistically increased risk of cardiovascular death during 10 years of followup among those with a baseline AHI ≥30 events/hr who were not treated with continuous positive airway pressure (CPAP). Those with a lower AHI or who were treated with CPAP were found to not be at an increased risk of cardiovascular death. Addition of the statistically significant predictor of existing cardiovascular disease, and the nonsignificant predictor of hypertension, did not substantially alter the association between AHI and cardiovascular death risk. The Wisconsin study found no association between AHI and cardiovascular death after 14 years of followup.

Summary

One of two studies (both quality A) found a significant independent association between an AHI ≥30 events/hr and the risk of cardiovascular death, but not lower baseline AHI, after longterm followup (10 years). The relationship was not altered by adjustment for existing cardiovascular disease or hypertension. In addition, an association was not seen in those treated with CPAP. No association was noted in the second study.

Nonfatal Cardiovascular Disease (Balk [AHRQ] (2011) Appendix D Tables 4.3 & 4.4)

Marin 2005,[6] a study of men with sleep disordered breathing, also evaluated the risk of nonfatal cardiovascular disease (myocardial infarction, stroke, or acute coronary insufficiency requiring an invasive intervention), and was also rated quality A for this outcome. The study found a similar association with nonfatal cardiovascular disease as for cardiovascular death. Only those participants with an AHI ≥30 events/hr who were not treated with CPAP were at a statistically significant increased risk of nonfatal cardiovascular disease. Adjustment for existing cardiovascular disease or hypertension did not substantially change the observed association.

Stroke (Balk [AHRQ] (2011) Appendix D Tables 4.3 & 4.4)

One study (Arzt 2005) evaluated the risk of stroke in adults aged 30 to 60 years without a previous history of stroke.[113] The participants were enrolled beginning in 1988. The study was rated quality B due to questions concerning the ascertainment of stroke. No statistically significant association was found between AHI and incident stroke during 12 years of followup. The low event rate (14/1,475) and the wide confidence intervals of the odds ratios, though, suggest that the study was highly underpowered to evaluate this outcome. However, in an analysis adjusted only for age and sex (not for body mass index [BMI]), the association between an AHI ≥20 events/hr and incident

stroke was statistically significant (OR = 4.48; 95 percent CI 1.31-15.3; P=0.02), thus suggesting that AHI and stroke are confounded with elevated BMI.

Hypertension (Balk [AHRQ] (2011) Appendix D Tables 4.5 & 4.6)
The association between AHI and risk of developing hypertension was evaluated in the two large cohort studies (SHHS and the Wisconsin Sleep Cohort Study).[11,114] The Wisconsin study excluded people with cardiovascular disease (but not hypertension). SHHS was rated quality A and the Wisconsin study was rated quality B for reasons discussed below.

In an overall analysis AHI was not an independent, significant predictor of incident hypertension in the SHHS at 5 years. However, AHI and hypertension were confounded by BMI. When BMI was not included in the model, an AHI of 15-30 events/hr and an AHI ≥30 events/hr were both significantly associated with incident hypertension (AHI = 15-30 events/hr: OR = 1.54, 95 percent CI 1.12-2.11; AHI ≥30 events/hr: OR = 2.19, 95 percent CI 1.39, 3.44).

Several subgroup analyses were also performed. Although the AHI x sex interaction term was not statistically significant (P=0.09), a significant association was found between an AHI≥30 events/hr and hypertension in women but not men. Similarly the AHI x BMI interaction term was not significant (P=0.36) but an AHI>30 events/hr was in those with a BMI less than, but not above, the median 27.3 kg/m^2. No consistent difference was found in associations of AHI and incident hypertension between those younger or older than the median age of 59 years, or with or without clinically significant sleepiness (defined as ESS ≤or >11, respectively).

The Wisconsin Sleep Cohort Study analyzed the risk of having hypertension at 4 and 8 years among people without cardiovascular disease. However, it should be noted that 28 percent of the participants had hypertension at baseline. Although the analysis adjusted for baseline hypertension, inclusion of these participants makes interpretation of the analysis unclear. Nevertheless, any AHI above 0 events/hr was found to be a statistically significant independent predictor of hypertension at 4 and 8 years of followup. Across AHI categories, it was observed that the higher the AHI the stronger the association. No interaction terms with other predictors were significant, and the results did not substantially change with the addition of sets of predictors.

Summary
In two studies, the association between AHI and future hypertension is unclear. One study found no overall independent association with incident hypertension, but found that BMI may have been a confounding factor. There were associations in subgroups of men and those with less than the median BMI, although the interaction terms were not statistically significant. The other study found that AHI was an independent predictor of future hypertension; however, the analysis included (and adjusted for) 28 percent of participants having hypertension at baseline.

Type 2 Diabetes (Balk [AHRQ] (2011) Appendix D Tables 4.7 & 4.8)
Two studies evaluated AHI as a predictor of incident type 2 diabetes mellitus in multivariable analyses.[115,116] The Wisconsin Sleep Cohort Study enrolled participants in 1988 while Botros 2009 recruited subjects with sleep disordered breathing in the early 2000s. Both excluded people with diabetes at baseline. The Wisconsin study was rated quality B due to unclear and incomplete reporting of the description of those included in the longitudinal analysis and of the results. The other study was rated quality A.

The Wisconsin study found no association between baseline AHI and the incidence of diabetes after 4 years. However, the association was confounded by waist girth. In an analysis without waist girth, a strong association was observed (AHI 5-15 events/hr: OR = 2.81, 95 percent CI 1.51-5.23, P=0.001; AHI ≥15 events/hr: OR = 4.06, 95 percent CI 1.86-8.85, P=0.0004). Botros 2009 found that

AHI ≥8 events/hr was significantly associated with incident diabetes after a mean of 2.7 years in an analysis controlled for BMI and change in BMI over the 2.7 years. The association was similar both with and without adjustment for other predictors.

Summary
Two studies suggest an association between higher AHI and incident type 2 diabetes. However, the Wisconsin study suggests that the association may be confounded by obesity, as measured by waist girth.

Quality of Life (Balk [AHRQ] (2011) Appendix D Tables 4.9 & 4.10)
The SHHS evaluated AHI as a predictor of quality of life as assessed with SF-36 after 5 years.[117] This analysis was rated quality A. The study found no statistically significant association between baseline AHI and changes in either the Physical or Mental Component Summaries.

Overall Summary
Three publications derived each from the Sleep Heart Health Study and the Wisconsin Sleep Cohort Study, and five other large cohort studies performed multivariable analyses of AHI as a predictor of long-term clinical outcomes.

A high strength of evidence indicates 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 years. 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. All other outcomes were analyzed by only one or two studies. Thus only a low strength of evidence exists that a higher AHI is associated with incident diabetes. This conclusion appears to be applicable for both the general population and specifically for patients diagnosed with sleep disordered breathing. This association, however, may be confounded with 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 longterm quality of life.

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

KQ#5. What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults?

KQ#5a. Does the comparative effect of treatments vary based on presenting patient characteristics, severity of OSA, or other pre-treatment factors?

- Characteristics: Age, sex, race, weight, bed partner, airway and other physical characteristics, specific comorbidities.
- OSA severity or characteristics: Baseline questionnaire (etc.) results, formal testing results (including hypoxemia levels), Baseline QoL; positional dependency, REM dependency
- Other: specific symptoms

KQ#5b. Are any of these characteristics or factors predictive of treatment success?

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for this Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 49-119). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

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). RCTs had to analyze at least 10 patients per intervention and the intervention had to be used for some period of time in the home setting (or equivalent). We also included prospective or retrospective studies that compared surgical interventions (including bariatric surgery) to nonsurgical treatments (with the same sample size restriction). In addition, we reviewed cohort (noncomparative) studies of surgical interventions with at least 100 patients with OSA that reported adverse event (or surgical complication) rates.

To address the subquestions to this Key Question, we sought within-study subgroup or regression analyses and, when the evidence base was sufficient and appropriate, looked for explanations of differences (heterogeneity) across studies.

In total, we found 155 eligible studies, reported in 167 articles. Of these, 132 were RCTs, 6 were prospective nonrandomized comparative studies, 5 were retrospective nonrandomized comparative studies, 2 were prospective surgical cohort studies, and 10 were retrospective surgical cohort studies.

Each section below focuses on a specific comparison between categories of interventions, with a final section focusing on adverse events. Most sections include a summary table describing the patient and study characteristics for all studies included in that section, and separate results summary tables for each outcome. We did not compile summary tables for comparisons evaluated by only one study.

Comparison of CPAP and Control

We identified 22 studies (reported in 23 articles) that compared a variety of CPAP devices with a control treatment. Twelve trials had a parallel design[118-130] and 10 were crossover trials.[131-140] One study[120] used C-Flex™ (a proprietary technology that reduces the pressure slightly at the beginning of exhalation) and the remaining trials used fixed continuous positive airway pressure (CPAP) devices. CPAP pressure was chosen manually in 13 studies, automatically determined in five, and was undefined in four. In 17 studies, it was reported that CPAP was introduced on a separate night than the diagnostic sleep study. The CPAP intervention was compared to no specific treatment in four studies, to placebo treatment (e.g., lactose tablets) in nine studies, to optimal drug treatment in one study, and to conservative measures (e.g., advice on sleep hygiene measures, weight loss) in seven studies. In four of these studies, the conservative measures were also applied to the CPAP arm.[118,127,129,130]

Mean baseline AHI ranged from 10 to 65 events/hr; nine trials included patients with an AHI ≥5, one with an AHI ≥10, seven with an AHI ≥15, two with an AHI ≥20, one with an AHI ≥30, and two did not report a lower AHI threshold. Most trials had unrestrictive eligibility criteria with the exception of Barbe 2010, which included hypertensive patients, Drager 2007, which included patients with severe OSA (mean baseline AHI = 65 events/hr), and two others (Kaneko 2003 and Mansfield 2004), which included only patients with symptomatic, stable, and optimally-treated congestive heart failure. The sample size of the studies ranged from 12 to 359 (total = 1,116 across studies). Eleven studies were rated quality B and 11 studies were rated quality C. The primary methodological concerns included small sample sizes with multiple comparisons, the lack of a power calculation, high dropout rates, incomplete reporting and, for certain crossover trials, the lack of a washout period. Overall, the studies are applicable to a broad range of patients with OSA.

Objective Clinical Outcomes

Mansfield 2004 evaluated the impact of CPAP treatment on heart failure symptomatology, as assessed by the New York Heart Association class.[126] No statistically significant improvement was found after 3 months of treatment with CPAP compared with no specific treatment for OSA. No studies evaluated other objective clinical outcomes.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.1.2; Balk [AHRQ] (2011) Figure 6)

Seven trials provided data on apnea-hypopnea index (AHI) during treatment.[119,121,123,126,127,129,140] All reported that AHI was statistically significantly lower in patients on CPAP than those on no treatment. Meta-analysis found that the difference in AHI between CPAP and control was statistically significant, favoring CPAP (difference = -20 events/hr; 95 percent CI -26, -14; P<0.001). Subgroup meta-analysis by minimum threshold AHI for study eligibility revealed that the single study with a minimum threshold of 20 events/hr found a larger difference in effect (Kaneko 2003: -28 events/hr) than the other studies that included patients with a lower AHI (range -10 to -22 events/hr); although, this difference did not fully account for the observed heterogeneity.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.1.3; Balk [AHRQ] (2011) Figure 7)

Fourteen trials provided data on the Epworth Sleepiness Scale (ESS).[118-120,126-131,135-138,140] Thirteen studies examined the comparison of CPAP versus control and one study of C-Flex versus control.[120] Nine studies reported statistically significant differences in ESS between CPAP and control, whereas the remaining five found no significant difference. Meta-analysis of all 12 studies with available data on the comparison of CPAP versus control revealed a statistically significant difference between CPAP and control, favoring CPAP (difference = -2.4; 95 percent CI -3.2, -1.5; P<0.001). However, the results were statistically heterogeneous.

Subgroup analysis by study design showed that synthesis of parallel trials (n=7) provided a significantly larger estimate of summary effect compared with crossover trials (n=5) (differences = -2.7 and -2.1, respectively, P for interaction = 0.04). A smaller effect was seen in the seven studies that included patients with an AHI as low as 5 events/hr (-2.2) as compared with the three studies that included only patients with at least 15 AHI events/hr (-4.4), but, again, this difference was not statistically significant. The single study that tested C-Flex versus no treatment (Drager 2007) demonstrated the biggest absolute reduction in ESS for the intervention arm (difference = -7.0; 95 percent CI -10.2, -3.7; P<0.001) compared with all other studies in this group.

Other Sleep Study Measures (Balk [AHRQ] (2011) Appendix D Table 5.1.4a-e; Balk [AHRQ] (2011) Figures 8 & 9)

Six studies evaluated arousal index.[119,121,123,129,139,140] All studies found greater reductions in arousal index for the CPAP arm; although in one study,[119] this difference was not statistically significant.

Meta-analysis of the five studies with sufficient data for analysis (Figure 8) revealed that arousals were significantly lower using CPAP compared with control interventions (difference = -15 events/hr; 95 percent CI -22, -7; P<0.001). Study results were found to be significantly heterogeneous. No significant difference in effect was found in the parallel design and crossover studies.

Five studies, all testing CPAP, evaluated minimum oxygen saturation (Figure 9).[121,123,126,129,140] Meta-analysis revealed the studies were heterogeneous and a statistically significant greater increase in minimum oxygen saturation while using CPAP compared with control (difference = 12 percent; 95 percent CI 6.4, 17.7; P<0.001). All studies found a statistically significant effect, although the small study by Ip 2004[121] detected a more pronounced increase of minimum oxygen saturation in favor of CPAP (difference = 27 percent; 95 percent CI 17.4, 35.8; P<0.001). Notably, this study enrolled severely hypoxemic patients (baseline minimum oxygen saturation was 65 percent in the patients randomized to the CPAP arm), which demonstrated a dramatic improvement when receiving CPAP treatment. The remaining studies were statistically homogeneous.

Sleep efficiency (measured as percent of total sleep time) was evaluated by two studies, neither of which detected a significant effect of CPAP treatment.[119,140] Five studies examined whether CPAP treatment increased the time in slow wave sleep (in absolute number of minutes or as a percentage of total sleep time) compared with control interventions.[119,123,126,139,140] Three studies found no significant differences. McArdle 2001 found a statistically significant difference of 18 minutes more when on CPAP and Mansfield 2004 reported a marginally significant net increase in the percentage of total sleep time with CPAP (4 percent, P=0.046). The same five studies found no significant differences for the outcome of rapid eye movement (REM) sleep (expressed in absolute number of minutes or as a percentage of total sleep time).

Objective Sleepiness and Wakefulness Tests (Balk [AHRQ] (2011) Appendix D Table 5.1.5a,b)
Six trials evaluated the Multiple Sleep Latency Test.[127,128,131,133,135,136] Four trials found no significant difference between CPAP and control, while Engleman 1998 and Engleman 1994 reported a statistically significant result favoring CPAP (respective net differences of 2.40 and 1.10 minutes). Meta-analysis of the six trials did not show a statistically significant difference between the interventions, but may suggest (nonsignificant) improvement with CPAP (difference = 0.78; 95 percent CI -0.07, 1.63; P=0.072).

Only Engleman 1999 evaluated the Maintenance of Wakefulness test sleep onset latency; no difference between CPAP and placebo intervention was found.

Quality of Life (Balk [AHRQ] (2011) Appendix D Table 5.1.6a,b)
Four studies evaluated results from the Functional Outcomes of Sleep Questionnaire (FOSQ).[127,131,138,140] The studies generally did not provide information on the exact FOSQ subscales that were analyzed, and the scores reported were generated by different methodologies (total summed score of responses, weighted average of subscale scores, or ratio of total summed score over maximum possible score). Thus, the reported FOSQ results appeared to be highly inconsistent (with baseline values ranging from 0.8 to 101 across studies) and a meta-analysis could not be performed. Regardless, none of the studies reported a statistically significant difference between CPAP and no treatment.

Ten studies reported on quality of life measures; five used the Short Form Health Survey 36 (SF-36),[126,129,131,137,140] four used various components of the Nottingham Health Profile,[118,127,136,137] three used the General Health Questionnaire-28,[133,135,136] two used the energetic arousal score of the University of Wales mood adjective list,[136,137] two used the sleep apnea hypopnea syndrome-

related symptoms questionnaire,[118,127] and one used the Calgary sleep apnea quality of life index (SAQLI).[129]

Overall, 29 comparisons of different quality of life measures were reported. In six trials, 11 quality of life measures reached statistical significance. In the studies that used SF-36, CPAP showed favorable results for the vitality scale in two studies,[126,137] the physical scale in two studies,[129,137] and the bodily pain in one study.[129] Among the various subscales of the Nottingham Health Profile, statistically significant differences in favor of CPAP for the physical scale were found only in one study.[118] Of the three studies using the General Health Questionnaire-28 scale, significant results were shown in only one study.[133] No significant findings were recorded for the University of Wales mood adjective list energetic arousal score in two studies, whereas one study reported significant differences for the SAQLI summary score.[129]

In summary, the impact of CPAP on quality of life is uncertain due to inconsistent findings across studies and the methodological issue of multiple testing of various quality of life subscales within these studies.

Neurocognitive and Psychological Tests (Balk [AHRQ] (2011) Appendix D Table 5.1.7)
Eight studies evaluated neurocognitive and psychological tests.[125,127,131,133,135-137,140] Of the 56 comparisons between CPAP and control, significant differences were detected in 10 comparisons across four studies; all significant differences were in favor of CPAP.[131,133,135,137] The tests with significant results included examinations of cognitive performance (intelligence quotient, digit symbol test), executive function (trailmaking), anxiety and depression scores, processing speed (Paced Auditory Serial Addition Test), and semantic fluency (the controlled oral word association test).

Blood Pressure and Hemoglobin A1c (Balk [AHRQ] (2011) Appendix D Table 5.1.8a,b)
Comparisons of daytime or nighttime blood pressure measurements between CPAP-treated patients and patients on control interventions were reported by seven studies.[120,123,127,130-132,134] No statistically significant differences were reported. Only one crossover trial (Comondore 2009) evaluated hemoglobin A1c; no difference was found between CPAP and no treatment.

Study Variability
For the main sleep study outcomes of interest (AHI, ESS, minimum oxygen saturation, and arousal index), the included studies were generally consistent in their findings, showing a beneficial effect of CPAP intervention. However, meta-analysis showed that the magnitude of the detected effects in the studies were heterogeneous. In subgroup meta-analyses by study design, there was evidence of larger effect magnitudes in parallel compared with crossover trials for the ESS outcome. The baseline severity of hypoxemia (for the minimum oxygen saturation outcome) were detected as factors influencing the magnitude of effect size for CPAP. No study reported subgroup analyses for the sleep outcomes of interest.

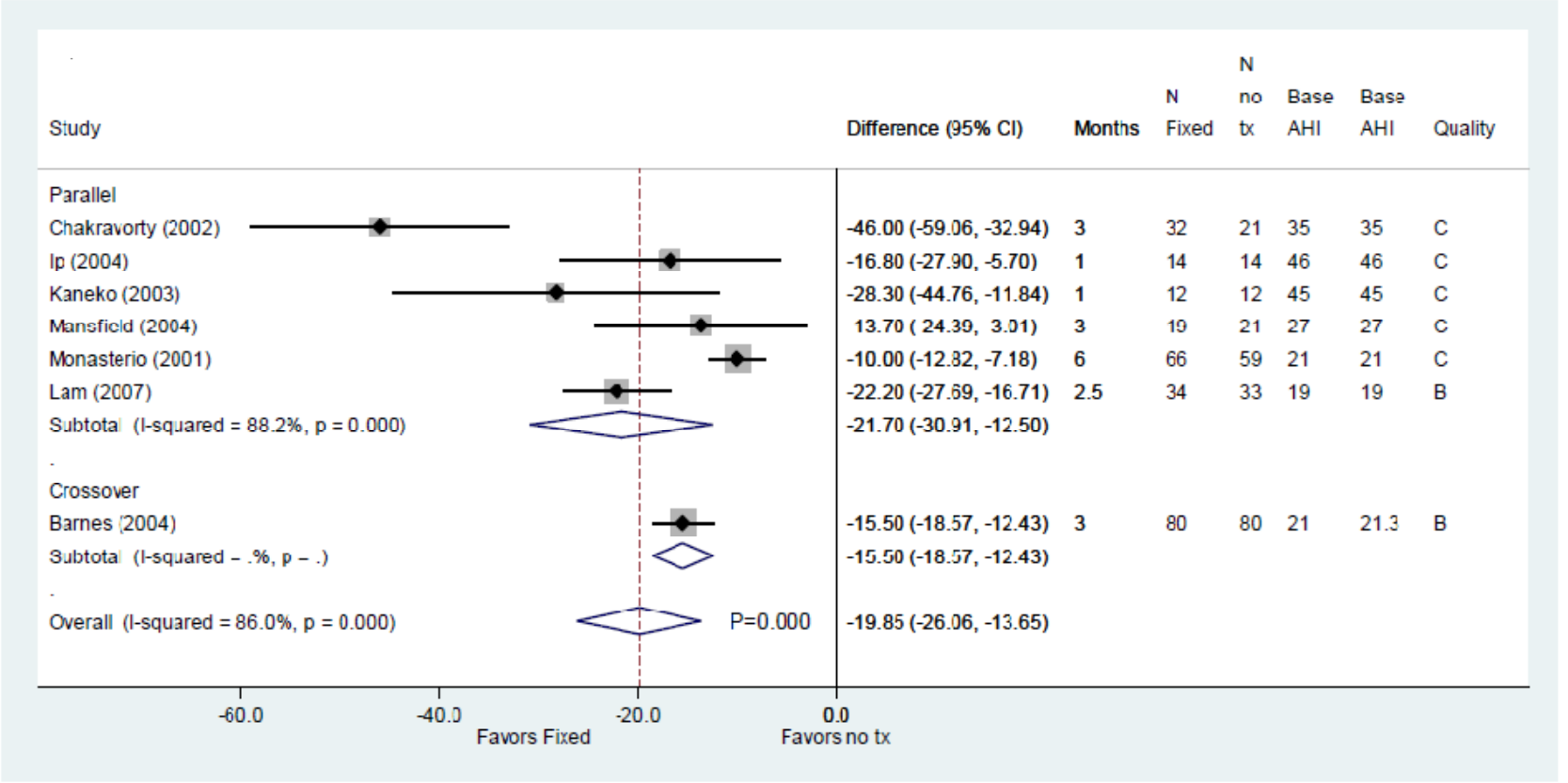
A wide range of measures were used in a small number of studies to assess quality of life, neurocognitive, and psychological outcomes. Most of these outcomes were explored as secondary endpoints. The majority of the comparisons did not report statistically significant differences in these assessments.

Summary
Eleven quality B trials and 11 quality C trials compared CPAP with control interventions. Most studies used fixed CPAP devices with manual choice of pressure. The studies reviewed generally found that CPAP was superior in reducing AHI, improving ESS, reducing arousal index, and raising the minimum oxygen saturation. These findings were confirmed by metaanalysis, although results were statistically heterogeneous. There was evidence that the magnitude of the demonstrated

efficacy of CPAP treatment may have been influenced by study design (parallel trials showed larger effect sizes), type of device, or baseline severity of disease. No consistent effect of CPAP versus control in improving other sleep study measures (slow wave and REM sleep or Multiple Sleep Latency Test) was observed. Most studies found no significant difference in quality of life or neurocognitive measures, although certain studies reported statistically significant results in favor of CPAP for the physical and vitality scales of SF-36 and various indices of cognitive performance. Generally, no consistent results were found for these measures. The wide variability in the quality of life and neurocognitive outcomes examined, and the multiple testing performed by small-sized studies, warrant cautious interpretation of any positive findings. A single study evaluated the impact of CPAP on the severity of symptoms of congestive heart failure and reported nonsignificant results. Similarly, no benefit from CPAP was found for lowering blood pressure.

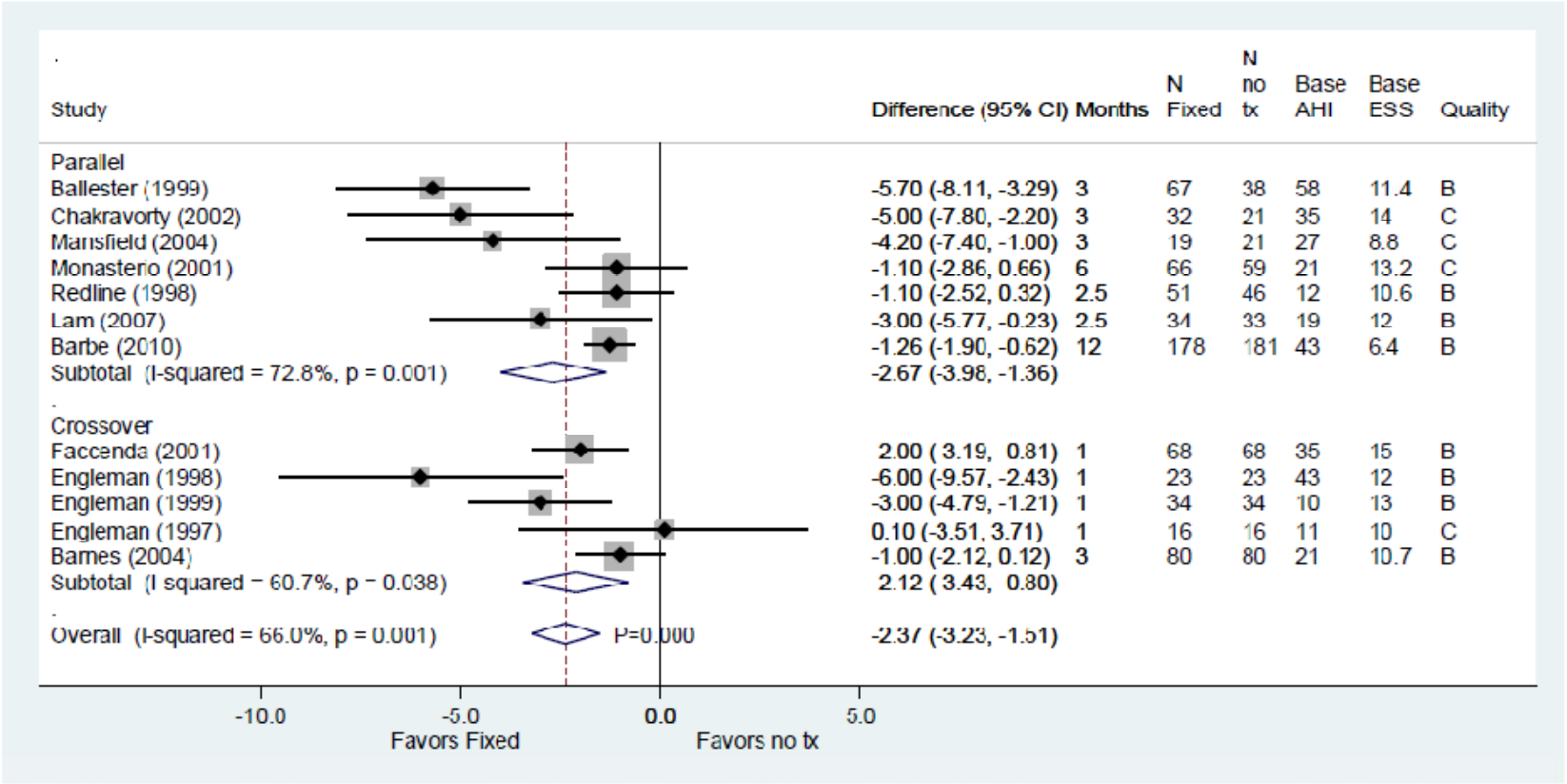
The reviewed studies report sufficient evidence supporting large improvements in sleep measures with CPAP compared with control. There is only weak evidence that demonstrated no consistent benefit in improving quality of life, neurocognitive measures or other intermediate outcomes. Despite no or weak evidence for an effect of CPAP on clinical outcomes, given the large magnitude of effect on the intermediate outcomes of AHI and ESS, the strength of evidence that CPAP is an effective treatment to alleviate sleep apnea signs and symptoms was rated moderate.

Figure 6. Meta-analysis of AHI (events/hr) in randomized controlled trials of CPAP vs. control, by study design



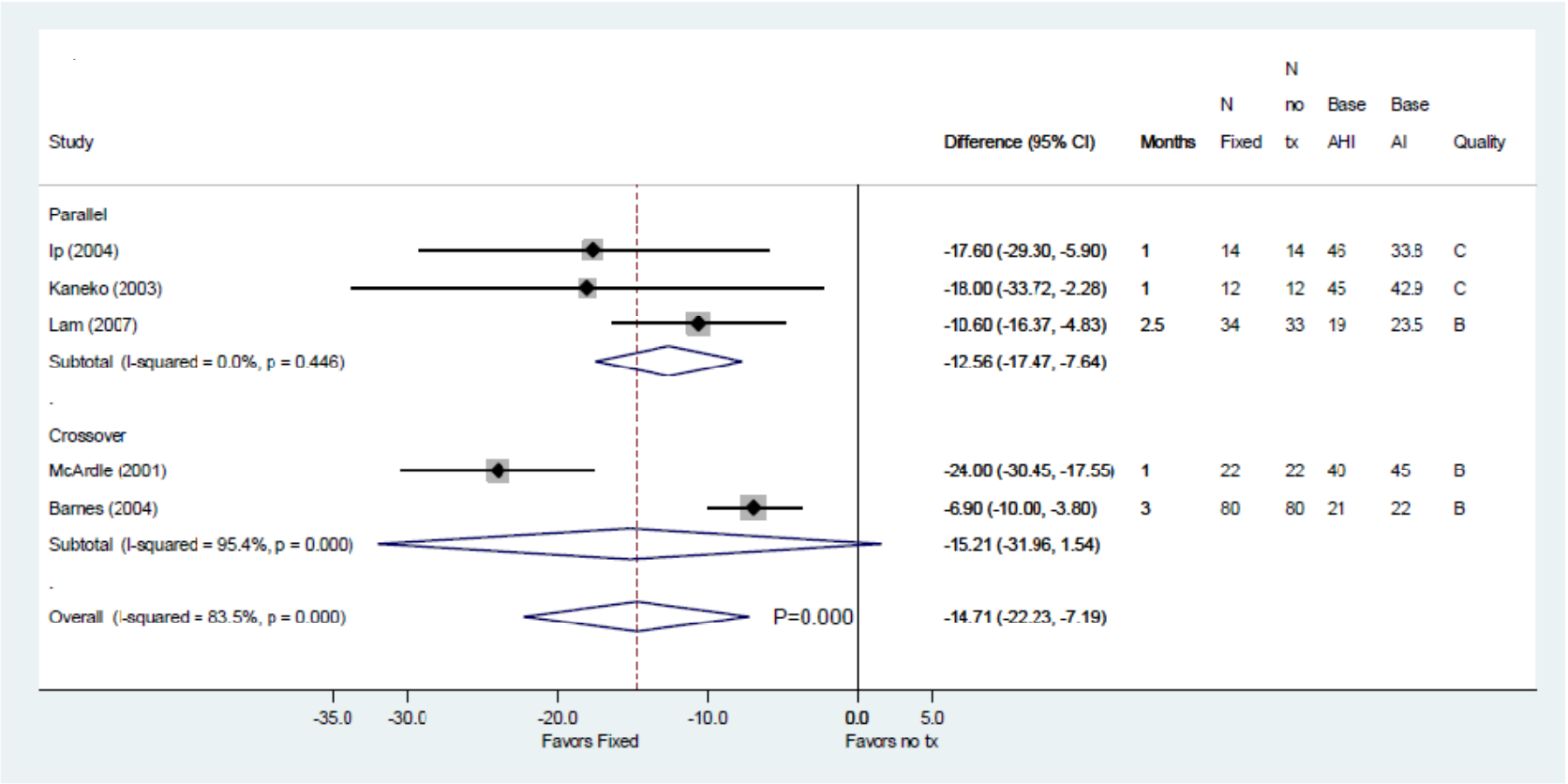
AHI = apnea-hypopnea index; CI = confidence interval; Fixed = fixed CPAP (continuous positive airway pressure); tx = treatment.

Figure 7. Meta-analysis of ESS in randomized controlled trials of CPAP vs. control, by study design



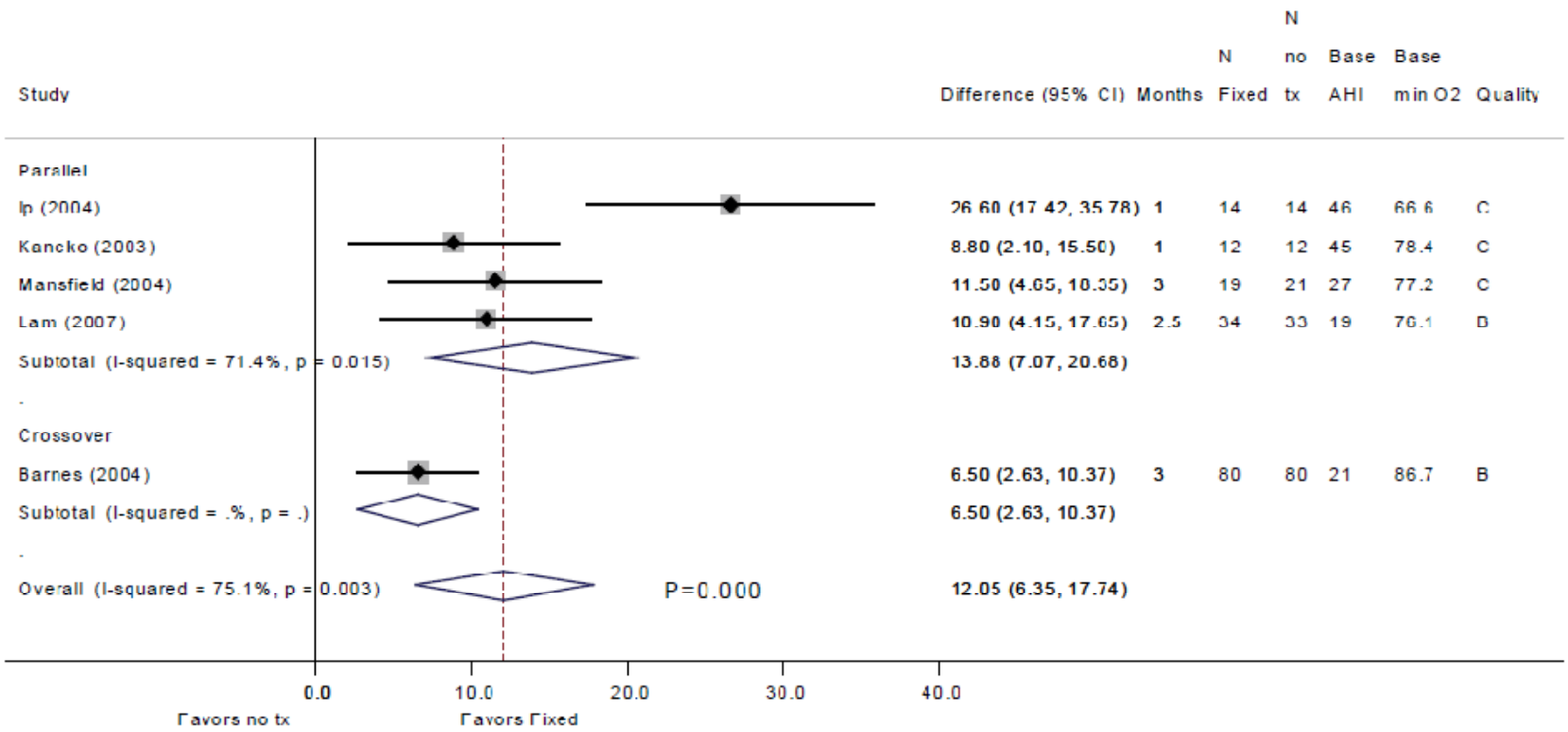
AHI = apnea-hypopnea index, CI = confidence interval, ESS = Epworth Sleepiness Scale, Fixed = fixed CPAP (continuous positive airway pressure), tx = treatment.

Figure 8. Meta-analysis of arousal index (events/hr) in randomized controlled trials of CPAP vs. control, by study design



AHI = apnea-hypopnea index, AI = arousal index, CI = confidence interval, Fixed = fixed CPAP (continuous positive airway pressure), tx = treatment.

Figure 9. Meta-analysis of minimum oxygen saturation (%) in randomized controlled trials of CPAP vs. control, by study design



AHI = apnea-hypopnea index, CI = confidence interval, Fixed = fixed CPAP (continuous positive airway pressure), min O2 = minimum oxygen saturation, tx = treatment.

Comparison of CPAP and Sham CPAP

There were 24 trials (reported in 30 articles) that compared CPAP devices with sham CPAP treatment (Balk [AHRQ] (2011) Appendix D Table 5.2.1).[141-170] Eighteen trials had a parallel design and six were crossover trials. The patients in these trials were treated with either fixed CPAP (8 trials[141,145,146,150,152,153,165,167]) or autoCPAP (16 trials[142-144,147-149,151,153-164,166,168-170]). In 19 of the 24 studies reviewed, it was reported that the CPAP was introduced on a separate full night from the night of the diagnostic sleep study.

Mean baseline AHI ranged from 22 to 68 events/hr; three trials included patients with an AHI ≥5 events/hr, five with an AHI ≥10, eight with an AHI ≥15, one with an AHI ≥20, one with an AHI ≥30, and six did not report a lower AHI threshold. Most trials had unrestrictive eligibility criteria. Exceptions were two studies (Egea 2008 and Smith 2007) that included only patients with stable and optimally-treated congestive heart failure, one study (Campos-Rodriguez 2006) that included only patients with primary hypertension and on hypertension treatment, and a final study (Robinson 2006) that included only hypertensive patients with significant OSA, but without sufficient daytime hypersomnolence. The reviewed studies were generally small with sample sizes ranging from 25 to 101 (total = 1,076 across studies), followed for 1 week to 3 months. Five studies were rated quality A, 13 studies quality B, and six studies quality C. The primary methodological concerns included small sample sizes with multiple comparisons, the lack of power calculations, high dropout rates, and incomplete reporting. Overall, the studies are applicable to a broad range of patients with OSA.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.2.2; Balk [AHRQ] (2011) Figure 10)

Nine trials provided data on AHI comparing CPAP with sham CPAP.[143,147,148,153,157,160,163,167,169] All trials had a parallel design, all except one (Lam 2010) evaluated fixed CPAP, and reported that AHI was statistically significantly lower in patients on CPAP than those on sham treatment. The one RCT that evaluated autoCPAP[169] did not report sufficient data to estimate the effect size; thus it was not included in meta-analysis. Metaanalysis revealed that the difference in AHI between CPAP and control was statistically significant, favoring CPAP (difference = -46 events/hr; 95 percent CI -57, -36; P<0.001). However, the results were statistically heterogeneous. Subgroup meta-analysis by minimum threshold AHI for study eligibility revealed that there was no statistical heterogeneity among four studies (Haensel 2007, Mills 2006, Loredo 2006, Norman 2006) that included patients with an AHI of at least 15 events/hr (difference = -58; 95 percent CI -68, -49; P<0.001). This difference was significantly larger than one study (Egea 2008) that included patients with an AHI of at least 10 events/hr (-25; P<0.0001), and another study (Loredo 1999) that included patients with an AHI of at least 20 events/hr (-37; P=0.02). Similarly, subgroup meta-analysis of two studies (Becker 2003 and Spicuzza 2006) that included patients with an AHI of at least 5 events/hr showed a lower net difference of AHI (difference = -43; 95 percent CI -65, -21); however, this effect was not statistically significant in difference compared with the five studies that included patients with an AHI of at least 15 events/hr.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.2.3; Balk [AHRQ] (2011) Figures 11 & 12)

Sixteen trials comparing CPAP with sham CPAP reported ESS data.[142-145,147,150,151,157,159,162,164-166,168-170] Eleven trials had a parallel design, and the remaining five had a crossover design. Five of the six trials

comparing autoCPAP versus sham autoCPAP reported statistically significant differences on the ESS, while only six of the 10 trials comparing fixed CPAP versus sham CPAP reported statistically significant findings. Meta-analysis of all 16 studies showed a statistically significant difference between CPAP and sham control, favoring CPAP (difference = -2.5; 95 percent CI -3.5, -1.5; P<0.001). However, the results were statistically heterogeneous.

Subgroup meta-analyses by study designs (Figure 11), by types of CPAP (Figure 12), and by minimum threshold ESS for study eligibility were conducted to explore possible factors that could explain the heterogeneity. We found that the same pooled estimates by trial designs (parallel versus crossover: -2.5 versus -2.5) but significant different pooled estimates by types of CPAP (autoCPAP versus CPAP: -1.9 versus -2.8; P=0.05). Subgroup meta-analysis by minimum threshold AHI for study eligibility showed significant net differences on the ESS among three studies including patients with an AHI of at least 10 events/hr (difference = -3.6; 95 percent CI -6.4, -0.9; P=0.01), among four studies including patients with an AHI of at least 15 events/hr (difference = -1.2; 95 percent CI -3.5, -0.3; P=0.02). The difference between the two subgroups (AHI ≥10 versus AHI ≥15) was marginally significant (P=0.08). However, the subgroup metaanalysis did not show significant differences on the ESS in two studies (Hui 2006 and Becker 2003) that included patients with an AHI of at least 5 events/hr (difference = -2.1; 95 percent CI -6.1, 1.9)

Other Sleep Study Measures (Balk [AHRQ] (2011) Appendix D Tables 5.2.4-5.2.7; Balk [AHRQ] (2011) Figure 13)

Three trials evaluated arousal index.[143,153,157] All three had a parallel design and evaluated fixed CPAP, and all three studies found greater reductions in arousal index for the CPAP arm. In one study (Becker 2003), however, this difference was not statistically significant. Meta-analysis revealed that arousals were significantly more reduced while using CPAP as compared with sham CPAP (difference = -27 events/hr; 95 percent CI -42, -12; P<0.001). Study results were significantly heterogeneous.

Only one trial evaluated minimum oxygen saturation; no significant difference in minimum oxygen saturation was observed in a comparison of CPAP with sham CPAP.[143] This trial was rated quality C due to a small sample size without a power calculation and a high dropout rate.

Sleep efficiency (measured as percent of total sleep time) was evaluated by two studies, neither of which detected a significant effect of CPAP treatment.[148,153] Four studies examined whether CPAP treatment increased the time in slow wave sleep (in absolute number of minutes or as a percentage of total sleep time) compared with sham CPAP, and all found no significant effect.[143,147,153,157] The same four studies also evaluated the outcome of REM sleep (expressed in absolute number of minutes or as a percentage of total sleep time). Three of the four studies did not find a significant effect of CPAP treatment; the other (Loredo 2006) reported that CPAP treatment significantly increased the time in REM sleep (difference = 7.5 percent of total sleep time; 95 percent CI 3.5, 11.5; P<0.05).

Objective Sleepiness and Wakefulness Tests (Balk [AHRQ] (2011) Appendix D Table 5.2.8)

One study evaluated the Multiple Sleep Latency Test outcome and found no significant difference in sleep latency test score comparing CPAP with sham CPAP.[142] Four studies evaluated the Maintenance of Wakefulness Test outcome,[151,159,168,170] with two reporting a statistically significant result, favoring autoCPAP. The remaining study (Marshall 2005) also showed a marginally significant increase in time maintaining alertness during the day in comparing CPAP with sham CPAP (P=0.09).

Quality of Life (Balk [AHRQ] (2011) Appendix D Tables 5.2.9 & 5.2.10)

Three studies administered the Functional Outcomes of Sleep Questionnaire (FOSQ), and all found no significant difference in test scores comparing CPAP with sham CPAP.[142,159,162]

Six studies (two using autoCPAP and four CPAP) measured quality of life using SF-36.[142,147,159,162,165,166] Five of the six studies did not find significant differences in physical and mental health component summary scores. The remaining study (Siccolli 2008) reported that patients who received autoCPAP treatment had significantly increased physical and mental health component summary scores compared with those who received sham treatment (differences = 8.2 and 10.8; P=0.01 and P=0.002, respectively) This study also found a similar result for the SAQLI summary score (difference = 0.9; P=0.001).

Neurocognitive and Psychological Tests (Balk [AHRQ] (2011) Appendix D Table 5.2.11)

Seven studies evaluated neurocognitive and psychological tests,[142,148,149,153,157,159,160] Of the 26 comparisons between CPAP and sham CPAP, a significant difference was detected only in one comparison of the digit vigilance test (measure of sustained attention and psychomotor speed) in one study, favoring CPAP.[160]

Blood Pressure (Balk [AHRQ] (2011) Appendix D Table 5.2.12)

Comparisons of daytime or nighttime blood pressure measurements between CPAP-treated patients and patients on sham CPAP were reported by 12 studies.[141-147,150,160,163,163,164,169] Six of these studies reported mean arterial pressure, and 10 reported systolic and diastolic blood pressure. The results were inconsistent across studies. About half of the studies reported significant blood pressure reduction favoring CPAP, and the other half reported no significant differences.

Study Variability

One study conducted a subgroup analysis of patients who had good compliance to CPAP use (≥3.5 hr/night) and found similar outcomes on the ESS and in blood pressure, favoring autoCPAP as compared with sham CPAP.[145] Trends toward larger reductions in blood pressure outcomes among this subgroup of patients were observed; however, the study did not perform a statistical analysis to test the differences between patients with good compliance and those with poor compliance.

For the main sleep study outcomes of interest (AHI, ESS, minimum oxygen saturation, and arousal index), the studies reviewed were generally consistent in their findings, showing a beneficial effect of CPAP intervention. However, our meta-analysis showed that the results of the studies were heterogeneous in terms of the magnitude of their detected effects. In subgroup meta-analyses by study designs, by types of CPAP, and by minimum threshold AHI for study eligibility to explore possible factors that may explain the heterogeneity, only minimum threshold AHI for study eligibility could account for some of the observed heterogeneity. However, no consistent patterns were seen with regard to the impacts of minimum threshold AHI for study eligibility on the main sleep study outcomes.

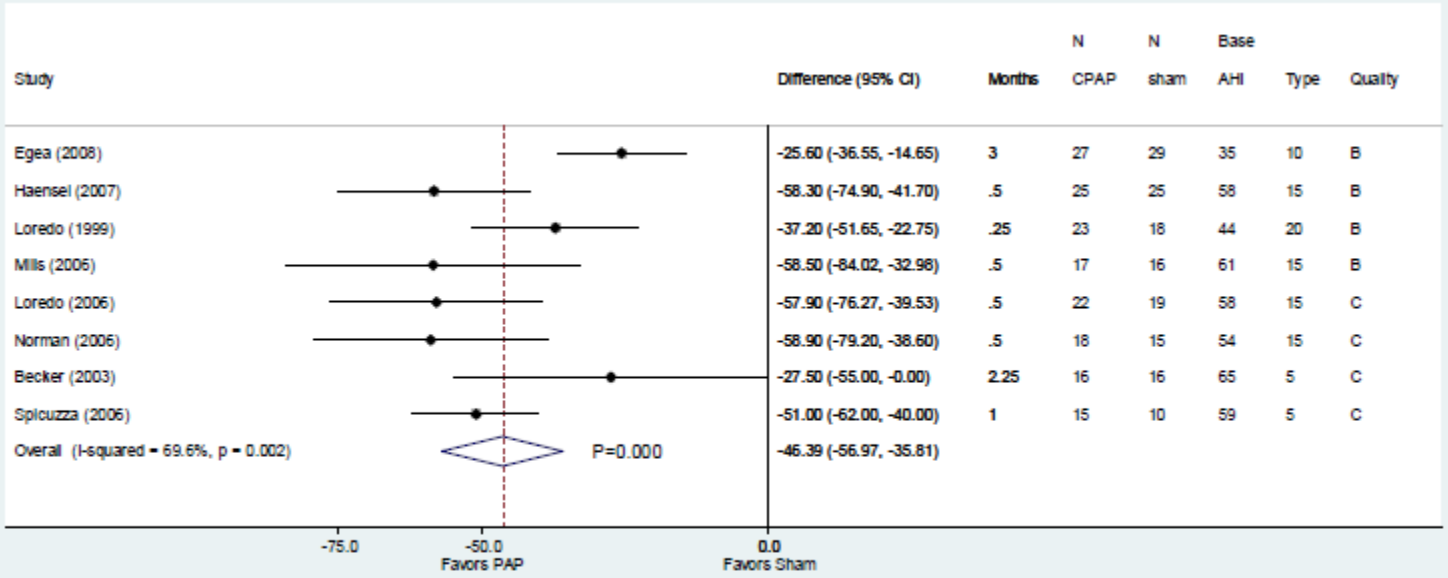
Regarding quality of life and neurocognitive outcomes, few studies used a wide range of tests and outcomes. In most cases, these outcomes were explored as secondary endpoints. Most of the comparisons performed did not reach statistical significance.

Summary

Five quality A, 13 quality B, and six quality C trials compared autoCPAP (16 trials) or fixed CPAP (8 trials) with sham treatments. The reviewed studies generally found that CPAP was superior in reducing AHI, improving ESS, and reducing arousal index. These findings were confirmed by meta-analysis, although the studies' results were statistically heterogeneous. There was evidence that the magnitude of the demonstrated efficacy of CPAP treatment may have been influenced by baseline severity of disease, although no consistent patterns were observed regarding the impacts of baseline severity of disease on the main sleep study outcomes. Most studies did not find a significant effect of CPAP versus sham in improving other sleep study measures (slow wave and REM sleep, Multiple Sleep Latency Test), but a small number of studies did show CPAP to significantly improve Maintenance of Wakefulness Test measures. Most studies also found no significant difference in effects on quality of life or neurocognitive function. The effects of CPAP on blood pressure outcomes were mixed. About half of the studies reported significant blood pressure reduction, favoring CPAP, and the other half reported no significant differences. No study evaluated objective clinical outcomes.

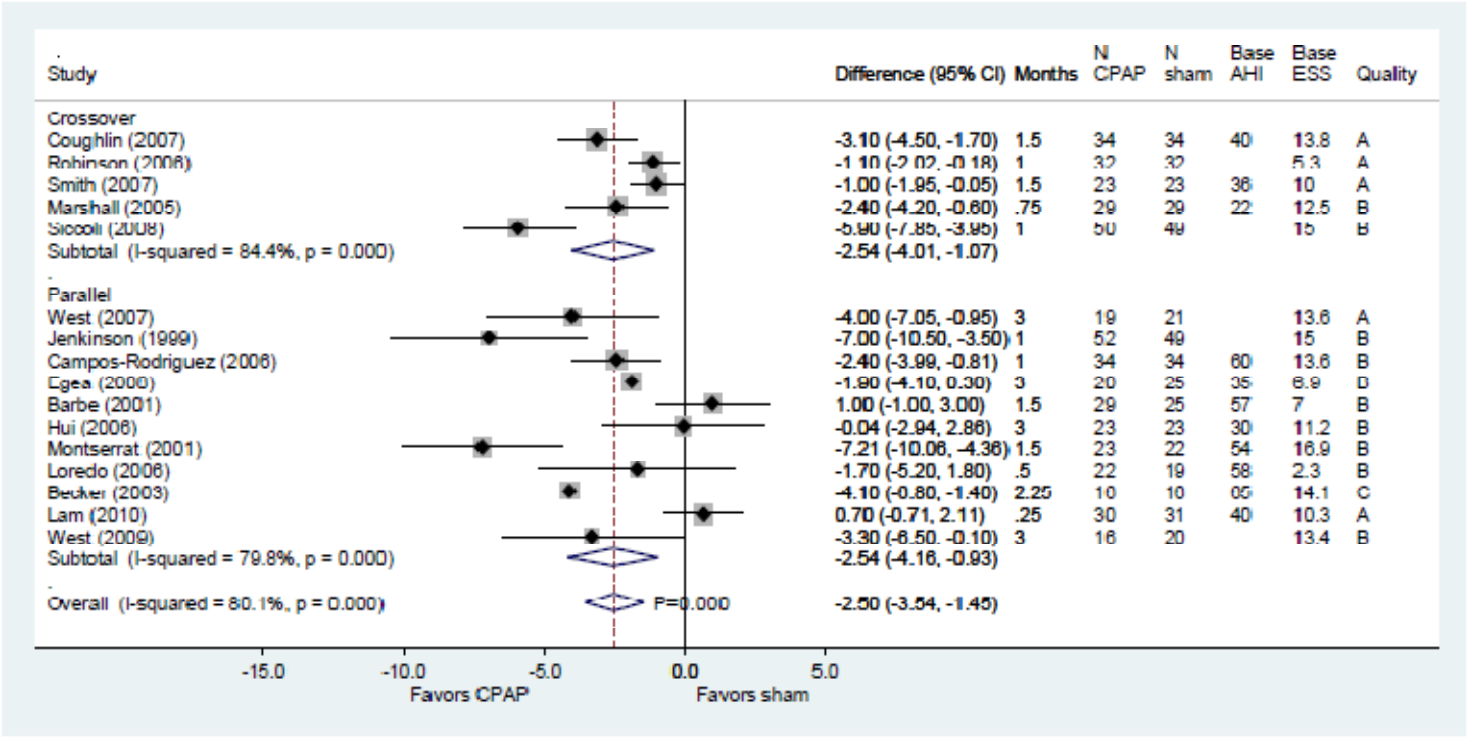
There was sufficient evidence supporting large improvements in sleep measures with CPAP compared with sham CPAP, but weak evidence that there is no difference between CPAP and sham CPAP in improving quality of life, neurocognitive measures, or other intermediate outcomes. Despite no or weak evidence 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 evidence that CPAP is an effective treatment for the relief of signs and symptoms of sleep apnea was rated moderate.

Figure 10. Meta-analysis of AHI (events/hr) in randomized controlled trials of CPAP vs. sham CPAP



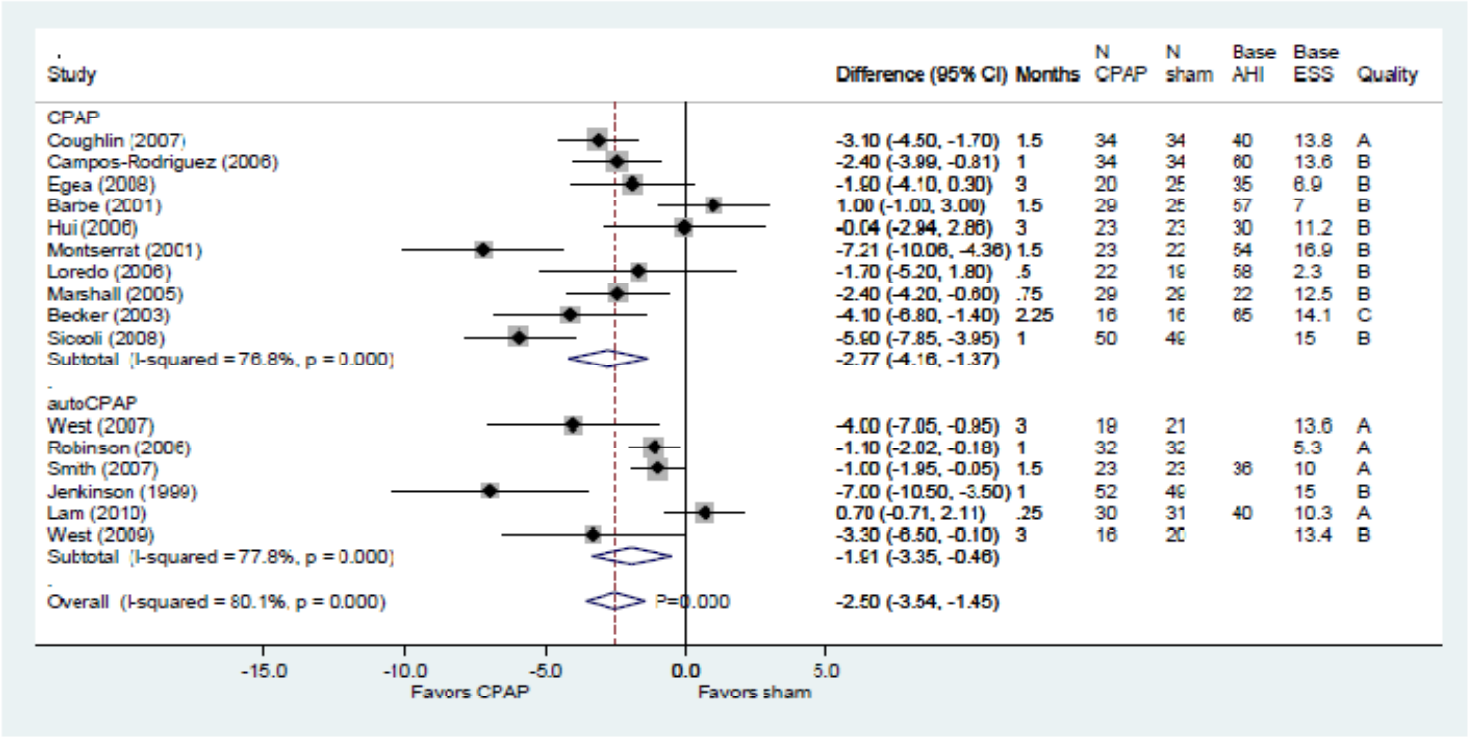
AHI = apnea-hypopnea index, CI = confidence interval, CPAP = continuous positive airway pressure.

Figure 11. Meta-analysis of ESS in randomized controlled trials of CPAP vs. sham CPAP, by study design



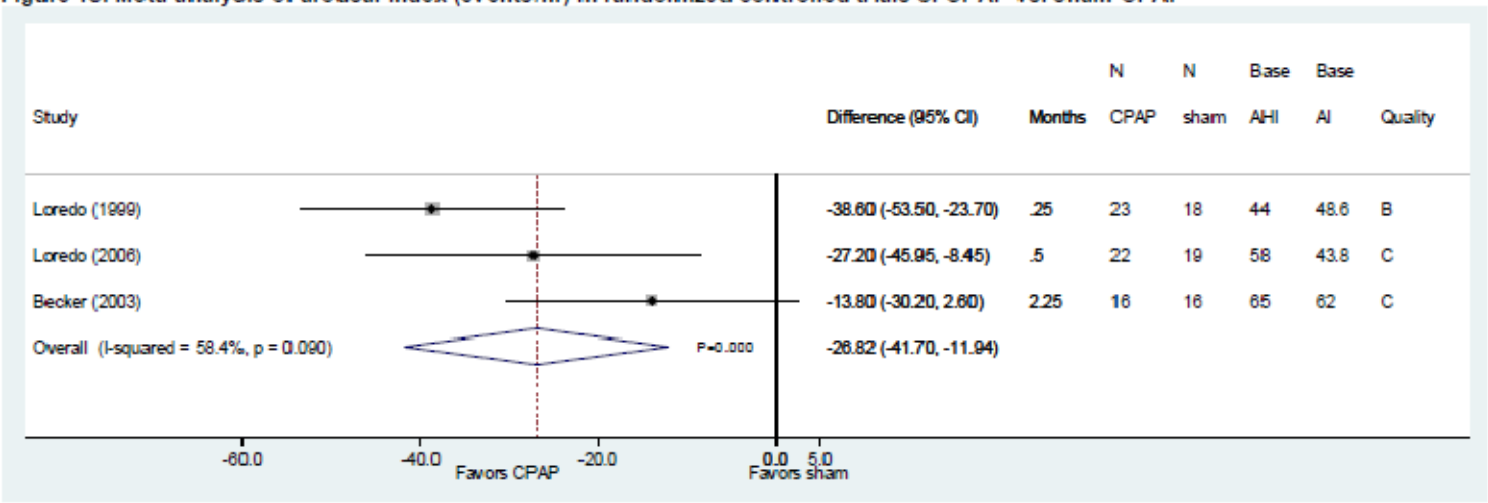
AHI = apnea hypopnea index, CI = confidence interval, CPAP = continuous positive airway pressure.

Figure 12. Meta-analysis of ESS in randomized controlled trials of CPAP vs. sham CPAP, by type of CPAP



AHI = apnea-hypopnea index, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure, ESS = Epworth Sleepiness Scale.

Figure 13. Meta-analysis of arousal index (events/hr) in randomized controlled trials of CPAP vs. sham CPAP



AHI = apnea-hypopnea index, AI = arousal index, CI = confidence interval, CPAP = continuous positive airway pressure.

Comparison of Oral and Nasal CPAP

One crossover trial[171] and one parallel trial[173] compared oral with nasal CPAP; one crossover trial[172] compared a face mask (covering both nose and mouth) with a nasal mask (Balk [AHRQ] (2011) Appendix D Table 5.3.1). Mean baseline AHI or respiratory disturbance index (RDI) in the studies were 35, 61, and 85 events/hr. Most included patients were obese; the mean body mass index (BMI) across studies ranged from 32 to 43 kg/m^2. None of the studies selectively focused on patients with other comorbidities. Study sample sizes ranged from 20 to 42 (total = 87 across studies).The duration of intervention was 1 month in two studies and 2 months in one study. One study was rated quality B and two were rated quality C. Small sample sizes and incomplete reporting were the main methodological concerns. These studies are applicable mainly to patients with AHI more than 30 events/hr and BMI more than 30 kg/m^2.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Compliance (Balk [AHRQ] (2011) Appendix D Table 5.3.2)

All three trials provided data on compliance. Mortimore 1998 reported a significant difference in compliance (hours of use per night) favoring nasal CPAP over face mask (nose and mouth) CPAP at 1 month (mean difference 1 hr/night; 95 percent CI 0.3, 1.8; P=0.01).[172] The other two studies did not find a significant difference in the number of hours of use with oral or nasal CPAP.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.3.3)

Two trials provided data on daytime sleepiness as assessed using ESS.[171,172] Anderson 2003 reported that both oral and nasal CPAP decreased daytime sleepiness, but that the difference between the two was not significant.[171] Mortimore 1998 did not provide baseline ESS data, but reported that patients in the face mask group had scored significantly higher on the ESS than those in the nasal group at followup (9.8 versus 8.2; P<0.01).

Other Outcomes

Anderson 2003 also provided outcomes on AHI, minimum oxygen saturation, arousal index, REM sleep, and sleep efficiency. The difference between oral and nasal CPAP was not statistically significant for any of these measures. Changes after 1 month within the two arms (oral versus nasal CPAP) were: -69 versus -74 events/hr for AHI; 16 versus 17 percent minimum for oxygen saturation; -54 versus -57 events/hr for arousal index; 16 versus 12 percent of total sleep time for REM sleep; and 11 versus 10 percent of total sleep time for sleep efficiency.

Study Variability

No study reported subgroup analyses with respect to the comparative effect of oral versus nasal CPAP for OSA in terms of patient characteristics (age, sex, race, weight, bed partner, and airway) or severity of OSA. The two studies that described minimum AHI or RDI enrollment criteria did not examine the same efficacy outcomes.[171,173] No conclusions could be drawn regarding indirect comparisons across studies on different patient characteristics or minimum AHI or RDI enrollment criteria.

Summary

Three small trials with inconsistent results preclude any substantive conclusions concerning the efficacy of oral 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

We found 21 RCTs that compared autoCPAP with fixed CPAP treatment in patients with OSA (Balk [AHRQ] (2011) Appendix D Table 5.4.1).[174-194] Fourteen used a crossover design and seven a parallel design. Across studies, patients' mean AHI ranged from 15 to 55 events/hr. All the studies reviewed included patients who were either overweight or obese (body mass index [BMI] ranged from 29.9 to 42 kg/m^2). None of the studies selectively focused on patients with other comorbidities. Study sample sizes ranged from 10 to 181 (total = 844 across studies). Study durations ranged from 0.75 to 9 months, the majority no longer than 3 months. One was rated quality A, 10 were rated quality B, and 10 quality C. Small sample sizes and incomplete data reporting were the main methodological concerns. These studies are applicable mainly to patients with AHI more than 15 events/hr and BMI more than 30 kg/m^2.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Compliance (Balk [AHRQ] (2011) Appendix D Table 5.4.2; Balk [AHRQ] (2011) Figure 14)

All 21 studies provided data on compliance. Seventeen studies did not find statistically significant differences in device usage (hours used per night) between autoCPAP and CPAP; four studies reported a significant increase in the use of autoCPAP compared with CPAP.[181,182,186,194] Meta-analysis revealed a statistically significant, but clinically marginal difference of 11 minutes per night favoring autoCPAP (difference = 0.19 hr; 95 percent CI 0.06, 0.33; P=0.006), without statistical heterogeneity.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.4.3; Balk [AHRQ] (2011) Figure 15)

Fourteen studies provided sufficient data on AHI after treatment.[174-180,184,186,188-190,192,193] Meta-analysis across all studies indicated a difference between autoCPAP and CPAP of 0.23 events/hr (95 percent CI -0.18, 0.64; P=0.27). The crossover and parallel design studies found similar results via meta-analysis (no significant difference by t test). No statistically significant heterogeneity was observed across studies, despite a broad range in the severity of OSA.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.4.4; Balk [AHRQ] (2011) Figure 16)

Seventeen studies provided sufficient ESS data for meta-analysis.[174,176-179,181,182,184-191,193,194] Meta-analysis across all studies yielded a difference between autoCPAP and CPAP of -0.48 (95 percent CI -0.86, -0.11; P=0.012), favoring autoCPAP. No significant difference between the study designs was shown by t- test. Despite the broad range of severity of OSA across studies, there was no statistically significant heterogeneity within the overall meta-analysis.

Arousal Index (Balk [AHRQ] (2011) Appendix D Table 5.4.5; Balk [AHRQ] (2011) Figure 17)

Seven studies provided sufficient data on arousal index after treatment.[174,176,178,184,188,190,193] Meta-analysis showed a difference of -1.09 events/hr (95 percent CI -2.4, 0.2; P=0.10), favoring autoCPAP. The summary estimates for the subgroups of studies with crossover or parallel designs were different, but neither found a statistically significant difference. Due to the large confidence intervals, no significant difference between the crossover and parallel design trials was shown (t-test, P=0.38). There was also no statistically significant heterogeneity within the overall meta-analysis as well as the subanalyses.

Minimum Oxygen Saturation (Balk [AHRQ] (2011) Appendix D Table 5.4.6; Balk [AHRQ] (2011) Figure 18)

Seven studies provided sufficient data on minimum oxygen saturation after treatment.[176-178,180,184,188,190] Meta-analysis of these trials resulted in a difference between autoCPAP and CPAP of -1.3 percent total sleep time (95 percent CI -2.2, -0.5; P=0.03), favoring CPAP. The crossover and parallel design trials had similar results. There was no statistically significant heterogeneity within the overall meta-analysis.

Sleep Efficiency (Balk [AHRQ] (2011) Appendix D Table 5.4.7)

Two studies provided data on sleep efficiency.[178,188] Both found no statistically significant difference between autoCPAP and CPAP for the improvement of sleep efficiency.

REM Sleep (Balk [AHRQ] (2011) Appendix D Table 5.4.8)

Seven studies provided data on REM sleep.[177,178,184,188,190,191,193] All but one study found no statistically significant difference between autoCPAP and CPAP for REM sleep. Only Nolan 2007 reported a greater reduction in REM sleep in patients treated with autoCPAP compared with those treated with CPAP (-0.5 versus +2 percent total sleep time; P=0.06).

Stage 3 or 4 sleep (Balk [AHRQ] (2011) Appendix D Table 5.4.9)

Six studies provided data on slow wave sleep (stage 3 or 4).[177,178,184,188,190,191] All reported no statistically significant difference between autoCPAP and CPAP for Stage 3 or 4 sleep.

Quality of Life (Balk [AHRQ] (2011) Appendix D Table 5.4.10)

Eight studies provided data on quality of life.[175,177,179,181,183,186,189,194] Seven used SF-36; one used the Sleep Apnea Quality of Life Index (SAQLI);[181] and two also added a modified Osler test.[189,194] Massie 2003 found a significant difference in the mental health (net difference 5; 95 percent CI 0.16, 9.8; P<0.05) and vitality (net difference 7; 95 percent CI 0.6, 13.4; P<0.05) components of SF-36, favoring those who had autoCPAP.[186] No other significant differences in quality of life measures between autoCPAP and CPAP were reported in the reviewed studies.

Blood Pressure (Balk [AHRQ] (2011) Appendix D Table 5.4.11)

Two studies reported changes in blood pressure.[178,180] Patruno 2007 reported significant reductions between baseline and followup in systolic and diastolic blood pressure in patients on CPAP, but not in those on autoCPAP; however, the study did not report a statistical analysis of the difference between the two interventions. Our estimates, based on the reported data, suggest a nonsignificant greater reduction in systolic blood pressure (net difference = 6 mm Hg; 95 percent CI -0.9, 12.9; P=0.09) and a significant greater reduction in diastolic blood pressure (net difference = 7.5 mm Hg; 95 percent CI 4.2, 10.8; P<0.001) with CPAP as compared to autoCPAP. Nolan 2007 reported no significant differences in systolic and diastolic blood pressure changes between autoCPAP and CPAP; however, no quantitative data were provided.

Study Variability

No study reported subgroup analyses with respect to the comparative effect of autoCPAP versus CPAP for OSA in terms of patient characteristics (age, sex, race, weight, bed partner, and airway characteristics) or severity of OSA. Only one study explicitly defined OSA.[185] Most other studies, though, provided explicit study enrollment criteria based on a minimum AHI threshold.

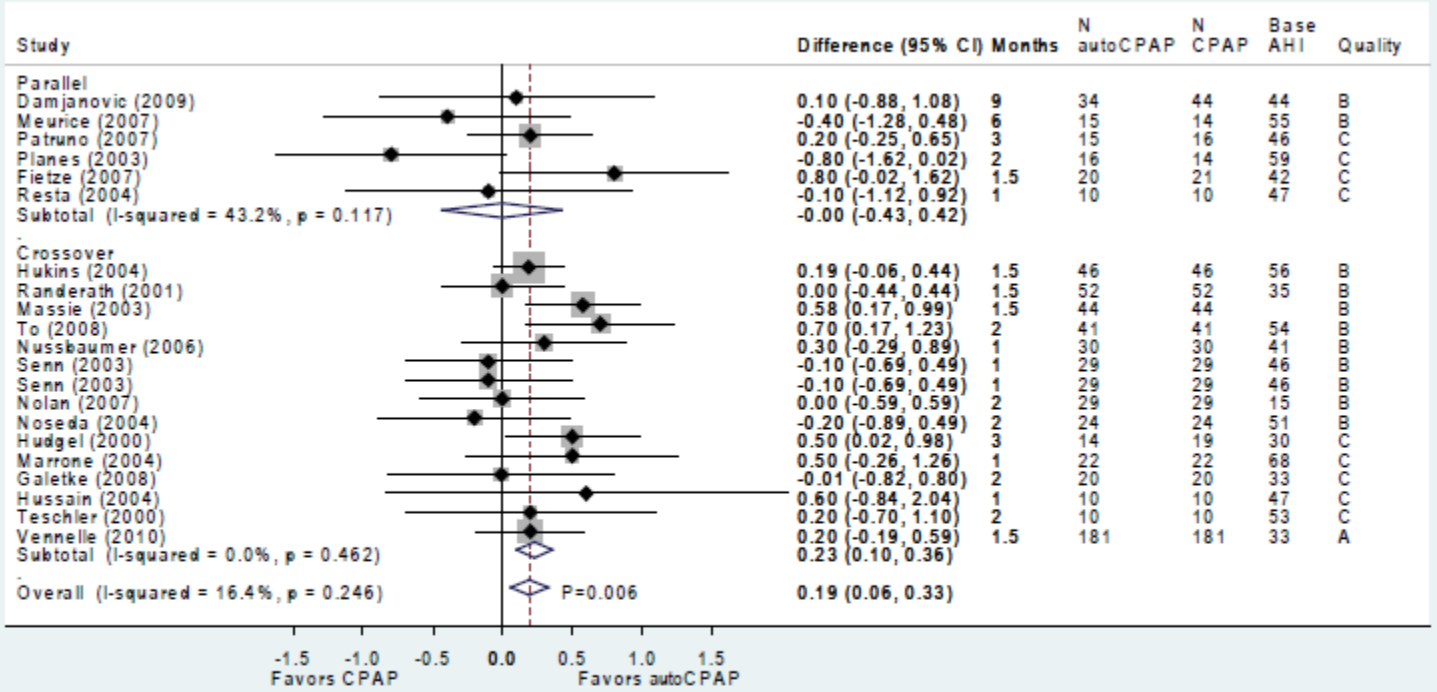
We performed subgroup meta-analyses stratified by different minimum AHI threshold for the AHI and ESS outcomes. No apparent difference in AHI outcomes was observed between autoCPAP and CPAP within any of the AHI subgroups (5, 10, 15, 20, or 30 events/hr). For the ESS, there were significant differences in favor of autoCPAP for the AHI subgroups of 20 and 30 events/hr, but not for the subgroups of studies that included patients with a lower AHI.

Summary

Twenty-one studies (mostly quality B or 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 clinically similar levels of compliance (hours used per night) 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 CPAP than with autoCPAP, although by only about 1 percent. Evidence is limited regarding the relative effect of CPAP and autoCPAP on blood pressure.

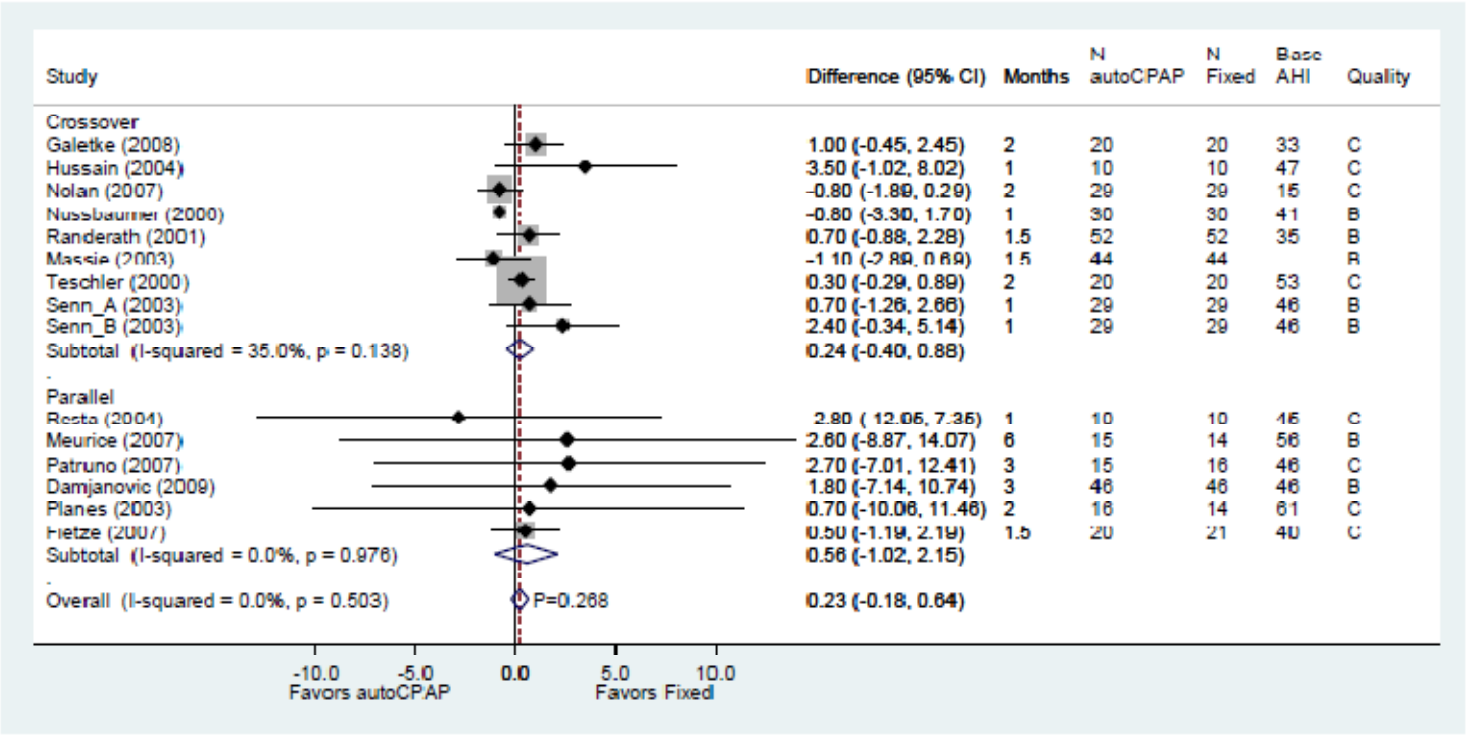
Overall, despite no or weak evidence on clinical outcomes, overall the strength of evidence is moderate that autoCPAP and fixed CPAP result in similar compliance and treatment effects for patients with OSA.

Figure 14. Meta-analysis of CPAP compliance (hr/night) in randomized controlled trials of autoCPAP vs. CPAP, by study design



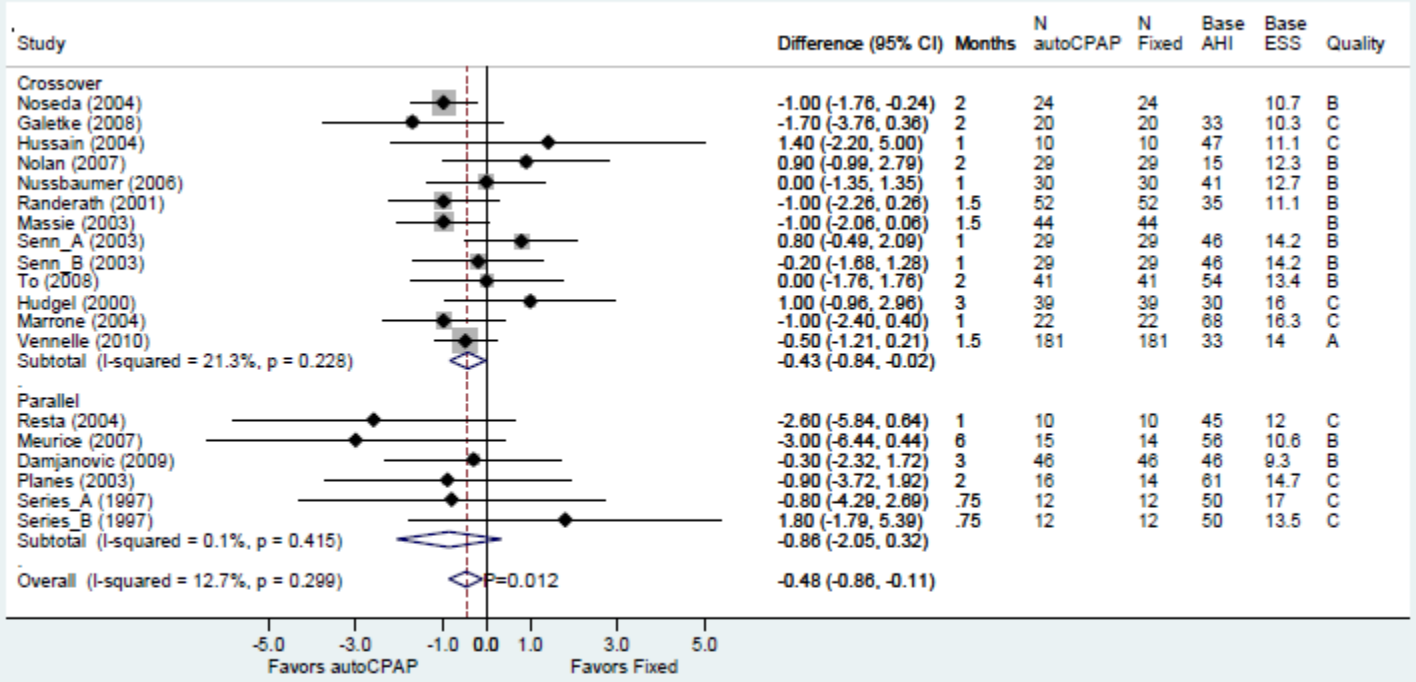
AHI = apnea-hypopnea index, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure.

Figure 15. Meta-analysis of AHI (events/hr) in randomized controlled trials of autoCPAP vs. CPAP, by study design



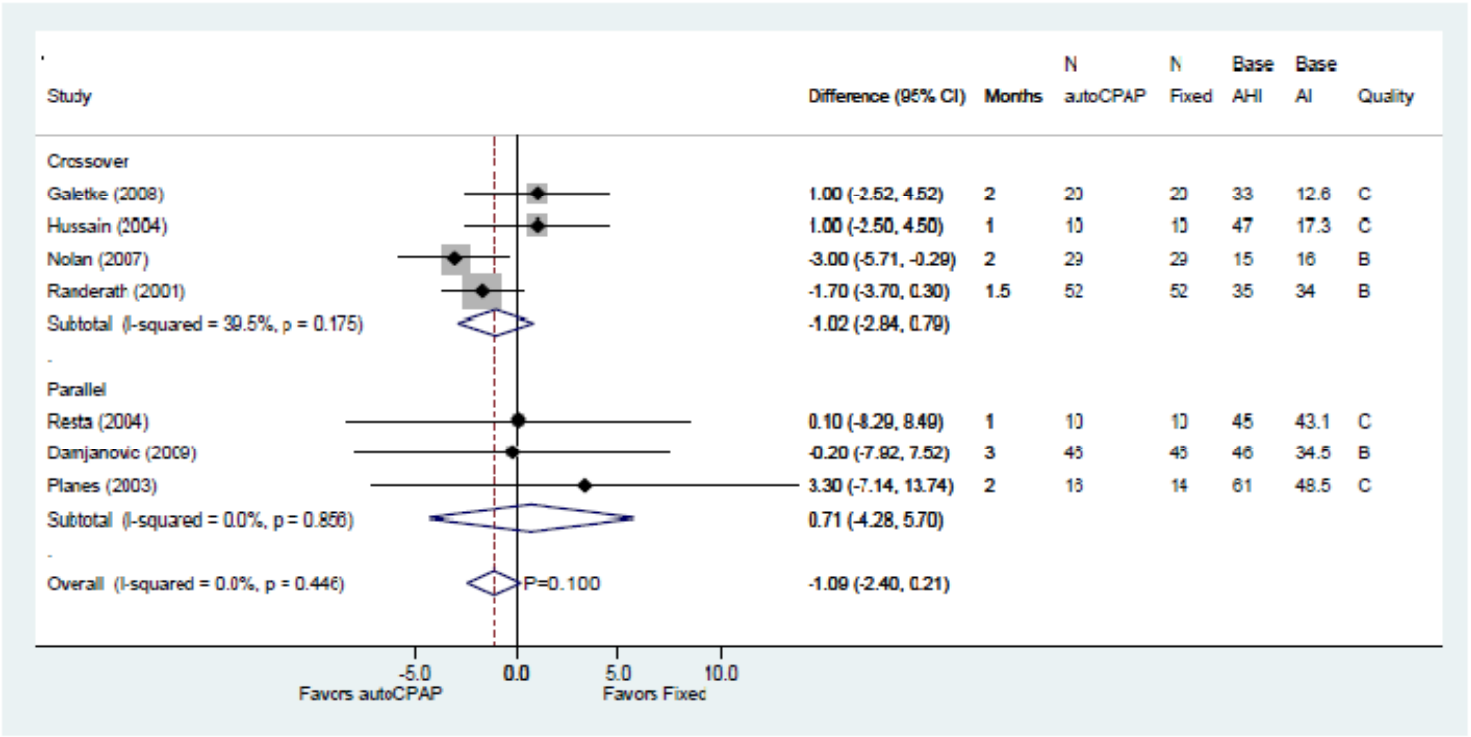
AHI = apnea-hypopnea index, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure, Fixed = fixed pressure CPAP.

Figure 16. Meta-analysis of ESS in randomized controlled trials of autoCPAP vs. CPAP, by study design



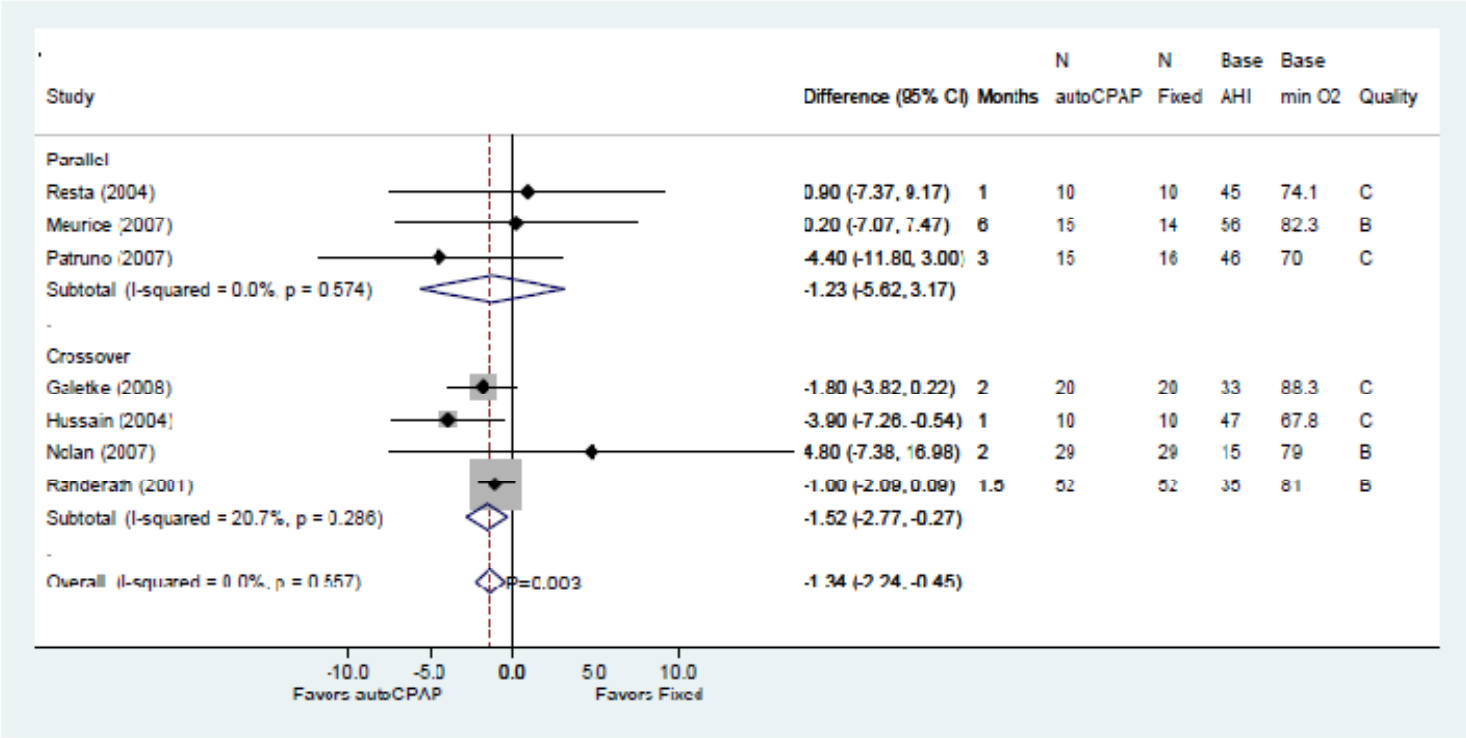
AHI = apnea-hypopnea index, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure, ESS = Epworth Sleepiness Scale, Fixed = fixed pressure CPAP.

Figure 17. Meta-analysis of arousal index (events/hr) in randomized controlled trials of autoCPAP vs. CPAP, by study design



AHI = apnea-hypopnea index, AI = arousal index, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure, Fixed = fixed pressure CPAP.

Figure 18. Meta-analysis of minimum oxygen saturation (%) in randomized controlled trials of autoCPAP vs. CPAP, by study design



AHI = apnea-hypopnea index, autoCPAP = autotitrating CPAP, CI = confidence interval, CPAP = continuous positive airway pressure, Fixed = fixed pressure CPAP, min O2 = minimum oxygen saturation

Comparison of Bilevel CPAP and Fixed CPAP

Four parallel trials compared bilevel CPAP with fixed CPAP[195-198] and one crossover trial compared bilevel CPAP with autoCPAP, in patients with OSA (Balk [AHRQ] (2011) Appendix D Table 5.5.1).[199] Baseline AHI in the four studies with reported data ranged from 32 to 52 events/hr. Piper 2008 included patients with concomitant morbid obesity (mean BMI = 53 kg/m^2) and obesity hypoventilation syndrome. Khayat 2008 included patients with concomitant heart failure (American Heart Association class II or III). Gay 2003 enrolled patients without comorbidities. About 10 percent of the patients in Reeves-Hoche 1995 had restrictive lung pattern on pulmonary function tests secondary to obesity. In the bilevel CPAP versus autoCPAP study, Randerath 2003 specifically enrolled patients who did not tolerate conventional CPAP. Study sample sizes ranged from 24 to 83 (total = 197 across studies). Study durations ranged from 1 to 12 months. One study was rated quality B[197] and the remaining four were rated quality C.[195,196,198,199] Small sample sizes and possible selection bias were the main methodological concerns. These studies are applicable mainly to patients with AHI more than 30 events/hr. Individual studies are applicable to patients with morbid obesity, heart failure, or no comorbidities. Only one study was restricted to patients who did not tolerate fixed CPAP.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Compliance (Balk [AHRQ] (2011) Appendix D Table 5.5.2)

All five trials provided data on compliance. None of them found a statistically significant difference in usage of the machine (hours used per night or percent days used) between bilevel CPAP and CPAP, or bilevel CPAP and autoCPAP, at followup. Piper 2008 and Gay 2003 reported that patients used the devices for about 6 hours a night, on average, Reeves-Hoche 1995 about 5 hours a night, and Khayat 2008 about 4 hours a night. Randerath 2003 reported that the patients used the machines about 90 percent of the time.

Apnea-Hypopnea Index

Two trials provided data on AHI outcome.[196,199] Khayat 2008 reported that both bilevel CPAP and CPAP decreased AHI after 3 months (-34 versus -26 events/hr, respectively). Randerath 2003 reported that both bilevel CPAP and autoCPAP decreased AHI after 1.5 months (-39 versus -35 events/hr, respectively). The difference between bilevel CPAP and CPAP or autoCPAP was not statistically significant in either trial.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.5.3)

Four trials provided data on changes in daytime sleepiness as assessed using ESS.[195-197,199] Each reported that both bilevel CPAP and CPAP decreased daytime sleepiness. The difference between bilevel CPAP and CPAP was not statistically significant in any trial.

Other Sleep Study Measures

Randerath 2003 also provided outcomes on minimum oxygen saturation, arousals, and sleep stages. [199] The difference between bilevel CPAP and autoCPAP was not statistically significant in any of these measures. Changes after 1.5 months within the two arms (bilevel CPAP versus autoCPAP) were 7.4 versus 9.4 percent for minimum oxygen saturation, -25.3 versus -22.5 events/hr for arousal index, -1 versus 0 percent of total sleep time for REM sleep, and 7.8 versus 4.7 percent of total sleep time for stages 3 or 4 sleep, respectively.

Quality of Life and Other Functional Outcomes (Balk [AHRQ] (2011) Appendix D Table 5.5.4)

Three trials provided data on quality of life outcomes.[195-197] Each study used a different instrument for assessment: the Minnesota Questionnaire for heart failure,[196] the Functional Outcomes of Sleep Questionnaire (FOSQ),[195] and the SF-36.[197] None of the trials found significant differences between bilevel CPAP and CPAP in any quality of life measure.

Neurocognitive and Psychological Tests

One trial reported on neurocognitive outcomes.[197] Piper 2008 found a significant difference in the ―mean of slowest 10 percent reaction‖ subtest of the Psychomotor Vigilance Test, favoring those patients who used bilevel CPAP (change from baseline: 0.32 versus 0.07 (unclear unit); P=0.03). No statistically significant difference was found in the other two subtests.

Blood Pressure

Khayat 2008, which included OSA patients with heart failure, was the only trial to report changes in blood pressure.[196] No significant differences were found between the two treatments; both bilevel CPAP and CPAP decreased systolic (6.3 versus 1.4 mm Hg, respectively; P=0.53) and diastolic blood pressure (7.5 versus 2.3 mm Hg, respectively; P=0.31).

Study Variability

No study reported subgroup analyses with respect to the comparative effect of bilevel CPAP versus CPAP for OSA in terms of patient characteristics (age, sex, race, weight, bed partner, and airway characteristics) or severity of OSA. All studies that reported a minimum AHI for inclusion eligibility used a threshold of 10 events/hr, thus no analysis of comparative effects across studies based on different AHI enrollment criteria was possible.

Summary

Five small trials with largely null findings did not support any substantive differences in the efficacy of bilevel CPAP versus CPAP in the treatment of patients with OSA. The studies were mostly of quality C but reported generally consistent results across outcomes. The studies were highly clinically heterogeneous in their populations, mostly with substantial comorbidities. Thus the studies, overall, have limited directness to the general OSA population. Largely due to small sample sizes, the studies mostly had imprecise estimates of the comparative effects. Due to the clinical heterogeneity and the imprecision, the overall strength of evidence was graded insufficient regarding any difference in compliance or other outcomes between bilevel CPAP and CPAP.

Comparison of Flexible Bilevel CPAP and Fixed CPAP

Only Ballard 2007, a quality B, parallel design RCT, compared flexible bilevel CPAP with fixed CPAP. The study enrolled 104 patients with OSA (mean AHI of 42 events/hr; mean BMI 34.2 kg/m^2) and self-estimated nightly use of CPAP of less than 4 hours.[200] After 3 months, significantly more patients had used flexible bilevel CPAP for more than 4 hours a night compared with CPAP (49 versus 28 percent, respectively; P=0.03). Mean hours used per night were similarly higher in the flexible bilevel CPAP group than the CPAP group (3.7 versus 2.9 hr/night, respectively; P<0.05) The study also reported that the patients treated with flexible bilevel CPAP displayed a significant increase in mean FOSQ total score of 1.45 (P=0.004); the increase (0.45) in the CPAP group was not significant. Statistical comparison between groups on FOSQ was not reported. By our calculation (based on the reported data) the difference

between the two treatments was not statistically significant. The study did not evaluate objective clinical outcomes. This study is applicable mainly to patients with AHI more than 30 events/hr and BMI more than 30 kg/m^2 who were poorly compliant with fixed CPAP.

In conclusion, while a single study found that flexible bilevel CPAP may yield increased compliance compared with fixed CPAP, overall the strength of evidence is insufficient regarding the relative effect of the two interventions.

Comparison of C-Flex™ and Fixed CPAP

Three parallel trials[201-203] and one crossover[204] trial compared C-Flex™ with fixed CPAP in patients with OSA (Balk [AHRQ] (2011) Appendix D Table 5.7.1). C-Flex is a proprietary CPAP technology that reduces the pressure slightly at the beginning of exhalation. Mean baseline AHI in these studies ranged from 35.4 to 53.3 events/hr. No comorbidities, with the exception of increased BMI (ranged from 31 to 34.9 kg/m^2), were reported. Study sample sizes ranged from 30 to 184 (total = 430 across studies). Study durations ranged from 1.5 to 6 months. Two studies were rated quality B and two were rated quality C. Incomplete and unclear reporting were the main methodological concerns. These studies are applicable mainly to patients with AHI more than 30 events/hr and BMI more than 30 kg/m^2.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Compliance (Balk [AHRQ] (2011) Appendix D Table 5.7.2)

All four trials provided data on compliance. One study prescreened patients for compliance before acceptance into the study; only those with 4 or more mean hours of nightly CPAP use during a one week screening were admitted.[201] None of the four trials found a statistically significant difference in the relative usage of the machines (hours used per night) at followup. Pepin 2009[203] and Nilius 2006[202] reported that patients used the machines for about 5 hours a night, on average. Dolan 2009[201] and Leidag 2008[204] reported a compliance of about 6 hours a night.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.7.3)

Three trials provided data on changes in daytime sleepiness as assessed using ESS.[201-203] Each reported that both C-Flex and CPAP decreased daytime sleepiness. The difference between C-Flex and CPAP was not statistically significant in any trial. Meta-analysis of ESS difference between C-Flex and CPAP in these three studies resulted in a difference of -0.23 (95 percent CI - 0.74, 0.27; P=0.36). No statistically significant heterogeneity was observed within the metaanalysis.

Other Sleep Study Measures

Leidag 2008 also provided outcomes on AHI, minimum oxygen saturation, arousals, and sleep stages.[204] Final values at 1.5 months between C-Flex and CPAP were not statistically significant in any of these measures: 6.2 versus 5.4 events/hr for AHI, 87.7 versus 88 percent for minimum oxygen saturation, 9.3 versus 8.9 events/hr for arousal, 19.5 versus 21.7 percent for REM sleep, and 9 versus 10.2 percent for stage 4 sleep.

Quality of Life

Pepin 2009 also provided data on quality of life outcomes.[203] With the exception of physical functioning and bodily pain in SF-36, both C-Flex and CPAP improved all domains in SF-36 and in the Grenoble Sleep

Apnea Quality of Life questionnaire. No significant differences between C-Flex and CPAP were shown in these assessments.

Study Variability

No study reported subgroup analyses with respect to the comparative effect of C-Flex versus CPAP for OSA in terms of patient characteristics (age, sex, race, weight, bed partner, and airway) or severity of OSA.

All the studies used AHI as either part of the definition of OSA or as a minimum study enrollment criterion. The AHI cutoffs used were 5,[204] 10,[201] 15,[203] or 20[202] events/hr. No apparent difference in ESS was noted between C-Flex and CPAP based on different minimum AHI enrollment criteria or OSA definitions across the three studies that provided these data.[201-203]

Summary

Four trials with largely null findings did not support any substantive differences in the efficacy of C-Flex versus fixed CPAP in improving compliance (hours used per night) in patients with OSA. Overall the studies were of quality B and C, but reported generally consistent results across outcomes and had no substantive issues regarding directness to the OSA population. No statistically significant differences in compliance or other outcomes were found between C-Flex and fixed CPAP. The strength of evidence for this finding is rated low because of the mixed quality (Bs and Cs) of the primary studies.

Comparison of Humidification in CPAP

Three parallel trials[205-207] and two crossover trials[208,209] compared different aspects of humidification in fixed CPAP or autoCPAP (Balk [AHRQ] (2011) Appendix D Table 5.8.1). Three trials compared heated humidified CPAP with dry CPAP.[205,208,209] One trial provided additional data on cold passover humidified CPAP compared with dry CPAP.[209] One trial compared heated-humidified autoCPAP with dry autoCPAP.[207] One trial compared —always on‖ with —as needed‖ heatedhumidified CPAP.[206] Mean baseline AHI in these studies ranged from 29 to 54 events/hr. No comorbidities, with the exception of increased BMI (ranging from 34.4 to 37.6 kg/m^2), were reported in these studies. Study sample sizes ranged from 42 to 123 (total = 360 across studies). Study durations ranged from 0.75 to 12 months. Three studies were rated quality B and two were rated quality C. Incomplete reporting and unclear analysis were some of the methodological concerns. These studies are applicable mainly to patients with AHI more than 30 events/hr and BMI more than 30 kg/m^2.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Compliance (Balk [AHRQ] (2011) Appendix D Table 5.8.2)

All five trials provided data on compliance (hours used per night). Two trials reported that patients who used heated-humidified CPAP had increased compliance compared with those who did not (5.7 versus 5.3 hr/night, P=0.03 in Neill 2003; 5.52 versus 4.93 hr/night, P=0.008 in Massie 1999). Ryan 2009 did not find a statistically significant difference in compliance between heated-humidified and dry CPAP. Mador 2005 did not find a statistically significant difference in compliance between —always on‖ and —as needed‖ heated-humidified CPAP. Massie 1999 did not find a statistically significant difference in compliance between cold passover CPAP and heated-humidified or dry CPAP. Salgado 2008 did not find a statistically significant difference in compliance between heated-humidified and dry autoCPAP. No consistent effect of humidification on compliance was observed across these studies.

Apnea-Hypopnea Index

Only Salgado 2008 provided outcomes on AHI.[207] Both heated-humidified and dry autoCPAP were effective in reducing AHI; there was no statistically significant difference between them (-23.5 versus -24.1 events/hr, respectively).

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.8.3)

All five trials provided data on changes in daytime sleepiness as assessed using ESS. The difference between the two intervention arms in each of the trials was not statistically significant. Both intervention arms in each trial reported decreased daytime sleepiness. Three trials were sufficiently similar and provided appropriate data to allow meta-analysis.[205,207,208] A metaanalysis showed the difference in ESS between CPAP with and without humidification in these 3 trials to be -0.31 (95 percent CI -1.16, 0.54; P=0.47). No statistically significant heterogeneity was observed within the meta-analysis.

Quality of Life

Two trials provided data on quality of life outcomes.[205,206] Ryan 2009 did not find any statistically significant difference in SF-36 between patients who had heated-humidified CPAP and those who had dry CPAP.[205] However, nasal symptoms were more common in the dry CPAP group compared with the heated humidified group (70 versus 28 percent, P=0.002). Mador 2005 did not find any statistically significant difference in Calgary Sleep Apnea Quality of Life Index between patients who had —always on‖ and those who had —as needed‖ heated-humidified CPAP.[206]

Study Variability

No study reported subgroup analyses with respect to the comparative effect of humidified versus dry autoCPAP or CPAP for OSA in terms of patient characteristics (age, sex, race, weight, bed partner, and airway characteristics) or severity of OSA.

For variability in minimum AHI or RDI enrollment criteria, three studies used 10 events/hr and two studies did not specify a minimum value. No cross study comparisons based on minimum AHI or RDI criteria were possible.

Summary

Five trials examined different aspects of humidified positive airway pressure treatment for patients with OSA. While some studies reported a benefit of added humidity in positive airway pressure treatment in improving patient compliance, this effect was not consistent across all the studies. Overall the studies were clinically heterogeneous, small, and not of quality A. Thus, 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.

Comparison of Mandibular Advancement Devices and No Treatment

Five trials compared mandibular advancement devices (MAD) with different controls (Balk [AHRQ] (2011) Appendix D Table 5.9.1). Three were crossover trials[140,210,211] and two had a parallel design.[129,212] All devices were designed to advance the mandible or otherwise mechanically splint the oropharynx during sleep.

Bloch 2000 compared a one-piece MAD or a two-piece MAD with no treatment.[210] Barnes 2004 compared MAD with a placebo tablet.[140] Kato 2000 compared oral appliances of 2 mm, 4 mm, or 6 mm with no treatment.[211] Lam 2007 compared MAD plus conservative management to conservative management alone,[129] and Petri 2008 compared MAD to no treatment.[212]

The mean AHI at baseline ranged from 19 to 34 events/hr. Common exclusion criteria included significant coexisting diseases such as heart disease and diabetes, an unsafe level of sleepiness, and other upper airway or jaw problems. Study sample sizes ranged from 24 to 80, with a total of 301 patients across studies. Kato 2000 did not provide clear outcome reporting, and was rated quality C; all other studies were rated quality B. The main methodological concerns were small sample sizes and lack of blinding in outcome assessors. The studies are generally applicable to patients with AHI ≥15 events/hr, though less so to patients with comorbidities or excessive sleepiness.

While acknowledging the large clinical heterogeneity due to the different devices being tested, data were examined via meta-analyses. Note that the meta-analysis figures include comparisons of MAD with both no treatment and sham MAD (discussed in the next section).

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.9.2) and Balk [AHRQ] (2011) Oxygen Desaturation Index (Figure 19)

Four trials with five device comparisons provided data on AHI as an outcome,[129,140,210,212] while one provided data on ODI.[211] All four trials reporting on AHI reported that AHI decreased significantly more in patients using a MAD compared with controls, with net differences ranging from -6.3 to -14.7 events/hr. Kato 2000 found that ODI decreased significantly in the MAD groups compared with control, with net differences of -8.7 for the 2 mm group, -11.3 for the 4 mm group, and -15.2 for the 6 mm group ($P<0.05$ for each comparison). Meta-analysis of AHI yielded a statistically significant effect (difference = -11 events/hr; 95 percent CI -15, -8), though with some statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices combined yielded similar results (though without statistical heterogeneity).

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.9.3; Balk [AHRQ] (2011) Figure 20)

Four trials with five device comparisons provided data on ESS as an outcome.[129,140,210,212] All four trials reported that ESS was significantly improved in patients using a MAD compared to controls, with differences ranging from -1 to -4.5. Meta-analysis of ESS yielded a statistically significant effect (difference = -1.2; 95 percent CI -1.7, -0.6), without statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices treatment combined yielded similar results.

Other Sleep Study Measures (Balk [AHRQ] (2011) Appendix D Table 5.9.4a-e; Balk [AHRQ] (2011) Figures 21 & 22)

Three trials reported on minimum oxygen saturation.[129,140,211] Kato 2000 and Barnes 2004 found a significantly higher minimum oxygen saturation in the MAD group, while Lam 2007 did not find a significant difference between any of the three MADs examined and the control group. Kato 2000 found that minimum oxygen saturation increased significantly in the MAD groups compared with control, with net differences of 2.0 percent for the 2 mm group, 2.3 percent for the 4 mm group, and 2.4 percent for

the 6 mm group (P<0.05 for each comparison). Barnes 2004 found a difference of 2.4 percent (95 percent CI 1.4, 3.4; P=0.001). Meta-analysis of minimum oxygen saturation yielded a statistically significant effect (difference = 3.0 percent; 95 percent CI 0.4, 5.5), without statistical heterogeneity (Appendix D Table 5.9.4a). Meta-analysis of MAD versus no treatment or inactive devices treatment combined yielded similar results (Figure 21).

Three trials reported on arousal index (Appendix D Table 5.9.4b). Barnes 2004 found no significant difference between MAD and control; Lam 2007 found a lower arousal index in MAD compared with control (difference = -8.2; 95 percent CI -9.3, 7.1; P<0.05). Bloch 2000 found a lower arousal index for patients using a one-piece MAD compared with control (difference = -14.5; 95 percent CI -22.6, -6.4; P<0.05), but no significant difference between patients using a two-piece MAD compared with control (difference = -10.1; 95 percent CI -17.9, 2.3; NS). Bloch 2000 and Barnes 2004 reported on sleep efficiency, and found no significant difference between groups (Appendix D Table 5.9.4c) Meta-analysis of arousal index yielded a statistically significant effect (difference = -7.9 events/hr; 95 percent CI -14, -1.3), though with some statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices treatment combined yielded similar results (Figure 22).

Bloch 2000 and Barnes 2004 both reported on slow wave sleep. Barnes 2000 found no significant difference between MAD and control. Bloch 2000 found no difference between groups when comparing 2-piece MAD with control, but found a higher percentage of slow wave sleep in the 1-piece MAD group compared with control (P<0.05). Three studies reported on REM sleep; no significant difference among groups was reported (Appendix D Table 5.9.4e).[140,210,212]

Quality of Life (Balk [AHRQ] (2011) Appendix D Table 5.9.5)

Barnes 2004 reported on SF-36, finding no significant difference between MAD and control in SF-36 mean score, physical component summary, or mental component summary. Lam 2007 found no difference in any SF-36 domain between MAD and control. Barnes 2004 also reported no significant difference between groups in Beck Depression Inventory score or Functional Outcomes of Sleep Questionnaire social domain score. Lam 2007 did not find a difference in Sleep Apnea Quality of Life Index (SAQLI) social interactions or treatment-related symptoms scores, but did find an improved overall SAQLI score in the MAD group compared with the control group (difference = 0.7; 95 percent CI 0.6, 0.8; P<0.001).

Neurocognitive Tests

Jokic 1999 did not find a significant difference between CPAP and positional therapy in the Wechsler Memory Scale, Purdue Pegboard, Trail-Making Test, Symbol Digit Modalities, Consonant Trigram, or Concentration Endurance Test scores.

Study Variability

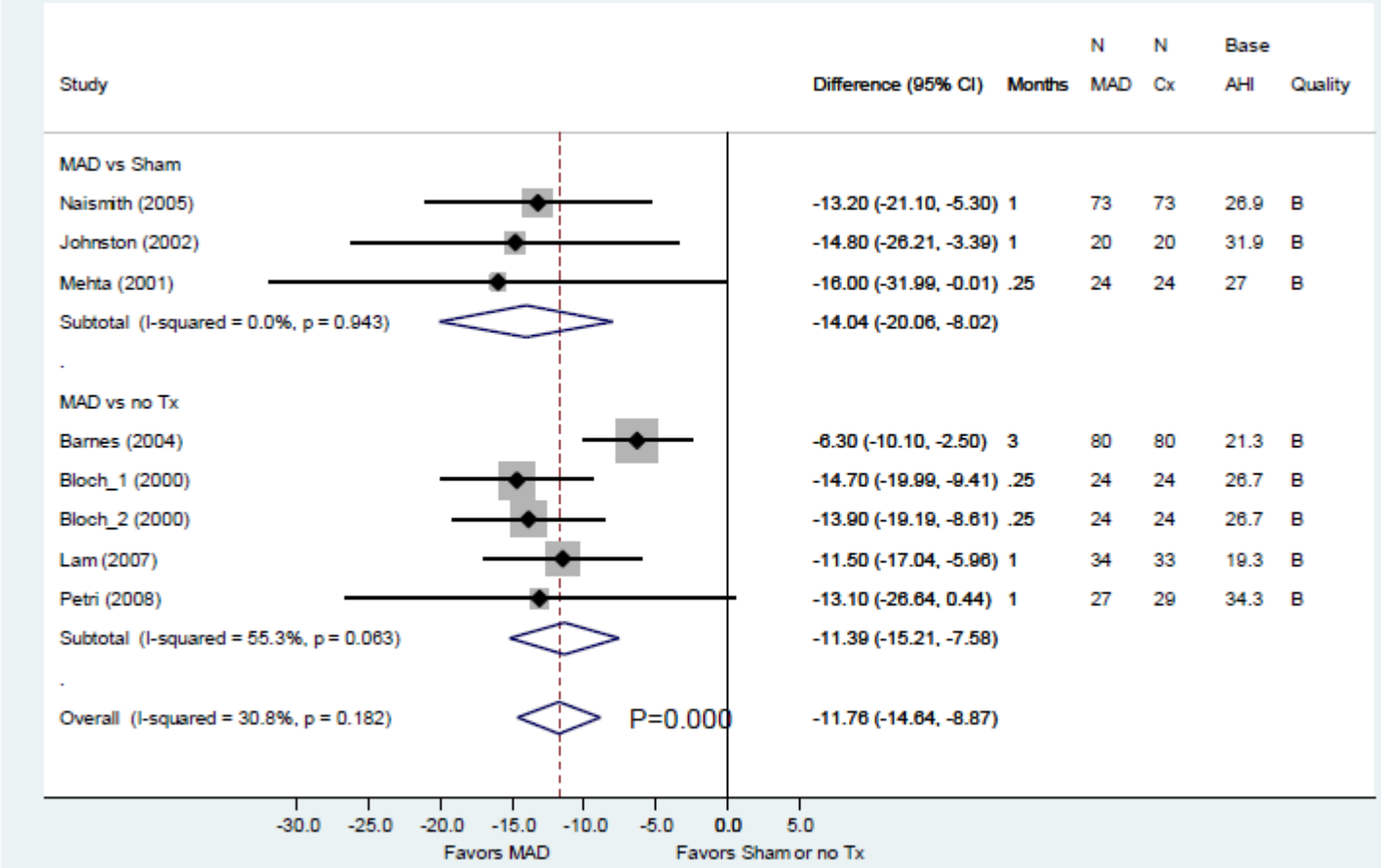
No studies reported subgroup analyses. Control treatments varied by study; Bloch 2000, Kato 2000, and Petri 2008 used no treatment as a control, whereas Barnes 2004 used a placebo tablet, and Lam 2007 used conservative management. Studies were mostly consistent in their findings.

Summary

Four quality B trials and one quality C trial compared MAD to no treatment, using a variety of different types of MAD. Individually and by meta-analysis, studies found significant improvements with MAD in AHI, ESS, and other sleep study measures. No trial evaluated longterm objective clinical outcomes. The results

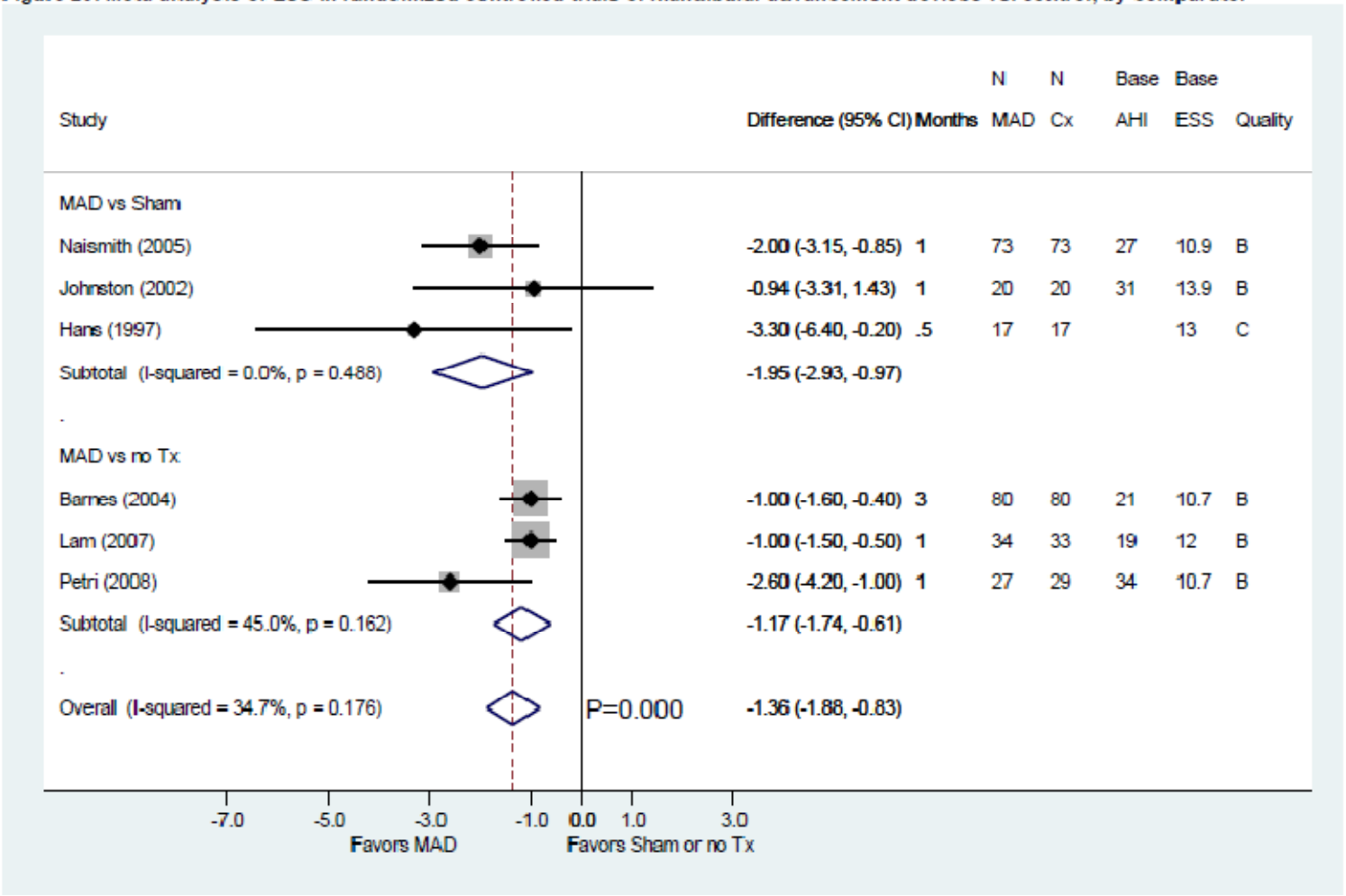
of quality of life measures, and neurocognitive tests were equivocal between groups. Overall, despite no or weak evidence on clinical outcomes, overall, the strength of evidence is moderate to show that the use of MAD improves sleep apnea signs and symptoms.

Figure 19. Meta-analysis of AHI (events/hr) in randomized controlled trials of mandibular advancement devices vs. control, by comparator



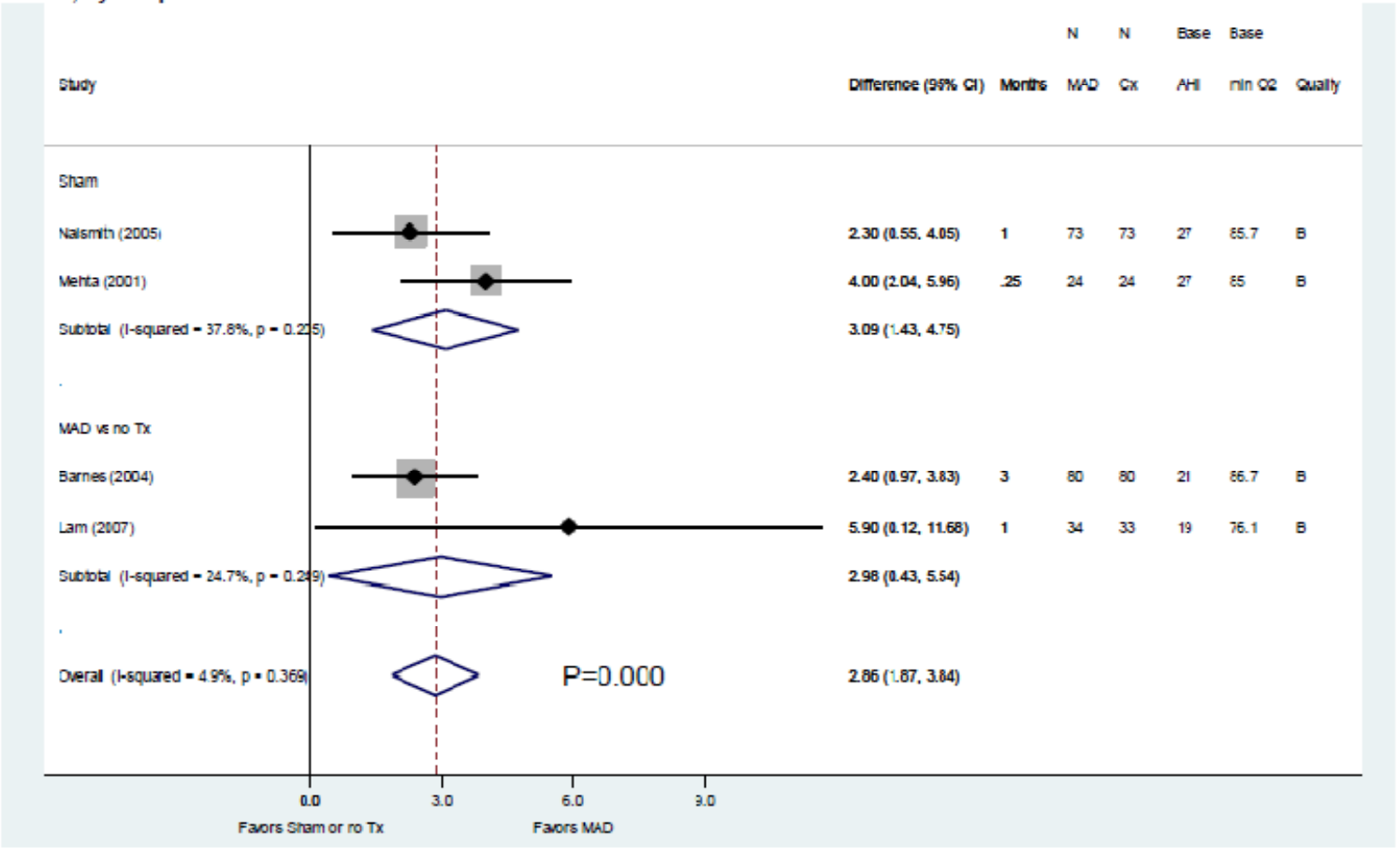
AHI = apnea-hypopnea index, CI = confidence interval, Cx = control, MAD = mandibular advancement device, Tx = treatment.

Figure 20. Meta-analysis of ESS in randomized controlled trials of mandibular advancement devices vs. control, by comparator



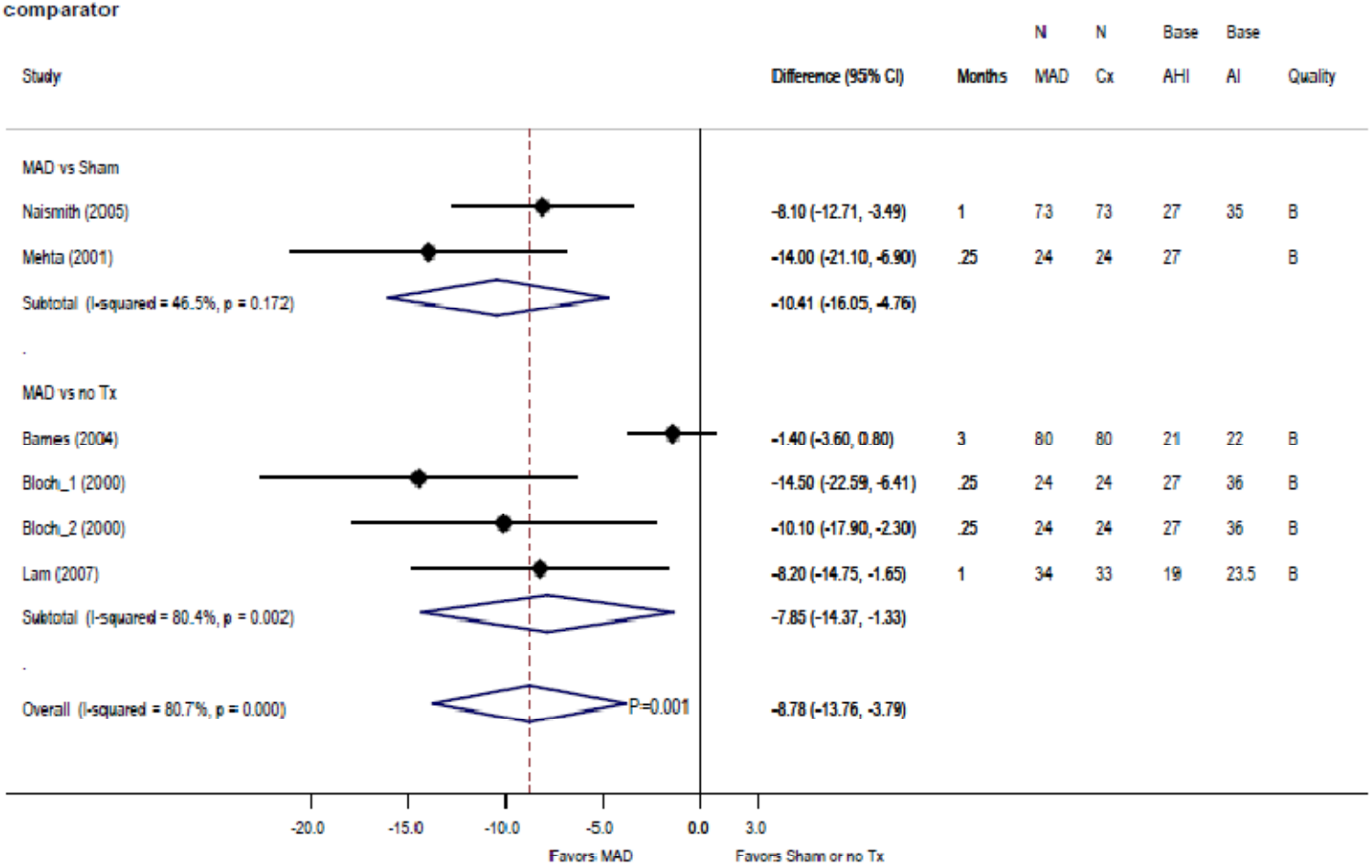
AHI = apnea-hypopnea index, CI = confidence interval, Cx = control, ESS = Epworth Sleepiness Scale, MAD = mandibular advancement device, Tx = treatment.

Figure 21. Meta-analysis of minimum oxygen saturation (%) in randomized controlled trials of mandibular advancement devices vs. control, by comparator



AHI = apnea-hypopnea index, CI = confidence interval, Cx = control, MAD = mandibular advancement device, min O2 = minimum oxygen saturation, Tx = treatment.

Figure 22. Meta-analysis of arousal index (events/hr) in randomized controlled trials of mandibular advancement devices vs. control, by comparator



AHI = apnea-hypopnea index, AI = arousal index, CI = confidence interval, Cx = control, MAD = mandibular advancement device, Tx = treatment.

Comparison of Mandibular Advancement Devices and Inactive (Sham) Oral Devices

One parallel design RCT[212] and four crossover trials compared the effects of MADs to inactive oral devices with no mandibular advancement across six publications (Balk [AHRQ] (2011) Appendix D Table 5.10.1).[213-218] The baseline AHI (or RDI) ranged from 25 to 36 events/hr. All studies included patients with no other significant comorbidities. Study sample sizes ranged from 17 to 73 (total = 186 patients). Study durations ranged from 8 days to 6 weeks. Hans 1997 was rated quality C due to a 30 percent dropout rate and the lack of a power analysis. The other studies were rated quality B. These studies are applicable primarily to patients with AHI of more than about 25 events/hr who do not have other significant comorbidities. All studies excluded edentulous patients or those with periodontal diseases.

While acknowledging the large clinical heterogeneity due to the different devices being tested, data were examined via meta-analyses.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Apnea-Hypopnea Index and Respiratory Disturbance Index (Balk [AHRQ] (2011) Appendix D Table 5.10.2; Balk [AHRQ] (2011) Figure 19)

Five studies provided data on AHI or RDI.[212-218] All found significant improvement in AHI or RDI with MAD compared with sham devices. Net changes in AHI or RDI ranged from -13 to -25 events/hr. Meta-analysis of AHI yielded a statistically significant effect (difference = -14 events/hr; 95 percent CI -20, -8), without statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices combined yielded similar results (see *Comparison of Mandibular Advancement Devices and No Treatment*, above).

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.10.3; Balk [AHRQ] (2011) Figure 20)

Four trials in six publications provided data on changes in daytime sleepiness assessed using ESS.[212-217] Gotsopoulos 2002, in a 4 week crossover trial with 73 patients (mean age = 48 yr, 80 percent male) compared a custom-made MAD to an inactive oral device (single upper plate). It found a statistically significant reduction in daytime sleepiness with MAD compared with the inactive oral device (net change in ESS -2; P<0.001). The other studies did not find statistically significant differences between MAD and inactive oral devices. Meta-analysis of ESS yielded a statistically significant effect (difference = -1.9; 95 percent CI -2.9, -1.0), without statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices treatment combined yielded similar results (see *Comparison of Mandibular Advancement Devices and No Treatment*, above).

Other Sleep Study Measures (Balk [AHRQ] (2011) Appendix D Tables 5.10.4-5.10.7; Balk [AHRQ] (2011) Figures 21 & 22)

Two trials in four publications reported changes in minimum oxygen saturation (Appendix D Table 5.10.4).[214,215,217,218] Gotsopoulos 2002 (and associated articles) and Mehta 2001 compared MAD with inactive oral devices (single upper plate and lower dental plate) and found statistically significant improvements in minimum oxygen saturation with MAD compared with the inactive oral devices (differences of 6 and 2 percent, respectively; P<0.0001). Meta-analysis of minimum oxygen saturation yielded a statistically significant effect (difference = 3.1 percent; 95 percent CI 1.4, 4.8), without statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices treatment combined yielded similar results (see Figure 21 and *Comparison of Mandibular Advancement Devices and No Treatment*, above).

The same trials also reported changes in the number of arousals (Appendix D Table 5.10.5).[214,215,217,218] Naismith 2005 found a significant decrease in the number of arousals in MAD compared with single plate devices (P<0.001). Meta-analysis of arousal index yielded a statistically significant effect (difference = -10 events/hr; 95 percent CI -16, -5; P=0.001), without statistical heterogeneity. Meta-analysis of MAD versus no treatment or inactive devices treatment combined yielded similar results (see Figure 22 and *Comparison of Mandibular Advancement Devices and No Treatment*, above).

Two trials in four publications reported sleep efficiency (Appendix D Table 5.10.6).[214,215,217,218] Neither trial found statistically significant differences in sleep efficiency between MAD and sham devices.

Two trials evaluated changes in sleep stages with MAD compared with sham devices.[212,218] The outcome of interest was the percentage of total sleep time spent in REM, stage 3 and stage 4 sleep. Mehta 2001 found a significant improvement in REM sleep with MAD compared with lower dental plate (P<0.005). Petri 2008 provided sufficient data for comparative calculations between MAD versus an appliance with no mandibular advancement. Our calculations showed a significant increase in percentage of total sleep time spent in stage 3 sleep with MAD compared with nonadvancement (sham) MAD (net difference 2.9 percent; P=0.045). However there was no significant difference in time spent in REM or stage 4 sleep between groups (Appendix D Table 5.10.7).

Quality of Life (Balk [AHRQ] (2011) Appendix D Table 5.10.9)

Petri 2008 reported quality of life outcomes measured using SF-36.[212] The study found significant improvement in the vitality dimension with MAD compared with sham MAD (P<0.001). There were no statistically significance differences between groups in other domains of SF-36.

Neurocognitive Tests (Balk [AHRQ] (2011) Appendix D Table 5.10.8)

Gotsopoulos 2002 (and related articles), which compared MAD with an inactive oral appliance (single upper plate) with no mandibular advancement, found significant improvements in the Multiple Sleep Latency Test (P=0.01) with MAD as compared with single upper plate. In addition, the study found significant improvements in somatic items on the Beck Depression Inventory scale (P<0.05) and the choice reaction time task (a speed & vigilance test) on the neuropsychological test (P<0.001) in MAD compared with single upper plate. There were no statistically significant differences between groups in other domains of Beck Depression Inventory or other subtests of the neuropsychological tests.

Neurocognitive Tests (Balk [AHRQ] (2011) Appendix D Table 5.10.8)

Gotsopoulos 2002 (and related articles) also found significant reductions in 24-hour systolic (P<0.05) and diastolic (P<0.001) blood pressures with MAD as compared with single upper plate.

Study Variability

None of these studies reported subgroup analyses. The studies were generally consistent in their findings and there were no clear differences in effect based on patient characteristics, severity of sleep apnea, other symptoms, or apparent differences in OSA definitions.

Summary

Five trials, most rated quality B, compared the effects of MAD with inactive devices. The studies individually and via meta-analysis showed sufficient evidence that most sleep study measures (AHI, minimum oxygen saturation, arousal index) and ESS were improved with MAD as compared with devices without mandibular advancement. No trials evaluated objective clinical outcomes. The strength of

evidence is insufficient concerning other evaluated outcomes due to inconsistent results or a limited number of studies per outcome. Overall, despite no or weak evidence on clinical outcomes, overall the strength of evidence is moderate to show that the use of MAD improves sleep apnea signs and symptoms.

Comparisons of Different Oral Devices

Two parallel design RCTs[219,220] and one crossover trial[221] compared the effects of different types of oral MAD in patients with OSA (Balk [AHRQ] (2011) Appendix D Table 5.11.1). A fourth study compared MAD with a novel tongue-stabilizing device[222] and a fifth study compared two types of tongueretaining devices.[223] This study was rated quality C due to inadequate methodology and poor statistical analysis reporting. The other studies were rated quality B. These studies are applicable mostly to patients with AHI of 15 to 30 events/hr and BMI less than 30 kg/m^2. However, one study included mainly obese patients (BMI >30 kg/m^2). All studies were restricted to patients with sufficient number of teeth to anchor the mandibular devices in place. Each study examined a unique comparison, and, as such, is presented separately.

Different Degrees of Mandibular Advancement (Balk [AHRQ] (2011) Appendix D Tables 5.11.1, 5.11.2, 5.11.4)

Walker-Engstrom 2003 compared the same MAD at different degrees of mandibular advancement: 75 percent (mean mandibular advancement 7.2 mm) versus 50 percent (mean of 5.0 mm). The trial enrolled 84 male patients, mostly obese (BMI >30 kg/m^2), with severe OSA (AHI ≥20 events/hr). The mean age was 50 years, mean baseline AHI was 50 events/hr, and mean ESS score was 11.5. After 6 months, AHI normalization (AHI<10 events/hr) was achieved by 52 percent of patients who had 75 percent mandibular advancement and 31 percent of patients with 50 percent advancement (P<0.04). However, the trial found no difference in mean AHI or ESS score between groups.

Self-Adjustment Versus Objective Adjustment of Devices (Balk [AHRQ] (2011) Appendix D Tables 5.11.1-5.11.5)

Campbell 2009 compared two methods of adjustment of the same MAD. One group of patients used self-adjustment of the MAD during the entire study duration. The other had an "objective adjustment" at 3 weeks following PSG-based feedback. The trial included 28, predominately male patients who had a BMI ≤35 kg/m^2. The mean age was 48 years, mean baseline AHI was 25 events/hr, and baseline ESS score was 11.6. At 6 weeks, the trial found no statistically significant difference between the two groups in the sleep study measures evaluated (AHI and minimum oxygen saturation) and no statistically significant difference in subjective sleepiness.

Custom-Made Versus Thermoplastic Devices (Balk [AHRQ] (2011) Appendix D Tables 5.11.1, 5.11.2, 5.11.4, 5.11.5)

Vanderveken 2008 compared custom-made MAD with a premolded thermoplastic MAD. Twenty-three, predominately male patients were evaluated in a crossover study with a 1 month washout period. The mean age was 49 years, mean baseline AHI was 13 events/hr, and baseline ESS score was 8. No statistically significant differences between treatment groups in AHI, sleep efficiency, ESS, or minimum oxygen saturation were found.

Mandibular Advancement Devices Versus Tongue-Stabilizing Device (Balk [AHRQ] (2011) Appendix D Tables 5.11.1 & 5.11.6)

Deane 2009 reported a 1 week crossover trial that compared the effects of a MAD with a novel tongue-stabilizing device.[222] The MAD produced 75 percent of maximal mandibular protrusion with a 4 mm vertical interincisal opening. The tongue-stabilizing device was a nonadjustable tongue-suction device made of silicone with no mandibular advancement. This study included 22 patients (73 percent male) with an AHI ≥10 events/hr (AHI mean 27 events/hr) and no other comorbid states. The mean age of patients was 49 yr and the mean BMI was 29 kg/m^2. The study reported that 91 percent of the subjects using MAD had a decrease in AHI compared with 77 percent of patients using a tongue-stabilizing device. Our calculations based on the reported data show no significant differences in mean AHI, minimum oxygen saturation, arousal index, or sleep efficiency between groups.

Tongue-Retaining Device With Versus Without Suction (Balk [AHRQ] (2011) Appendix D Tables 5.11.1 & 5.11.7)

Dort 2008 reported a crossover trial that compared the effects of a tongue-retaining device with or without suction in 32 patients (69 percent male) with primary snoring (RDI <5 events/hr) or mild to moderate OSA (RDI <30 events/hr).[223] The active suction device was designed to allow suction formation on the tip of the tongue when placed in the mouth. After 1 week with each device (and a 1 week washout period), a significant improvement in RDI was observed with the suction tongue-retaining device as compared with the nonsuction device (difference = -4.9 events/hr; 95 percent CI -8.9, -0.85; P=0.019). However, there were no significant differences in ESS score, SAQLI, and compliance (mean hours of use per night) between groups.

Study Variability

A subgroup analysis comparing males and females in Deane 2009 showed no difference between MAD and a tongue-stabilizing device. As there is only one study per comparison, we were unable to assess potential differences due to factors of interest, such as patient characteristics and severity of OSA.

Summary

Five studies 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 groups.

As the reviewed studies were generally small, and each concerned with a unique comparison, the strength of evidence is insufficient to draw conclusions with regards to the relative efficacy of different types of oral MAD in patients with OSA.

Comparison of Mandibular Advancement Devices and CPAP

Ten trials (11 articles) compared different MAD with CPAP (Balk [AHRQ] (2011) Appendix D Table 5.12.1). Seven were crossover trials[140,224-229] and three had a parallel design.[129,230-232] All devices were designed to advance the mandible or otherwise mechanically splint the oropharynx during sleep.

Five trials tested branded oral devices, four used custom-made oral devices, and one (Skinner 2004) used a cervicomandibular support collar, which was worn around the neck and shoulders. This latter device was compared with autoCPAP; all other devices were compared with fixed (or undefined) CPAP.

Mean baseline AHI in the reviewed trials ranged from 18 to 40 events/hr; four trials included patients with an AHI ≥5, three with an AHI ≥10, one with an AHI ≥15 events/hr, and two did not report a lower AHI threshold. Four trials included only patients with relatively less severe OSA, with an AHI <30-50 events/hr. Most trials had otherwise unrestricted eligibility criteria with the exception of Barnes 2004, which excluded patients with diabetes, and Tan 2002, which excluded patients with recent cardiovascular disease. The sample size of the studies ranged from 10 to 94 (total = 384 across studies). The smallest study, Skinner 2004, was stopped early after analyzing half the planned participants because of significant results favoring the control (autoCPAP). This study was rated quality C; the remaining studies were all rated quality B. Small sample sizes, lack of outcome assessor blinding, and incomplete reporting were the main methodological concerns. The studies are generally applicable to patients with AHI >5-10 events/hr.

While acknowledging the large clinical heterogeneity due to the different devices being tested, data were examined via meta-analyses.

Objective Clinical Outcomes
No study evaluated objective clinical outcomes.

Compliance
Only Gagnadoux 2009, in a crossover trial of 28 patients, assessed compliance. Patients reported statistically significantly more hours of use per night (7.0 versus 6.0 hr/night; P<0.01) and more nights of use (98 versus 90 percent of nights; P>0.01) with MAD as compared with CPAP.

Treatment Response (Balk [AHRQ] (2011) Appendix D Table 5.12.2)
Two studies measured treatment response (as a dichotomous outcome). In a 2 month crossover study of 28 patients, Gagnadoux 2009 found that significantly more patients on CPAP, as compared with MAD, had a complete response, defined as a ≥50 percent reduction in AHI to <5 events/hr (risk difference = -29 percent; 95 percent CI -53, -4; P=0.02). However, the large majority of the remaining patients experienced a partial response (≥50 percent reduction in AHI to >5 events/hr) and thus no significant difference in combined or partial response was observed.

Hoekema 2008 evaluated several related outcomes in a 2-3 month parallel trial of 103 patients comparing CPAP with oral appliance. ―Effective treatment‖, defined as a final AHI <5 events/hr or a >50 percent reduction to an AHI<20 events/hr without symptoms, was similar in both groups and in the subgroups of patients with baseline AHI below and above 30 events/hr. However, CPAP was more effective than oral appliance at achieving an AHI of <5 events/hr in all patients (risk difference= -20 percent; 95 percent CI -37, -2; P=0.02).

In the subgroup analyses, a larger, significant effect was found in patients with a baseline AHI >30 events/hr; no difference was observed in those with less severe OSA. The study also did not find a difference in this subgroup in achieving an AHI <10 events/hr.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.12.3; Balk [AHRQ] (2011) Figure 23)
Nine trials provided data on AHI outcomes.[129,140,224,226-232] All trials reported that AHI was lower in patients on CPAP than when using MADs. The results were statistically significant in seven of the trials. Meta-analysis of the eight trials with adequate data found that the difference in AHI between MAD and CPAP was statistically significant, favoring CPAP (difference = 7.7 events/hr; 95 percent CI 5.3, 10.1; P<0.001).

Analysis of the net difference assessed in the two parallel trials and of the difference of final values in the six crossover trials yielded similar findings. However, the trial results were statistically heterogeneous.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.12.4; Balk [AHRQ] (2011) Figure 24)
Seven trials provided data on ESS outcomes.[129,140,225,227,229,230,232] Four trials found no significant difference in ESS between the two interventions, while three found a significantly lower ESS in patients on CPAP (net difference 2 and 6 units). However, only Gagnadoux 2009 found a statistically significant lower ESS (-0.5 units) while patients were using MAD. Metaanalysis revealed no significant difference between the two interventions but indicated a trend somewhat favoring CPAP (difference = 1.3; 95 percent CI -0.2, 2.8; P=0.098). A large degree of the statistical heterogeneity across studies was due to Engleman 2002, which found a considerably larger difference favoring CPAP (net difference = 6; estimated 95 percent CI 4.2, 7.8; P<0.001). It is unclear why this study found a different magnitude effect; the study shared features with several other studies that had reported smaller effects. Excluding this one study, meta-analysis indicated no difference between the interventions, though statistical heterogeneity remained (difference = 0.4, 95 percent CI -0.6, 1.3).

Other Sleep Study Measures (Balk [AHRQ] (2011) Appendix D Table 5.12.5a-c; Balk [AHRQ] (2011) Figures 25 & 26)
Five studies evaluated arousal index (Figure 25). Meta-analysis revealed that arousals were significantly more common while using MAD than CPAP (difference = 3.5 events/hr; 95 percent CI 1.5, 5.5; P=0.001). All studies reported a higher arousal index on MAD than CPAP, though only Barnes 2004 found a statistically significant difference (and Skinner 2004, discontinued early, which found a large, but marginally nonsignificant difference between the cervicomandibular support collar and autoCPAP) (Appendix D Table 5.12.5a).

Seven studies evaluated minimum oxygen saturation (Figure 26). Meta-analysis revealed that the studies were homogeneous and indicated a statistically significant lower oxygen saturation while using MAD than CPAP (difference = -3.5 percent; 95 percent CI -4.6, -2.4; P<0.001). All studies found a consistent effect, though only two trials were statistically significant (Appendix D Table 5.12.5b).

Six studies found no significant difference in sleep efficiency (range of effects -2.9 percent, 0.4 percent total sleep time) (Appendix D Table 5.12.5c). Five of these studies found a consistent (though nonsignificant) trend toward less time in slow wave sleep with MAD (range of effects -3.9 percent, -0.6 percent total sleep time) (Appendix D Table 5.12.5d).

Seven studies evaluated REM sleep (Appendix D Table 5.12.5e). No statistically significant difference in percentage of time spent in REM sleep was reported, although the range of differences between the two interventions was large (-4.7 to 6.1). There was no clear explanation for the different results across studies.

Objective Sleepiness and Wakefulness Tests (Balk [AHRQ] (2011) Appendix D Table 5.12.6)
Two studies evaluated wakefulness tests. Engleman 2002 found no difference in the Maintenance of Wakefulness Test sleep onset latency. Gagnadoux 2009 found no difference in the Oxford Sleep Resistance test sleep latency.

Quality of Life (Balk [AHRQ] (2011) Appendix D Table 5.12.7a-b; Balk [AHRQ] (2011) Figure 27)

Three studies measured FOSQ (Appendix D Table 5.12.7a) with inconsistent findings. Hoekema 2008 (in a trial of an oral MAD versus CPAP) and Skinner 2004 (in an aborted trial of a cervicomandibular support collar versus autoCPAP) found no difference in quality of life as measured by FOSQ. In contrast, Engleman 2002 (in a trial of oral MAD versus CPAP) found that quality of life improved significantly less while patients were using MAD than while using CPAP. Meta-analysis revealed no significant effect on FOSQ (difference = -0.86; 95 percent CI -2.5, 0.8).

Seven studies measured various quality of life tests (Appendix D Table 5.12.7b); five used SF-36, two used the Hospital Anxiety and Depression Scale, and one study each used the Beck Depression Index, the Calgary Sleep Apnea Quality of Life Index (SAQLI), the Nottingham Health Profile, a ―General Health‖ measure, and the Scottish National Sleep Laboratory symptom questionnaire. Five of the studies found no significant difference in quality of life among patients using MAD or CPAP. The remaining two studies found differences in components of SF-36 favoring CPAP: Engleman 2002 found significant differences in various components of SF-36, and Lam 2007 found a large net difference in only the Bodily Pain component (-16 points). Lam 2007 also found differences in SAQLI, which separately measures the effect of treatment on quality of life and any treatment-related symptoms (adverse effects). The study found that, overall, CPAP was better at improving quality of life, but that patients treated with CPAP had more treatment-related symptoms. Combining quality of life findings and treatment-related symptoms (the analysis SAQLI A-E), neither intervention was superior.

Neurocognitive Tests (Balk [AHRQ] (2011) Appendix D Table 5.12.8)

Two studies evaluated neurocognitive tests. Neither Engleman 2002 nor Gagnadoux 2009 found any significant differences in a range of tests of cognitive performance (IQ), executive function (Trailmaking), processing speed (Paced Auditory Serial Addition Test), error making (Oxford sleep resistance), or driving skills (SteerClear).

Study Variability

Only one study reported subgroup analyses. As discussed above, Hoekema 2008 found no difference in the effective treatment rate between interventions in either those with more or less severe OSA (at an AHI threshold of 30 events/hr) after 2 to 3 months. However, those patients with a baseline AHI >30 events/hr were more likely to achieve an AHI of <5 events/hr with CPAP than MAD, as compared with those with a lower baseline AHI. An evaluation using a final AHI of <10 events/hr did not confirm this difference. This trial was a parallel design RCT, enrolling 103 patients with a minimum AHI of 5 events/hr but a relatively high mean AHI of 40 events/hr and with relatively severe sleepiness (mean ESS = 14.2).

For most evaluated outcomes, the reviewed studies were generally consistent in their findings. Where there were outliers, no clear differences in effect based on patient characteristics, severity of sleep apnea, other symptoms, or apparent differences in OSA definitions (particularly minimum AHI threshold) were observed.

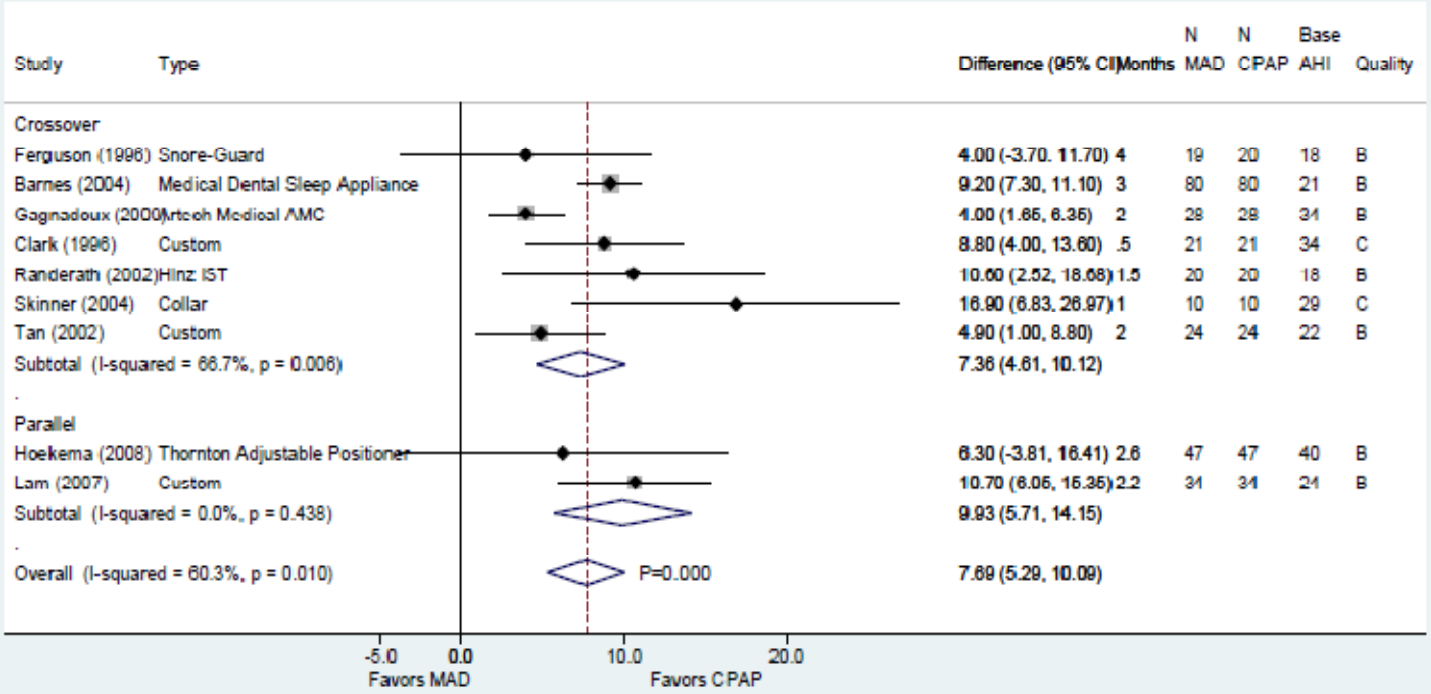
The only consistent difference across studies and outcomes reported was that of the aborted study comparing a cervicomandibular support collar with autoCPAP. This study reported differences in effects (which favored autoCPAP) that were generally larger than the differences for the intraoral MADs compared with CPAP. This, apparently, was the reason that the study was prematurely terminated.

Summary

Ten trials (most quality B) compared MAD with CPAP; nine compared intraoral devices with CPAP and one compared an extraoral device with autoCPAP. The reviewed studies generally found that CPAP was superior in reducing AHI, reducing arousal index, raising minimum oxygen saturation. The evidence regarding relative effects on ESS is unclear due to heterogeneity of results across studies. These findings were confirmed by meta-analysis. No difference in effect was found for other sleep measures. Most studies found no significant difference in effects on quality of life or neurocognitive function, although one study found that the benefits of CPAP over MAD, as measured by SAQLI, were counterbalanced by an increase in perceived treatment-related symptoms under CPAP. Only one of two studies found a difference (favoring CPAP) in treatment response. Only one study evaluated compliance, finding that patients were more compliant with MAD than CPAP. No consistent or substantive differences in effects were found based on patient characteristics, disease severity, or other baseline symptoms.

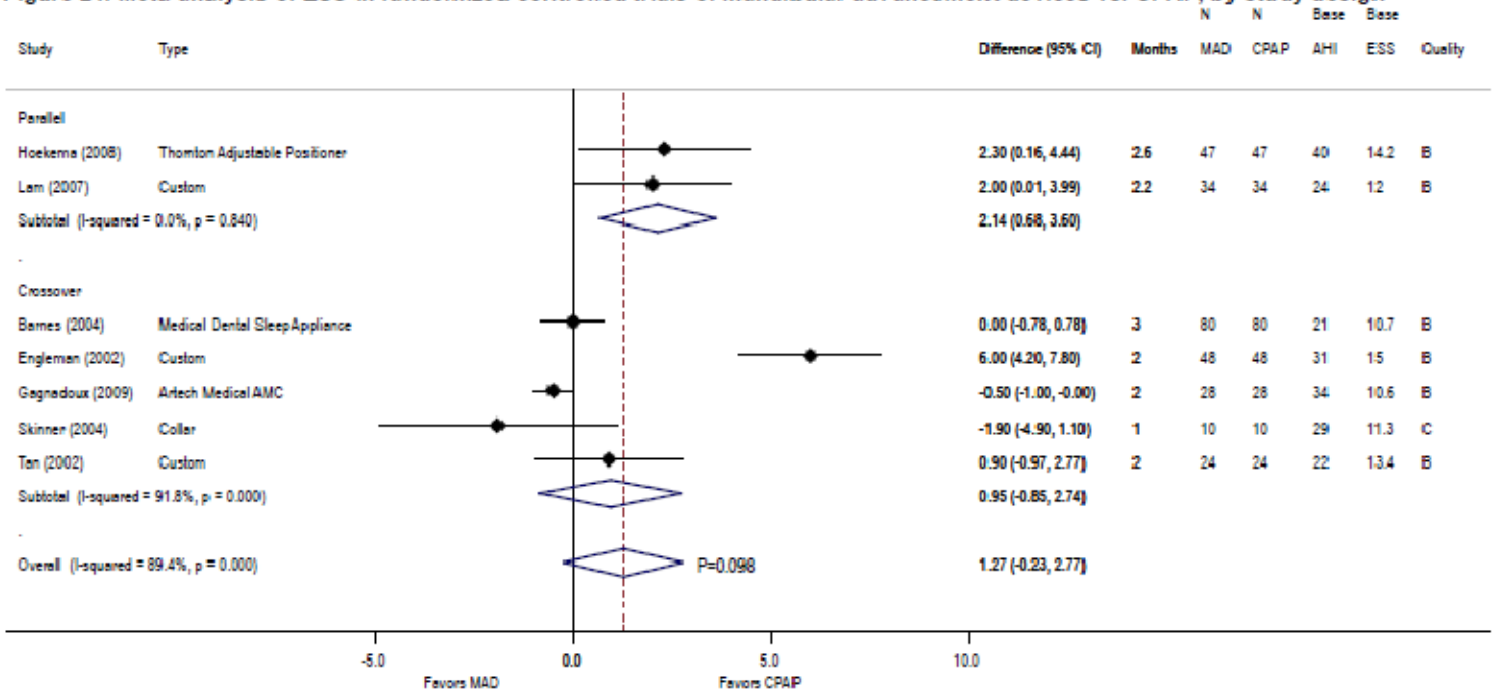
There was sufficient evidence supporting greater improvements in sleep measures with CPAP as compared to MAD, but only weak evidence indicating no or only small differences favoring CPAP for improving compliance, treatment response, quality of life, or neurocognitive measures. There were no data on objective clinical outcomes. Nevertheless, overall there remains a moderate strength of evidence that CPAP is superior to MAD to improve sleep study measures. However, the strength of evidence is insufficient to address which patients might benefit most from either treatment.

Figure 23. Meta-analysis of AHI (events/hr) in randomized controlled trials of mandibular advancement devices vs. CPAP, by study design



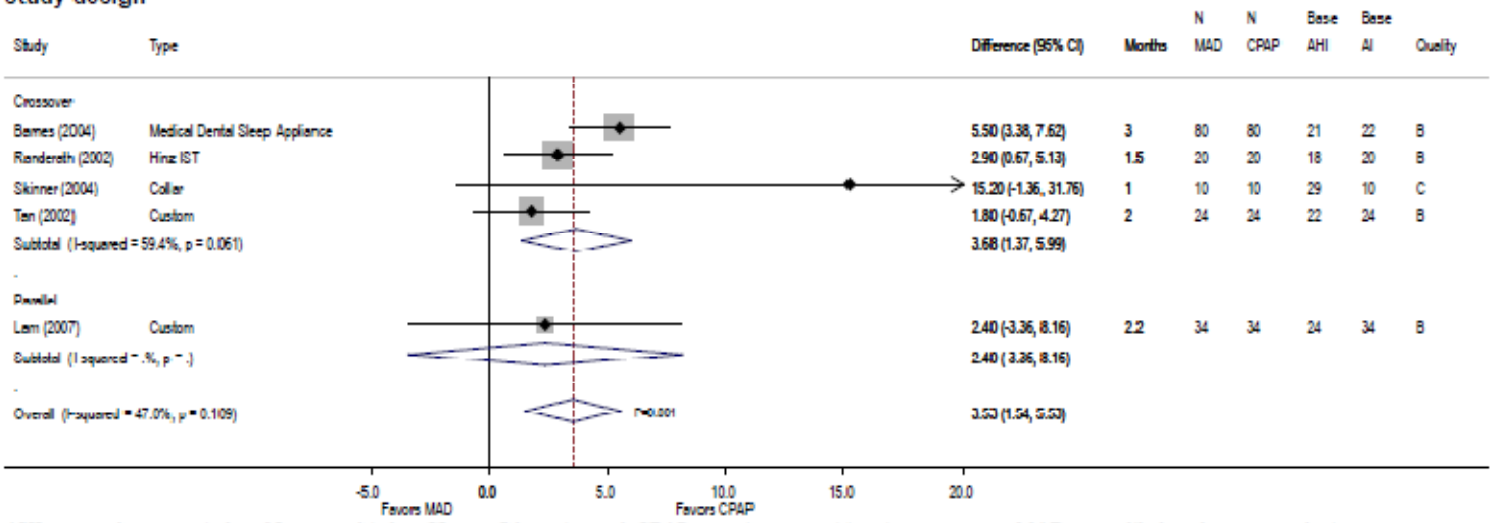
AHI = apnea-hypopnea index, CI = confidence interval, CPAP = continuous positive airway pressure, MAD = mandibular advancement device.

Figure 24. Meta-analysis of ESS in randomized controlled trials of mandibular advancement devices vs. CPAP, by study design



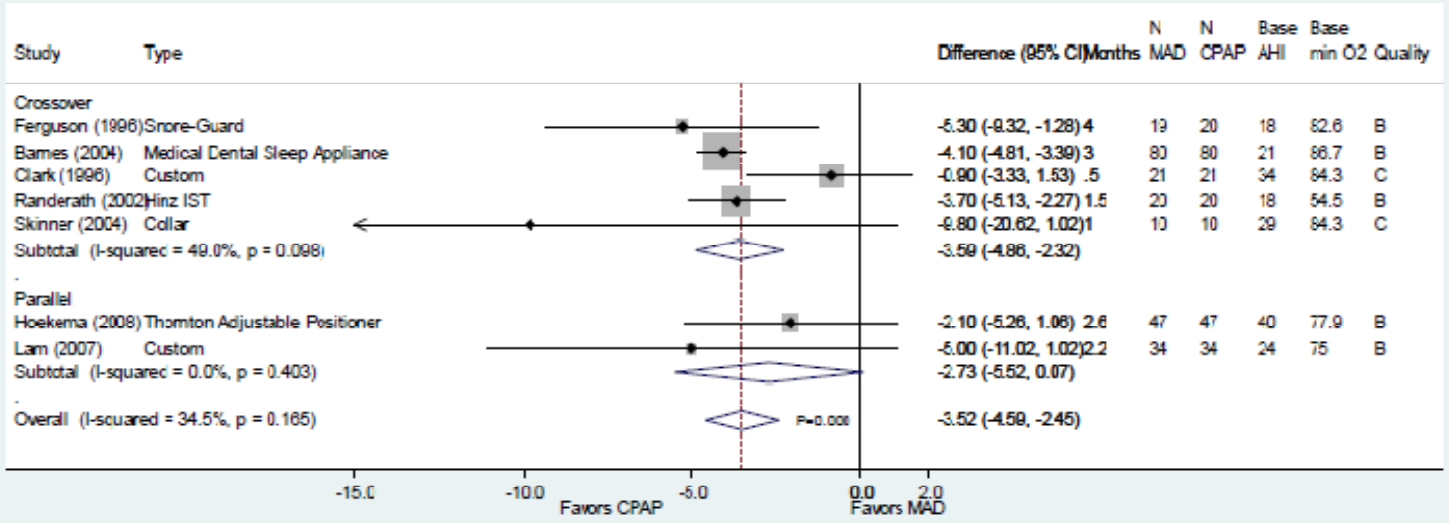
AHI = apnea-hypopnea index, CI = confidence interval, CPAP = continuous positive airway pressure, ESS = Epworth Sleepiness Scale, MAD = mandibular advancement device.

Figure 25. Meta-analysis of arousal index (events/hr) in randomized controlled trials of mandibular advancement devices vs. CPAP, by study design



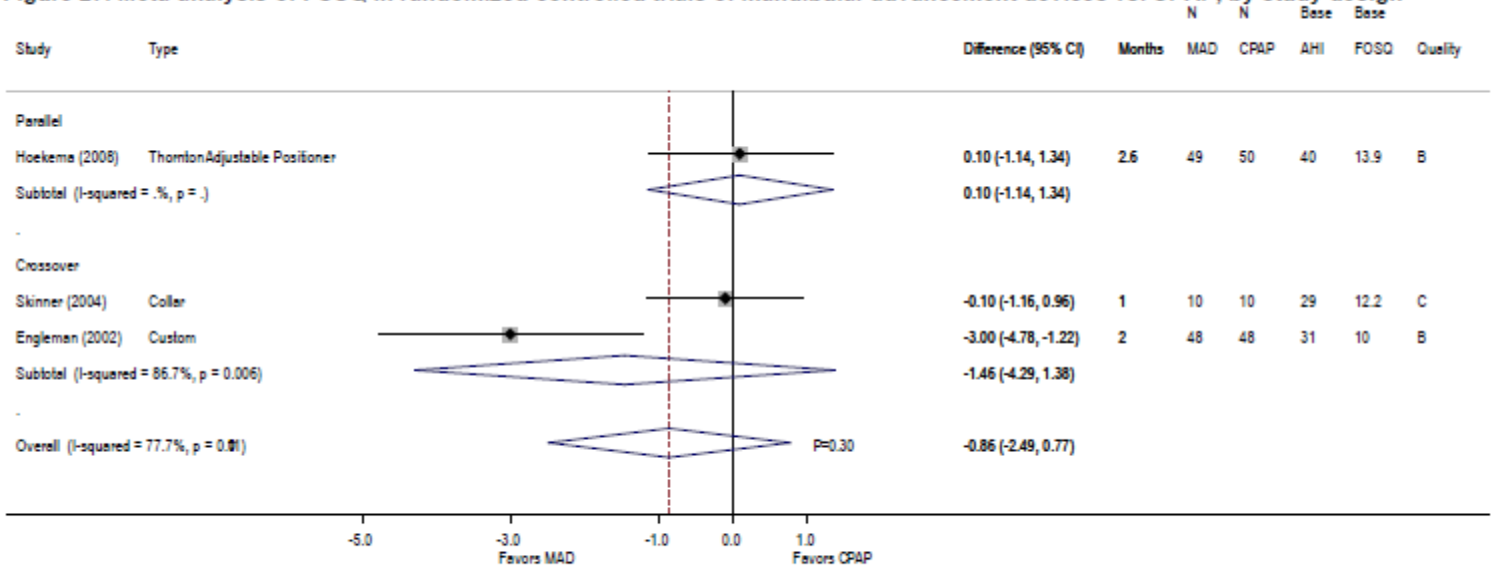
AHI = apnea-hypopnea index, AI = arousal index, CI = confidence interval, CPAP = continuous positive airway pressure, MAD = mandibular advancement device.

Figure 26. Meta-analysis of minimum oxygen saturation (%) in randomized controlled trials of mandibular advancement devices vs. CPAP, by study design



AHI = apnea-hypopnea index, CI = confidence interval, CPAP = continuous positive airway pressure, MAD = mandibular advancement device, min O2 = minimum oxygen saturation

Figure 27. Meta-analysis of FOSQ in randomized controlled trials of mandibular advancement devices vs. CPAP, by study design



AHI = apnea-hypopnea index, CI = confidence interval, CPAP = continuous positive airway pressure, FOSQ = Functional Outcomes Sleep Questionnaire, MAD = mandibular advancement device.

Comparison of Positional Therapy and CPAP

Three crossover trials compared three different positional devices with CPAP. One trial compared a shoulder head elevation pillow with autoCPAP,[233] and two compared devices worn on the back to prevent sleeping supine with either autoCPAP[234] or CPAP[235] (Balk [AHRQ] (2011) Appendix D Table 5.16.1). Across studies, mean baseline AHI ranged from 18 to 27 events/hr. Skinner 2004 included patients with an AHI ≥10 events/hr, Skinner 2008 included patients with AHI ≥5, and Jokic 1999 included only patients who were shown to have an AHI <15 events/hr while in the lateral position. All patients had positional OSA; none were patients for whom positional therapy might be contraindicated due to conditions such as chronic musculoskeletal pain or other conditions affecting sleep. Study sample sizes ranged from 13 to 20 patients (total = 47 across studies). All studies were rated quality B. Small sample sizes and lack of patient and outcome assessor blinding were the main methodological quality concerns. These studies are applicable mainly to patients with AHI less than 30 events/hr who have positional OSA, and for whom positional therapy would not be contraindicated due to comorbid conditions.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Compliance

No study evaluated compliance.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.16.2a-b)

All three trials provided data on AHI as an outcome. Each trial reported that AHI decreased significantly more in patients on CPAP as compared with those on positional therapy. Jokic 1999 found a difference of 6.1 events/hr (95 percent CI 2.0, 10; P=0.007), Skinner 2004 found a difference of 16 events/hr (95 percent CI 4.2, 28; P=0.008), and Skinner 2008 found a difference of 7.1 (95 percent CI 1.1, 13; P=0.02). Skinner 2008 also reported statistically significantly more patients achieved an AHI of ≤10 events/hr with CPAP (89 percent) than with a Thoracic antisupine band (72 percent; P=0.004, by Wilcoxin sign-rank test), though the relative risk of achieving a low AHI was nonsignificant (0.81; 95 percent CI 0.58, 1.13).

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.16.3)

All three trials provided data on ESS as an outcome. Each trial reported that ESS scores were higher in patients on positional therapy than those on CPAP (differences ranged from 0.7 to 1.5), although none of these findings were statistically significant.

Other Sleep Measures (Balk [AHRQ] (2011) Appendix D Table 5.16.4)

Jokic 1999 reported a nonsignificantly larger drop in arousal index in patients on CPAP as compared with those on positional therapy. No significant differences in maintenance of wakefulness testing, sleep latency, sleep efficiency, percentage of time spent in stage 3-4 sleep, and percentage of time spent in REM sleep were observed. Arousal index was nonsignificantly higher in patients on positional therapy (difference = 4.5 events/hr; 95 percent CI -0.7, 9.4; P=0.08).

Quality of Life (Balk [AHRQ] (2011) Appendix D Tables 5.16.5 & 5.16.6)

Skinner 2004 and Skinner 2008 both reported no significant difference in SF-36 mental component and physical component summaries for patients in the CPAP group compared with those in the positional therapy group (Appendix D Table 5.16.5a-b). Skinner 2004 also found no significant difference in FOSQ score between the two groups (Appendix D Table 5.16.6).

Jokic 1999 found a lower score in the Nottingham Health Profile energy subscale in the positional therapy group (difference = -1; P=0.04; Appendix D Appendix D Table 5.16.7a) , but no difference between treatments in the Hospital Anxiety and Depression Scale (Appendix D Table 5.16.7b), University of Wales Institute of Science and Technology (UWIST) mood adjective checklist (Appendix D Table 5.16.7c), or General Health Questionnaire (Appendix D Table 5.16.7d).

Neurocognitive Tests (Balk [AHRQ] (2011) Appendix D Table 5.16.7)

Jokic 1999 found no significant difference between CPAP and positional therapy in the Wechsler Memory Scale, Purdue Pegboard, Trail-Making Test, Symbol Digit Modalities, Consonant Trigram, or Concentration Endurance Test scores.

Study Variability

No studies performed subgroup analyses. As study treatments were heterogeneous, we were unable to examine differences in outcomes based on patient characteristics.

Summary

Three small quality B crossover trials compared different positional treatments with CPAP. AHI was found to be lower in patients using CPAP than in those on positional therapy. ESS scores were not significantly different between groups. Additionally, quality of life measurements and neurocognitive tests showed no difference between positional therapy and CPAP.

Because of the small number of studies and because each study evaluated a different device, the strength of evidence is insufficient to determine the relative merit of positional therapy compared with CPAP in the treatment of OSA.

Comparison of Weight Loss Interventions and Control Interventions

Three parallel trials compared weight loss interventions with control interventions (Balk [AHRQ] (2011) Appendix D Table 5.17.1).[236-238] Foster 2009 enrolled patients with type 2 diabetes and randomized them to an intensive lifestyle intervention (a behavioral weight loss program involving portion-controlled diets and physical activity prescription) or a diabetes support and education program (three educational sessions on diabetes management over a 1 year period on diet, physical activity, and social support). Johannson 2009 randomized patients to a group following a 9 week low energy diet or a group that was instructed to adhere to their usual diet. Tuomilehto 2009 enrolled obese patients and randomized them to a group following a very low calorie diet complemented with lifestyle changes or a group subject to general counseling on diet and exercise only. Mean baseline AHI in these studies ranged from 9 to 37 events/hr. Study sample sizes ranged from 63 to 264 (total = 345 across studies). Study durations ranged from 2.3 to 12 months. Johansson 2009 was rated quality A, while the other two were rated quality B. The main methodological concerns were the lack of clarity on whether the outcome data included all initial participants and unclear reporting of outcomes. The inclusion criteria used in these studies varied considerably 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.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Treatment Response (Balk [AHRQ] (2011) Appendix D Table 5.17.2)

Tuomilehto 2009 examined cure from OSA as a dichotomous outcome. OSA was considered objectively cured when the AHI was <5 events/hr at 1 year. Treatment with a very low calorie diet was associated with a 4-fold increase in the odds of being cured from OSA at 1 year compared with the control intervention (adjusted odds ratio 4.2; 95 percent CI 1.4, 12; P=0.011).

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.17.3)

All three studies examined AHI and demonstrated statistically significant reductions in AHI for the arms randomized to weight loss interventions. The reductions ranged from -4 to -23 events/hr. Johansson 2009 showed the largest net reduction in AHI. This study enrolled patients with no comorbidities but with more severe OSA (baseline AHI = 37) as compared to the other two studies; it also involved a much shorter duration of followup (9 weeks).

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.17.4)

Two trials provided data on changes in daytime sleepiness as assessed using the ESS. Johansson 2009 reported a statistically significant reduction in ESS scores for the low energy diet group, while Tuomilehto 2009 found no significant difference.

Minimum Oxygen Saturation (Balk [AHRQ] (2011) Appendix D Table 5.17.5)

Only Johansson 2009 reported changes in minimum oxygen saturation; the lower energy diet was associated with a statistically significant net increase in the minimum oxygen saturation as compared to usual diet (5 percent; 95 percent CI 2, 7; P=0.002).

Other Outcomes (Balk [AHRQ] (2011) Appendix D Tables 5.17.6-5.17.8)

Tuomilehto 2009 examined the impact of a weight loss intervention on blood pressure measurements (Appendix D Table 5.17.6). No statistically significant changes were detected for systolic or diastolic blood pressure. Foster 2009, which was conducted exclusively in diabetic patients, examined the impact of an intensive lifestyle intervention on hemoglobin A1c concentration (Appendix D Table 5.17.7) and reported a statistically significant net difference (-0.5 percent; P<0.001) at 1 year followup.

In all three studies, the weight loss program resulted in large reductions in weight (Appendix D Table 5.17.8) of -10.7, -10.8, and -18.7 kg; the control interventions resulted in near stable weight (changes ranging from -2.4 to +1.1 kg). These differences were all highly statistically significant (P<0.001).

Study Variability

No study reported subgroup analyses with respect to the comparative effect of weight loss interventions versus control interventions for OSA in terms of patient characteristics (age, sex, race, weight, bed partner, and airway) or severity of OSA.

Given the small number of studies and the variability of interventions, no conclusions could be reached regarding whether effects of weight loss interventions varied for different subgroups of patients.

Summary

Findings from three parallel RCTs supported a benefit of intensive weight loss interventions in reducing AHI. The reviewed studies were quality A or B and reported consistent results supporting the improvement of AHI with weight loss interventions, either as a continuous outcome (three studies) or as a dichotomous outcome for cure based on an AHI of less than 5 events/hr (one study). It should be noted,

however, that the study that showed the largest benefit had relatively few participants. Conclusive statements cannot be made about other outcomes evaluated due to inconsistent results or a limited number of studies per outcome. No data on objective clinical outcomes were reported. Overall, there is a low strength of evidence to show that some intensive weight loss programs may be effective in relieving the signs and symptoms of sleep apnea in obese patients with OSA.

Comparison of Oropharyngeal Exercise and Control

Three trials compared different methods of oropharyngeal exercise with CPAP (Balk [AHRQ] (2011) Appendix D Table 5.18.1).[239-241] All three had a parallel design, and tested methods intended to train aspects of the upper airway and reduce symptoms of OSA. These methods included didgeridoo training, oropharyngeal exercise, and tongue training.

Mean baseline AHI ranged from 20 to 27 events/hr. Study sample sizes ranged from 25 to 57 patients. Both Puhan 2005 and Randerath 2004 were rated quality A, while Guimaraes 2009 was rated quality B due to a small sample size and unclear reporting. The studies are generally applicable to patients with AHI ≥15 events/hr.

Study Results (Balk [AHRQ] (2011) Appendix D Tables 5.18.2-5.18.10)

As the devices compared varied considerably and each study examined different outcomes, trials are described separately below. No study evaluated objective clinical outcomes.

Guimaraes 2009[240] compared oropharyngeal exercise (consisting of exercise of the soft palate, tongue, and facial muscles plus stomatognathic function exercises) to sham therapy (consisting of deep breathing, nasal lavage, and recommendations for bilateral chewing). The sample consisted of 31 patients with moderate OSA (AHI 15-30 events/hr). Patients were excluded if they had a BMI >40 kg/m^2 or major comorbidities. Patients in the oropharyngeal exercise group were 64 percent male and had a mean age of 52 years. Those in the control group were 73 percent male and had a mean age of 48 years. The study found that oropharyngeal exercise resulted in a significantly lower AHI (difference = -12 events/hr, 95 percent CI -19, -5; P<0.001) (Appendix D Table 5.18.2), as well as lower ESS scores (difference = -4.0; 95 percent CI -8, -0.02; P<0.05) (Appendix D Table 5.18.3). No significant differences between groups were observed in minimum oxygen saturation (Appendix D Table 5.18.4) or sleep efficiency (Appendix D Table 5.18.5). The oropharyngeal exercise group had a significantly lower Pittsburgh Quality of Sleep Index score (difference = -3.4; P<0.01) (Appendix D Table 5.18.9).

Randerath 2004[241] compared tongue training (using a muscle stimulator placed under the tongue and chin) to sham training (using the same device but without electrical stimulation). The study consisted of 57 newly diagnosed OSA patients with an AHI of 10-40 events/hr. Patients had no other significant comorbidities. Patients in the tongue training group were 57 percent male and had a mean age of 51 years. Those in the control group were 73 percent male and had a mean age of 53 years. The study found no significant difference between groups in AHI, ESS, minimum oxygen saturation (Appendix D Tables 5.18.2-5.18.4), slow-wave or REM sleep (Appendix D Table 5.18.6), arousal index (Appendix D Table 5.18.6), FOSQ score (Appendix D Table 5.18.8), or Attention Test score (Appendix D Table 5.18.10).

Puhan 2005[239] compared didgeridoo training to no treatment. The study consisted of 25 mostly male patients (mean age: 49 years) with an AHI range of 15-30 events/hr and a mean BMI ≤30 kg/m^2. All patients complained of snoring. Training consisted of instruction on the didgeridoo, which involves

learning a breathing technique called circular breathing. Patients practiced for 30 minutes daily, 6 days a week. After 4 months, the didgeridoo group had a significantly lower AHI (difference = -6.2 events/hr; 95 percent CI -12.3, -0.1; P=0.05; Appendix D Table 5.18.2) and ESS score (difference = -2.8; 95 percent CI -5.7, -0.3; P=0.04; Appendix D Table 5.18.3). No differences between groups were observed in any domain of SF- 36 or in the Pittsburgh Quality of Sleep Index (Appendix D Table 5.18.9).

Study Variability

None of the studies reviewed performed subgroup analyses. No comparisons could be made across studies.

Summary

Three trials with unique comparisons compared oropharyngeal exercise to control for treatment of patients with OSA. One study on a specific form of oropharyngeal exercise and one study on didgeridoo training reported improved sleep study measures. A third study found tongue training to not be beneficial in relieving the symptoms of OSA. Overall, due to the limited number of studies, the strength of evidence is insufficient to determine a definitive benefit of oropharyngeal exercise in the treatment of OSA.

Comparison of Palatal Implant and Placebo Implant

Two RCTs compared palatal implants and placebo implants in patients with OSA (Balk [AHRQ] (2011) Appendix D Table 5.19.1). [242,243] Both studies included only patients with mild to moderate sleep apnea and no other significant comorbidities. Mean baseline AHI was 20 events/hr in Friedman 2008 and 16 events/hr in Steward 2008; mean ESS values were 11.7 and 10.6, respectively. While Friedman 2008 had an equal sex distribution, Steward 2008 included a majority (79 percent) of men most of whom had retropalatal pharyngeal obstruction. The mean ages of patients in the studies were 39 years (Friedman 2008) and 49 years (Steward 2008). Friedman 2008, a quality A study, enrolled 62 patients and Steward 2008, a quality B study, enrolled 100 patients. Both studies were double-blinded and had a 3 month followup (Appendix D Table 5.19.1). Neither study evaluated objective clinical outcomes. These studies are applicable to patients with AHI of 5 to 40 events/hr and BMI less than 30 kg/m^2.

Study Results (Balk [AHRQ] (2011) Appendix D Tables 5.19.2-5.19.6)

Friedman 2008 found significant improvements in AHI (P<0.0001; Appendix D Table 5.19.2a), ESS values (P=0.0002; Appendix D Table 5.19.3), and SF-36 total score (P<0.0001) with palatal implants as compared to placebo. The study did not find significant differences in minimum oxygen saturation (Appendix D Table 5.19.4) or REM sleep as a percentage of total sleep time (Appendix D Table 5.19.5) between groups. This study was rated quality A.

In contrast, Steward 2008 did not find statistically significant differences in mean AHI (Appendix D Table 5.19.2a) or ESS values (Appendix D Table 5.19.3) between palatal implants and placebo. However, the study did find that a clinically meaningful reduction in AHI (≥50 percent reduction to <20 events/hr) was more common in the palatal implant group as compared to placebo (26 versus 10 percent, P=0.04; Appendix D Table 5.19.2b). The study also reported significant improvements in minimum oxygen saturation (P=0.007; Appendix D Table 5.19.4) and FOSQ (P<0.05; Appendix D Table 5.19.6) with palatal implants as compared to placebo. This study was rated quality B.

Study Variability

Neither study performed subanalyses. Due to the limited number of studies, we were unable to assess potential differences with regards to factors of interest such as patient characteristics and severity of obstructive sleep apnea.

Summary

Two studies in patients with mild to moderate OSA compared treatment effects of palatal implants to placebo implants. Both studies found significantly greater improvements in sleep study measures and quality of life with palatal implants; however, the studies disagreed as to which specific outcomes palatal implants significantly improved. Overall, due to the limited number of studies reviewed, the strength of evidence is insufficient to determine the relative efficacy of palatal implants versus sham implants in patients with mild to moderate OSA.

Comparison of Surgery and Control Treatments

Six trials in seven publications and one prospective nonrandomized comparative study investigated the effects of several surgical interventions compared to control (Balk [AHRQ] (2011) Appendix D Table 5.20.1).[124,125,244-249] Each study used a different intervention: uvulopalatopharyngoplasty (UPPP), laser-assisted uvulopalatoplasty (LAUP), radiofrequency ablation (RFA), and combinations of pharyngoplasty, tonsillectomy, adenoidectomy, genioglossal advancement septoplasty, radiofrequency ablation of the inferior nasal turbinates, or combination nasal surgery. The control treatments were sham surgery, conservative therapy, or no treatment.

Patients included in the surgery comparisons were reported to have prior treatment failures with nonsurgical techniques or declined their usage. The mean baseline AHI ranged from 5 to 40 events/hr; three trials included patients with an AHI ≥5, one with an AHI ≥10, and one did not report an AHI threshold. One trial reported ODI (mean at baseline 21- 72 events/hr). All included only patients with relatively less severe OSA (AHI <30-50). Study sample sizes ranged from 26 to 52 (total = 223 across studies). Three studies were rated quality A, one quality B, and two quality C. Guilleminault 2008 was reported as a crossover study comparing several surgical combinations to cognitive behavioral therapy. This study was rated quality C due to an inappropriate study design as the effects of surgery could not be reversed. These studies are applicable mostly to patients with a range of baseline AHI and BMI less than 35 kg/m^2.

Study Results (Balk [AHRQ] (2011) Appendix D Tables 5.20.2-5.20.7)

As each study evaluated a different surgical technique, each study is described individually. No studies evaluated objective clinical outcomes.

Back 2009 compared a single session of RFA surgery of the soft palate to sham surgery (simulated surgery with no energy administered). The study included 32 male patients with mild sleep apnea (AHI 5-15 events/hr) and habitual snoring following a failed trial of conservative treatment (weight loss, positional therapy, restriction of alcohol and sedatives). Patients were between the ages of 30 and 65 years. At 4 month followup, no statistically significant difference between groups in AHI (Appendix D Table 5.20.2), ESS (Appendix D Table 5.20.3), minimum oxygen saturation (Appendix D Table 5.20.4), and quality of life (as measured by SF-36; Appendix D Table 5.20.7) were found. This study was rated quality A.

Koutsourelakis 2008 randomized patients to either nasal surgery (submucous resection of the deviated septum and bilateral resection of inferior turbinates) or sham surgery (simulated nasal surgery under anesthesia). In addition to OSA (defined as AHI ≥5 events/hr), all patients had fixed nasal obstruction due to deviated nasal septum. The study was conducted on 49, predominately male patients with a mean baseline AHI of 31 events/hr. After 4 months followup, the study found no statistically significant difference between groups in AHI (Appendix D Table 5.20.2) or on ESS (Appendix D Table 5.20.3). This study was rated quality A.

Woodson 2003 conducted a three-arm RCT that included a comparison of multilevel temperature controlled RFA of the soft palate with sham surgery (simulated RFA with no energy delivered). The study was conducted in 51, predominately male patients. Notably, the age of participants between groups was significantly different at baseline. (49 years (RFA) versus 51 years (sham), P=0.04). The mean baseline AHI also differed among groups (21 (RFA) versus 15 (sham) events/hr; P=0.06, including the CPAP study group). After 8 weeks followup, the study found a significantly greater improvement in sleep quality as measured by FOSQ with RFA as compared to sham surgery (P=0.04; Appendix D Table 5.20.6), but no statistically significant difference in AHI (Appendix D Table 5.20.2), ESS (Appendix D Table 5.20.3), minimum oxygen saturation (Appendix D Table 5.20.4), or quality of life as measured by SF-36 (Appendix D Table 5.20.7). This study was rated quality A.

Ferguson 2003 randomized patients to either LAUP or no treatment. In LAUP, the uvula and a specified portion of the palate is vaporized under local anesthesia in an outpatient setting. The goal is to relieve obstruction in patients with mild OSA or snoring. The study included 44 mostly male patients with mild OSA (AHI 10-27 events/hr) and snoring. The patients had a mean age of 45 years and a mean BMI of 31.6 kg/m^2. This study reported disparate followup durations of 15 months in the LAUP group and 8 months in the control group. A statistically significant improvement in AHI was observed following LAUP as compared with no treatment (net change -10.5 events/hr; P=0.04; Appendix D Table 5.20.2). However, there was no statistically significant difference between groups on the ESS (Appendix D Table 5.20.3) or in quality of life as measured by SAQLI (Appendix D Table 5.20.7). This study was rated quality B.

Guilleminault 2008 was reported as a crossover study comparing several surgical combinations to cognitive behavioral therapy in 30 patients with insomnia and mild OSA (mean AHI 10 events/hr). Based on anatomy, disease severity, and comorbidity, patients received combinations of pharyngoplasty, tonsillectomy, adenoidectomy, genioglossal advancement septoplasty, and RFA of the inferior nasal turbinates. Since the surgery could not be undone during the second phase of the study, we evaluated only the first phase as a parallel trial. Results showed that surgery led to improvements in AHI (-6.2 events/hr; P=0.0001; Appendix D Table 5.20.2), ESS (-1.1; P=0.002; Appendix D Table 5.20.3), minimum oxygen saturation (4.4 percent; P=0.0001; Appendix D Table 5.20.4), REM (2.9 percent of total sleep time; P=0.0001; Appendix D Table 5.20.5), and slow wave sleep (3.5 percent of total sleep time; P=0.0001; Appendix D Table 5.20.5) as compared to cognitive behavioral therapy. This study was rated quality C due to the design issues described above.

Lojander 1996 & 1999 compared UPPP with or without mandibular osteotomy to conservative treatment (weight loss, positional therapy, and avoidance of tranquilizers and alcohol at bedtime). The study included 32, predominately male patients with a mean age of 47 years and a mean baseline BMI of 31 kg/m^2. Baseline ODI ranged from 10 to 72 events/hr. A significant improvement in daytime somnolence (net difference -25 on a visual analogue scale ranging from 0 (no somnolence) to 100 (worst); P<0.05) was

observed after 12 months; no statistically significant difference was found between groups in cognitive function (Wechsler test; Appendix D Table 5.20.7). This study was rated quality C due to problems with the power calculation, a small sample size, and a possible selection bias stemming from the use of an expert panel to determine which patients would be most suitable for UPPP.

Li 2009, in a nonrandomized prospective study (quality C), compared correction of nasal septum and volume reduction of the inferior turbinates to conservative nasal treatments in patients with snoring, nasal obstruction, and OSA. The study included 66 patients, 44 of whom had surgery. The patients were almost all male, with a mean age of 38 years and a mean BMI of 26.2 kg/m^2 . Baseline AHI was 38 events/hr in the surgically treated group and 26 in the conservative treatment group (no significant difference), and baseline ESS was 10.6. The article did not report at what timepoint followup data were collected. The study found a statistically significant difference in ESS, favoring surgery (net difference -3.6; 95 percent CI -6.1, -1.1; P=0.02; Appendix D Table 5.20.3). The study found no difference in AHI, minimum oxygen saturation, slow wave sleep, or REM sleep (Appendix D Tables 5.20.2, 5.20.4, 5.20.5) However, seven of 44 patients receiving surgery had success by the Sher criteria (followup AHI <30 events/hr and reduction in AHI of at least 50 percent) and none of the 22 patients on conservative treatment (P=0.048 per the article). The study did note that six of the seven patients with surgical success had, at baseline low ESS (<10.5), a low Friedman tongue position (grade II or III), and a low BMI (<25.8 kg/m^2).

Study Variability

None of the studies performed subgroup analyses. As there is only one study per comparison, we were unable to assess potential differences with regards to factors of interest such as patient characteristics and severity of OSA

Summary

Seven studies with unique interventions compared surgery with control treatment for the management of patients with OSA. Due to the heterogeneity of the studies reviewed and inconsistency as to which outcomes were improved with surgery as compared to no or sham surgery, the strength of evidence is insufficient to evaluate the relative efficacy of surgical interventions for the treatment of OSA.

Comparison of Surgery and CPAP Treatments

Two parallel RCTs,[247,250] four prospective studies,[251-254] and six retrospective studies[255-260] investigated the effects of several surgical interventions compared with CPAP in adults with OSA (Balk [AHRQ] (2011) Appendix D Table 5.21.1). The surgery modalities compared include temperaturecontrolled radiofrequency tissue volume reduction of the soft palate, UPPP, maxillomandibular advancement osteotomy, and radiofrequency ablation (RFA). Only one trial (Woodson 2003) included patients who had neither prior surgery nor prior CPAP. The other trial (Vicini 2010) excluded patients with prior surgery but did not report on prior CPAP use. The remaining studies either explicitly or implicitly were biased in that the patients receiving surgery had already failed or refused CPAP or other nonsurgical interventions, in contrast with the patients who were being treated with CPAP.

Mean baseline AHI ranged from 5 to 80 events/hr across the studies. Most studies had a mean age above 45 years and a mean BMI ≤35 kg/m^2. The studies enrolled predominately male subjects (≥70 percent). Although Conradt 1998 included patients with craniofacial abnormalities, all other studies included patients with no important comorbidity. Study sample sizes ranged from 25 to 22,898 patients (total = 24,215 across studies). One study was rated quality A and the remainder rated quality C due to

inadequate reporting of eligibility criteria, inconsistent reporting, small sample sizes, and discrepancies between followup periods. Studies included patients with a wide range of baseline AHI, but were heterogeneous in the severity of OSA within each study, thus limiting the applicability of most studies. Studies mostly included patients with BMI ≤35 kg/m^2.

Mortality (Balk [AHRQ] (2011) Appendix D Tables 5.21.2 & 5.21.3)

Two retrospective studies evaluated the effects of UPPP compared with CPAP on long-term survival.[257,260] o studies evaluated any other objective clinical outcomes.

Weaver 2004 compared 20,826 patients using CPAP to 2072 patients who had UPPP. All patients were followed for at least 6 years in a database at Veterans Affairs medical facilities. In addition to UPPP, about one-quarter to one-third of patients also received tonsillectomy, septoplasty, or, turbinate procedures, and about 2 percent of patients had tracheotomy or tongue procedures, each. After adjusting for age, sex, race, date of initial treatment, and comorbidities, the study found a higher mortality in the CPAP than in the UPPP group at all time periods throughout the study. The adjusted hazard ratio of death for CPAP versus UPPP was 1.31 (95 percent CI 1.03, 1.67; P=0.03).

Keenan 1994 found no difference in age-adjusted 5 year survival between the cohorts of 275 patients who had received UPPP or CPAP. Compared with those who used CPAP, patients who received UPPP had a significantly lower BMI (30 versus 36 kg/m^2; P<0.001) and a higher arousal index at baseline (25 versus 20 events/hr; P<0.01). However, the results are difficult to interpret as the followup for UPPP patients was significantly longer than that of patients receiving CPAP (43 versus 28 months; P<0.001).

Apnea-Hypopnea Index or Respiratory Disturbance Index (Balk [AHRQ] (2011) Appendix D Table 5.21.4)

One RCT[250] and four prospective studies reported outcomes on AHI, RDI, or a combination of AHI and RDI.[251,254,256,258] Vicini 2010 randomly assigned 50 patients to either maxillomandibular advancement or autoCPAP therapy and compared treatment effects at the end of 12 months. They found no statistically significant difference in AHI between groups (-48.7 versus -44 events/hr; P=0.21).Ceylan 2009, a prospective, nonrandomized comparative study, reported AHI. This study compared single-stage, multilevel temperature-controlled radiofrequency tissue volume reduction of the soft palate and base of the tongue to CPAP and found no significant difference in changes in AHI between the two groups after 12 months followup. The other studies reported RDI. As each study evaluated a different surgical technique, each study result is described separately.

After 3 months followup, Conradt 1998 found essentially the same large declines in RDI (-54 events/hr) after maxillomandibular advancement osteotomy and CPAP in prospectively followed cohorts of patients. Lin 2006, in a retrospective analysis comparing extended uvulopalatoplasty (UPP) found a significantly larger improvement in RDI in patients on CPAP as compared to surgery (-63 versus -32 events/hr; P<0.001). However this study also had significant differences between groups in several baseline characteristics including age (51 yr versus 45 yr; P=0.005), BMI (28.1 versus 26.4 kg/m^2; P=0.025), RDI (65.3 versus 43.6 events/hr; P<0.001), and ESS (14.1 versus 11.8; P=0.005).

Katsantonis 1988, using retrospective data, compared patients who had UPPP with those who were treated with CPAP (other reported interventions are not included from this retrospective study). They analyzed 98 mostly male patients with moderate to severe OSA who had UPPP and a sample of 44 of 138 patients who received CPAP. Patients were categorized as good responders (>50 percent improvement in

AHI and 85 percent improvement in severity index [number of abnormal breathing events/hr with <85 percent oxygen saturation]), poor responders (<50 percent decrease in AHI and severity index), and moderate responders (those in between).

After 18 months followup, the study reported 100 percent of patients using CPAP were good responders. In contrast, of those who received UPPP , 38 percent were good responders, 34 percent moderate responders, and 28 percent poor responders. The study was rated quality C due to poor reporting and a lack of reported eligibility criteria.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.21.5)

Two RCTs,[247,250] one prospective nonrandomized study,[252] and three retrospective studies[254,258,259] compared various surgical techniques with CPAP for the treatment of OSA None of the studies found statistically significant differences in ESS values between surgery and CPAP.

Objective Sleepiness and Wakefulness Tests (Balk [AHRQ] (2011) Appendix D Tables 5.21.2 & 5.21.6)

Two studies reported objective sleepiness using the Multiple Sleep Latency Test.[253,255] Zorick 1990 compared UPPP with CPAP and found no statistically significant improvement in excessive daytime sleepiness (net difference -4.5; P<0.05). Anand 1991 found that 30 percent of UPPP patients and 41 percent of CPAP patients increased their Multiple Sleep Latency Test score by at least 3 minutes. No statistical analysis was reported.

Other Sleep Study measures (Balk [AHRQ] (2011) Appendix D Tables 5.21.6 & 5.21.7)

Two studies reported REM and stage 3 and 4 sleep as a percentage of total sleep time,[251,253] one study reported arousal index,[251] and one study reported minimum oxygen saturation.[254] Zorick 1990 found that after 6 weeks of followup there was a statistically significant relative increase in REM and stage 3 or 4 sleep in the CPAP as compared to the UPPP group. Conradt 1998 found no difference in arousal index, sleep efficiency, REM sleep, or stage 3 and 4 sleep 3 months after maxillomandibular advancement osteotomy or CPAP. Ceylan 2009 found a nonsignificant difference of 2.7 percent between temperature-controlled radiofrequency tissue volume reduction of the soft palate and CPAP.

Quality of Life (Balk [AHRQ] (2011) Appendix D Tables 5.21.6 & 5.21.8)

Three studies, two of which compared temperature-controlled radiofrequency tissue volume reduction of the soft palate and base of the tongue to CPAP and one which compared extended uvulopalatoplasty to CPAP, found no difference between groups in all domains of SF-36**.**[247,252,258] Both Woodson 2001 and Woodson 2003 found no difference in FOSQ after 2 to 3 months of follow up.[247,252]

Study Variability

One study reported a subgroup analysis. Keenan 1994 retrospectively analyzed data from 208 patients with OSA over a 6 year period. Patients were stratified by apnea index. For patients with an apnea index >20 events/hr, the study found no significant differences in cumulative survival between UPPP and CPAP. No data were reported separately for those patients with apnea index ≤20 events/hr.

Summary

Of 12 studies comparing surgical modalities with CPAP, two were RCTs. The quality A trial was the only unbiased comparison of surgery and CPAP (patients had previously received neither treatment) did not find statistically significant differences in ESS and quality of life measures between patients with mild to moderate OSA (AHI 10 to 30 events/hr) who had temperature-controlled radiofrequency tissue volume

reduction of the soft palate and those who had CPAP at 2 months followup. Similarly, the other trial found nonstatistically significant differences in AHI and ESS in patients with severe OSA (AHI ≥30 events/hr) between maxillomandibular advancement osteotomy and CPAP.

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 (and not always in a consistent manner, i.e., the surgical group could have a higher AHI than the CPAP group in one study and vice versa in another study). The reported findings on sleep study and quality of life measures were heterogeneous across studies.

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 to draw useful conclusions comparing surgical interventions with CPAP in the treatment of patients with OSA at this time. Therefore the strength of evidence is insufficient to determine the relative merits of surgical treatments versus CPAP.

Comparison of Surgery and Mandibular Advancement Devices

One parallel design RCT across three publications compared the effects of a MAD with uvulopalatopharyngoplasty (UPPP) in patients with mild to moderate OSA and no other significant comorbidities.[261-263] Subjects were 95 men with a mean age of 50 years and a mean BMI of 27 kg/m^2. Mean baseline AHI was 19 events/hr. Patients were followed for up to 4 years. Results at 12 months showed that 80 percent of patients using MAD achieved a decrease in AHI of ≥50 percent compared to 60 percent who had UPPP ($P<0.05$). A statistically significant reduction in AHI was also observed in the MAD group as compared to the UPPP group at 4 years (-11 versus -6 events/hr; net difference -5 events/hr; 95 percent CI [estimated] -9, -1; $P<0.001$ [P analyzed for final values, not net difference]). Objective clinical outcomes were not evaluated. This study was rated quality B. This study is applicable mainly to patients with apnea index scores between 5 and 25 events/hr. It was restricted to patients with sufficient number of teeth to anchor the mandibular devices in place. With only one study that evaluated only AHI, the strength of evidence is insufficient regarding the relative merit of MAD versus surgery in the treatment of OSA.

Comparison of Drug Therapy and Control

Seven RCTs compared different drug treatments with controls (Balk [AHRQ] (2011) Appendix D Table 5.23.1).[205,264,265,265-268] All but Ryan 2009 were crossover trials. The studied drugs included mirtazapine, xylometazoline, fluticasone, paroxetine, pantoprazole, steroid plus CPAP (versus CPAP alone), acetazolamide, and protriptyline (Appendix D Table 5.23.1). All trials used placebo controls except for Ryan 2009, which used CPAP without steroid as a control.

Mean baseline AHI ranged from 10 to 36 events/hr. Study sample sizes ranged from 10 to 81, with a total of 231 across the studies. One study was rated quality A, five were rated quality B, and one was rated quality C. Whyte 1988 was rated quality C because of its lack of exclusion criteria and a washout period.

Study Results (Balk [AHRQ] (2011) Appendix D Tables 5.23.2-5.23.9)

As each study evaluated a different drug therapy, each study is described individually. No study evaluated objective clinical outcomes.

Carley 2007[264] compared two mirtazapine doses (4.5 mg and 15 mg) to control. Both groups on mirtazapine had a significantly lower AHI than the control group (P=0.004). The 15 mg mirtazapine group had a significantly lower arousal index (P=0.02), higher sleep efficiency (P=0.05), and lower REM sleep percentage (P=0.04) than the controls; however, the 4.5 mg group did not differ from the control group in these outcomes. Neither drug group differed from controls in slow wave sleep, minimum oxygen saturation, or Stanford Sleepiness Scale score.

Clarenbach 2008[269] did not find a difference in AHI, ESS, arousal index, sleep efficiency, slow wave sleep, or REM sleep between the xylometazoline group and control.

Kiely 2004[265] found a significantly lower AHI in the fluticasone group as compared to the placebo group, both in patients with an AHI ≥10 events/hr (median difference = -6.5 events/hr; P<0.05) and in patients with an AHI ≥5 events/hr (median difference = -5.6 events/hr; P=0.01). The drug group did no differ from controls in REM sleep or minimum oxygen saturation.

Kraiczi 1999[266] found a lower AHI in the paroxetine group than in the control group (95 percent CI -17.9, 0.6; P=0.02113FP13F[i]). The drug group did not differ from the control group in sleep efficiency, slow wave sleep, REM sleep, or Comprehensive Psychopathological Rating Scale (CPRS) score.

Whyte 1988 did not find a significant difference in AHI, arousal index, sleep efficiency, slow wave sleep, REM sleep, or minimum oxygen saturation between acetazolamide and control, or between protriptyline and control.

Suurna 2008 found a lower ESS score in the pantoprazole group as compared to control (difference = -0.5; 95 percent CI -0.98,-0.02; P=0.04), but no significant difference in FOSQ score (difference = 0.06; 95 percent CI -5.3, 0.1; P=0.06).

Ryan 2009 did not find a statistically significant difference in ESS, SF-36 score, or Mini Rhinoconjunctivitis Quality of Life Questionnaire score between the steroid plus CPAP and dry CPAP groups.

Study Variability

None of the studies reviewed performed subgroup analyses. As the drugs used were different in each study, we were not able to examine differences with regard to patient characteristics across studies.

Summary

Seven trials compared different drugs with control for the treatment of patients with OSA. Due to the heterogeneous nature of the drugs examined and the different findings reported, it is not possible to draw any general conclusions about the effects of drugs on the treatment of OSA at this time. As only one study examined each drug, the strength of evidence is insufficient to evaluate the effectiveness of any individual drug for the treatment of OSA.

Comparison of Atrial Overdrive Pacing and Control or CPAP

Two crossover trials examined atrial overdrive pacing in the treatment of OSA (Balk [AHRQ] (2011) Appendix D Table 5.24.1).[270,271] Both trials evaluated patients who had pacemakers that had been implante for an underlying arrhythmia. The pacemakers were capable of specific scheduling for overnight atrial overdrive pacing. Melzer 2006 compared atrial overdrive pacing of 75 beats per minute with sham pacing of 45 beats per minute in 20 patients.[270] Simantirakis 2009 compared atrial overdrive pacing

(pacing at 14 beats per minute greater than spontaneous mean nocturnal heart rate) with CPAP (and no atrial overdrive pacing) in 16 patients.[271] The mean baseline AHI in the trials were 27[270] and 49 events/hr.[271] Melzer 2006 excluded patients with other ventilatory OSA interventions. This study was rated quality A. Simantirakis 2009 excluded those with left ventricular dysfunction or heart failure. This study was rated quality B, as it did not provide a description of how CPAP pressure was titrated and had a small sample size. The studies are applicable to patients who already have implanted pacemakers without cardiac dysfunction.

Objective Clinical Outcomes

No study evaluated objective clinical outcomes.

Apnea-Hypopnea Index (Balk [AHRQ] (2011) Appendix D Table 5.24.2)

Both trials provided data on AHI outcomes. Melzer 2006 did not find a statistically significant difference between atrial overdrive pacing and control. Simantirakis 2009 did not find a statistically significant difference between atrial overdrive pacing and CPAP.

Epworth Sleepiness Scale (Balk [AHRQ] (2011) Appendix D Table 5.24.3)

Simantirakis 2009 did not find a statistically significant difference between atrial overdrive pacing and CPAP in ESS score.

Other Sleep Measures (Balk [AHRQ] (2011) Appendix D Table 5.24.4)

Melzer 2006 did not find a statistically significant difference between atrial overdrive pacing and CPAP in slow wave sleep or REM sleep.

Study Variability

No study reported subgroup analyses with respect to the comparative effect of atrial overdrive pacing versus no pacing in terms of patient characteristics (age, sex, race, weight, bed partner, and airway) or severity of OSA.

Summary

Two trials examined atrial overdrive pacing in the treatment of OSA. Each trial used a different control comparator (sham pacing or CPAP). Neither trial reported a benefit in sleep study measures with atrial overdrive pacing as compared to the control. As each comparison was unique and the respective sample sizes small, the strength of evidence is insufficient to determine the effect of atrial overdrive pacing on sleep apnea signs and symptoms.

Comparison of Other Interventions and Controls

Five trials, each a parallel design, compared a variety of miscellaneous interventions with different controls.[272-276] Freire 2006 compared acupuncture to sham acupuncture. Wang 2009 compared auricular plaster to vitamin C. Cartwright 1991 compared a tongue-retaining device, a posture alarm, or a combination of the two with no intervention. Krakow 2006 compared nasal dilator strip therapy to no treatment. Grunstein 2007 compared bariatric surgery to another weight loss protocol.

As each study evaluated different, unrelated interventions, each study is described individually. No study evaluated objective clinical outcomes. Each study's applicability is suggested by its eligibility criteria.

Tongue-Retaining Device, Posture Alarm, or Combination Versus No Treatment

Cartwright 1991 compared a tongue-retaining device, posture alarm, or a combination of these therapies against no intervention in an RCT. [273] The study consisted of 60 male patients with positional sleep apnea and an AHI >12.5 events/hr. Neither of the devices nor their combination resulted in significantly different changes in AHI compared to control. From reported data, the odds ratio for achieving an AHI <5.5 events/hr was nonsignificant for each intervention.

The study was rated quality C due to unclear reporting, the lack of an appropriate statistical analysis, and an inadequate description of the interventions.

Bariatric Surgery Versus Routine Management

The Grunstein 2007 study was a nonrandomized comparison of bariatric surgery (gastric bypass, vertical banded gastroplasty, or gastric banding) and routine obesity management (consisting of diet and exercise advice and behavior modification).[272] Patients were included if they had a BMI ≥38 kg/m^2, and were excluded if they had undergone previous bariatric surgery. Patients had previously responded to a baseline questionnaire that they had frequent apneas. There were 694 total patients, with 382 patients in the bariatric surgery group and 312 in the control group. Patients were able to choose which treatment they would receive, and were computer-matched to patients in the other treatment group. No baseline data were collected on OSA severity.

After 2 years of followup, patients who had bariatric surgery experienced significantly less persistence of sleep apnea, as defined by fewer symptoms noted on a followup questionnaire (OR = 0.16; 95 percent CI 0.10, 0.23; P<0.001).

The study was rated quality C due to a lack of randomization, a high dropout rate, and dissimilar baseline characteristics between groups.

Nasal Dilator Strip Versus No Treatment

Krakow 2006 compared nasal dilator strip therapy with no treatment in an RCT.[275] Enrolled were 80 patients with nonsevere OSA. The nasal dilator strip group had a significantly better Pittsburgh Sleep Quality Index score (difference = 2.7; P<0.001), better Quality of Life Enjoyment and Satisfaction Questionnaire score (difference = 0.46; P=0.01), improved Insomnia Severity Index score (difference = -0.78; P<0.001), and better FOSQ score (difference = 1.3; P<0.02) than the no treatment group. The study was rated quality C due to the lack of an objective measurement of sleep apnea and unclear reporting.

Acupuncture Versus Control

Freire 2006 compared acupuncture (10 weekly sessions including needle manipulation) to sham acupuncture (10 weekly sessions with needles, but not at acupuncture sites, and no manipulation) or no treatment.[274] Patients (N=26) were included if they had not received acupuncture before and had an AHI of 15-30 events/hr, and were excluded if they had a history of CPAP or oral device use. The mean baseline AHI was 19 events/hr.

Treatment with acupuncture resulted in statistically significant net differences in AHI compared with both sham acupuncture (net difference = -13 events/hr; 95 percent CI -21, -5; P<0.05) and no treatment (net difference = -18 events/hr; 95 percent CI -29, -6; P=0.002). The acupuncture group had no significant difference in ESS scores as compared to sham acupuncture, but a significant net reduction in ESS compared with no treatment (net difference = -5.9 events/hr; 95 percent CI -11.1, -0.7; P<0.05). Patients

in the acupuncture group did not differ in sleep efficiency, REM sleep, or SF-36 total score as compared to the other groups.

The study was rated quality C due to a small sample size, an unequal number of dropouts per group, and lower quality of life measurements in the controls at baseline compared with the active treatment group.

Auricular Plaster Therapy

Wang 2009 compared auricular plaster therapy with vitamin C in 45 males with OSA.[276] After 10 days of followup, the group randomized to auricular plaster group was found to have a lower AHI than the vitamin C group (net difference = -13 events/hr; 95 percent CI -18, -8). The study was rated quality B due to incomplete reporting.

Summary

Five studies examined miscellaneous interventions compared with controls in the treatment of OSA. Four of these studies were rated quality C and one was rated quality B. No consistent effects on sleep study measures were reported across different interventions as compared to inactive controls or routine treatments. As each intervention was studied only once, the strength of evidence is insufficient to determine the benefit of each intervention compared with control in the treatment of patients with OSA.

Adverse Events

Across all studies and interventions, the reporting of adverse events (or side effects) was sparse. Almost no RCT was sufficiently large to adequately compare rates of adverse events between different interventions, particularly if analysis is focused on RCTs of actual treatments (as opposed to RCTs with placebo or sham treatment groups). Furthermore, as will be described, the types of adverse events related to different categories of treatments vary considerably, further hampering direct comparisons.

Adverse events are, therefore, evaluated here based on the cohorts of patients who received specific treatments within RCTs (e.g., CPAP), rather than by the RCT comparisons (e.g., CPAP versus surgery). In addition, based on discussions with the Technical Expert Panel about the likely dearth of RCTs and other comparative studies of surgical treatments, it was also decided that adverse events data would be collected from prospective or retrospective cohort studies of surgical treatments for OSA with at least 100 patients. It should also be noted that the summary tables include adverse event rate data for only those findings study authors reported to be (or we determined to be) clinically important and/or severe outcomes. Less clinically significant adverse events were listed for each intervention in table footnotes. In addition, data concerning a lack of adverse events (e.g., no perioperative deaths) or general results (e.g., —major adverse events‖) were extracted only from studies with at least 100 patients. We did not collect data on adverse events from control, placebo, or sham treatments.

Of the 143 otherwise eligible comparative studies of two or more interventions for OSA and 13 surgical cohort studies with at least 100 treated patients, 19 comparative studies, and 12 surgical cohort studies reported adverse event data.

Positive Airway Pressure Devices (Balk [AHRQ] (2011) Appendix D Table 5.25.1)

Only six trials of CPAP reported adverse event (or side effect) data.[171,173,179,183,207,259] Trials enrolled between 21 and 73 patients using CPAP. Four of the trials compared different CPAP devices to each other; the remaining two compared CPAP to other interventions. Four studies evaluated CPAP (two compared

nasal to oral CPAP) and three evaluated autoCPAP (one compared humidified to nonhumidified autoCPAP). No study of other types of CPAP reported adverse event data.

The most commonly reported adverse event was claustrophobia. In four studies with 1 to 4 month followups, claustrophobia was reported by one to three patients, representing 1.4 to 23 percent of patients. Epistaxis was reported among patients in two studies: two of 22 patients (9 percent) using nonhumidified autoCPAP, but none of the patients using humidified autoCPAP, and two of 17 patients (12 percent) using nasal CPAP (but implied no patients using oral CPAP). Excessive pressure or pressure intolerance was reported in two studies: five of 55 patients (9 percent) on CPAP in one study and 2 (4 percent) on autoCPAP in the other. Major or excessive oral dryness was reported in two studies of oral CPAP, with one study noting 11 patients (52 percent) and the other 3 (14 percent), complaining of excessive oral dryness. Only one trial reported excessive nasal dryness, with 2 (12 percent) patients noting the complaint. Severe gum pain was also reported in one study of oral CPAP in 3 of 21 (14 percent) patients. A major excess of salivation and sore gums or lips were reported in one trial of oral CPAP in 1 (5 percent) and 2 (10 percent) patients, respectively. Other more minor adverse events reported included skin irritation, nasal irritation or obstruction, dry nose or mouth, excess salivation, minor or moderate sore gums or lips, minor aerophagia, abdominal distension, minor chest wall discomfort, pressure discomfort, and transient or minor epistaxis.

Generally, about 5 to 15 percent of patients reported specific adverse events they considered to be a major problem while using CPAP. However, no study reported a severe adverse event that would not resolve quickly upon discontinuing CPAP or that may be amenable to alleviation with ancillary treatments (such as humidification).

Mandibular Advancement Devices (Balk [AHRQ] (2011) Appendix D Table 5.25.2)

Only five RCTs of MAD reported adverse event data.[212,216,225,226,262] The trials included between 19 and 48 patients using these devices. Four studies evaluated custom-made devices, with ranges of maximal mandibular advancement from 50 to 100 percent and 2-5 mm interdental clearance, and one study evaluated the Snore-Guard™ (mandible set at 3 mm posterior to maximal acceptable advance with a 7 mm opening). Four studies lasted 1 to 4 months, while one study followed patients for 4 years. All major adverse events were related to tooth, mouth, or jaw pain or damage. In one study with a device with 80 percent mandibular advancement, 3 of 48 patients (6 percent) had a dental crown damaged. One of 31 patients using a maximal advancement device had loosening of teeth. Temporomandibular joint (TMJ) or jaw pain was reported in one patient each of four studies (between 2.2 and 5.2 percent of patients). An aphthous ulcer due to acrylic polymer allergy was also reported by one patient (2.2 percent) in one study. Other more minor adverse events included a sensation of pressure in the mouth, transient morning mouth and TMJ discomfort or sounds, minor sore teeth or jaw, transient mild mucosal erosions, minor excessive salivation, tooth grinding, and sleep disruption.

Overall, about 2 to 4 percent of patients complained of jaw or temporomandibular joint pain with MAD. There were an insufficient number of patients evaluated to determine whether the likelihood of jaw pain might be related to the degree of jaw opening. More permanent damage, namely dental crown damage, occurred in 6 percent of patients in one study, but was not reported in other studies. One patient had an allergic reaction to acrylic polymer.

Airway Surgery Interventions

Ten eligible studies of UPPP (and related surgeries), two studies of RFA, six studies of combinations or other surgeries, and one study of palatal implants alone reported adverse events (or complications).

Uvulopalatopharyngoplasty (Balk [AHRQ] (2011) Appendix D Table 5.25.3)

Ten eligible studies reported adverse events related to UPPP.[124,245,255,262,277-282] The largest cohort study analyzed 3,130 patients who received UPPP with or without tonsil, nasal, or turbinate surgery. The remaining nine studies ranged in sample size from 18 to 158 patients and generally included similar surgeries, or tracheostomy, or, in one study, osteotomy; one study performed laser assistant uvulopalatoplasty.

Perioperative death (up to 30 days of surgery) was reported by five studies and ranged from 0/158 to 2/132 (1.5 percent) of patients. The largest cohort (Kezirian 2004) reported 7/3,130 (0.2 percent) perioperative mortality. This study reported serious complications (including death) in 51/3,130 (1.6 percent) of patients. These complications included reintubation (17 patients), pneumonia (11), hemorrhage (9), cardiovascular complication (8), emergency tracheotomy (7), and mechanical ventilation for >48 hr (6). No patients suffered deep vein thromboses or kidney failure.

Across studies, reintubation was reported in 0.5 to 5.2 percent of patients (three studies, with no long-term sequelae in one study), pneumonia in 0.4 and 1.5 percent of patients (two studies), major hemorrhage in 0.3 to 5.5 percent of patients (eight studies), and tracheotomy in 0.2 to 5.6 percent of patients (four studies). Other major perioperative adverse events reported across studies included respiratory events (six patients), substantial laryngeal edema (two patients), pulmonary edema (one patient), and postextubation asystole (one patient). Individual studies reported no perioperative airway complications, abscesses requiring surgical interventions, or rehospitalizations (in 134 patients), or infections or arrhythmias (in 101 patients).

Adverse events reported over the long term (3 months to 1, 4, or 7.25 years) included difficulty with speech or change in voice (0.6 to 15 percent; three studies); velopharyngeal incompetence (11 and 12 percent; two studies); infection requiring surgical intervention (0 and 11 percent; two studies); difficulty swallowing (5 to 10 percent; three studies); pronounced nasal regurgitation of fluids (8 percent; one study); and breathing difficulties, nasal synechiae, loss of taste, and tracheal stenosis in 5 percent of patients or fewer. One study with 158 patients reported no long-term sequelae from complications were reported.

Other adverse events (or side effects or harms) reported by studies included: unplanned medications, mild transient pain and swallowing difficulty, postoperative (minor) hematomas or ulcerations, mild bleeding, mild and transient tongue deviation, transient swelling sensation, pharyngeal dryness, nasal regurgitation (transient), increased mucus secretion, gagging, cough, infection (self-limited), antibiotic-related diarrhea, burning sensation, anosmia, temporary vocal quality change, and difficulty singing, playing saxophone, etc.

Radiofrequency Ablation (Balk [AHRQ] (2011) Appendix D Table 5.25.4)

Two studies reported adverse events following radiofrequency ablation of the tongue base (or other sites in one study) in 497 and 73 patients, respectively.[252,283] The larger cohort experienced no long-term complications (after 8 days) and the following short-term adverse events: dysphagia requiring hospitalization (4 patients; 0.8 percent), tongue base ulceration requiring surgical intervention (3 patients; 0.6 percent), and in one patient each (0.2 percent) soft palate mucosa ulceration requiring surgical

intervention, temporary hypoglossal nerve palsy, and tongue base abscess requiring surgical intervention. The smaller study reported that seven patients (10 percent) had an infection or cellulitis during 6 weeks of followup, four patients (5.5 percent) had severe, suppurative tongue base infections (two of which required surgical intervention), and one patient (1.4 percent) had a tongue abscess.

Other adverse events (or side effects or harms) reported by studies included: unplanned medications, mild transient pain and swallowing difficulty, postoperative (minor) hematomas or ulcerations, mild and transient tongue deviation, transient swelling sensation, and asymptomatic fibrotic narrowing.

Combination or Various Surgeries (Balk [AHRQ] (2011) Appendix D Table 5.25.5)

Six studies reported adverse events in patients who received a variety of other surgeries.[280,284-287] These included combinations of UPPP and geniotubercle advancement, hyoid suspension, maxillary and/or mandibular osteotomy, and tongue RFA, and multilevel surgeries without UPPP, or stepwise multilevel surgeries.

The studies analyzed between 64 and 233 patients. Only one study specifically reported on perioperative death, noting that no deaths occurred. Two studies reported no major complications, though one also reported five patients (4 percent) with Pillar implant extrusion requiring removal and replacement, two patients (1.6 percent) with turbinate bone exposure, and one patient (0.8 percent) with nasal septum perforation, tongue mucosal ulceration, and hypoglossal nerve weakness lasting less than 1 month. With the exception of the smallest study, all other adverse events were reported in <2 percent of patients, including undescribed bleeding (1.9 percent), new onset atrial fibrillation (1.9 percent), transient nerve paralysis (1.4 percent), bleeding requiring anesthesia (1.3 percent), hypoglossal nerve paralysis (0.7 percent), and new unstable angina (0.5 percent). The largest study reported no long-term speech or swallowing problems and another study reported no airway complications, abscesses requiring surgical interventions, or rehospitalizations. The smallest study, examining stepwise surgery in 64 patients, had the highest reported complication rates, including paresthesia (not described; 17 percent), dysphagia (not described; 11 percent), voice change (3 percent), infection (not described; 3 percent), taste alteration (1.6 percent), wound dehiscence (1.6 percent), and transient palatal fistula (1.6 percent).

Other adverse events (or side effects or harms) reported by studies included: aspiration, neck seroma, transient dysphagia, transient tongue base ulceration, suture removal for foreign body reaction, and transient facial anesthesia.

Palatal Implants Alone (Balk [AHRQ] (2011) Appendix D Table 5.25.6)

One study reported adverse events following insertion of Pillar palatal implants in 50 patients.[243] During 1 week of followup, one patient had an undefined infection and two had extrusion of their implants. Other reported adverse events included sore throat and foreign body sensation.

Bariatric surgery (Balk [AHRQ] (2011) Appendix D Table 5.25.7)

One large study of 1,592 patients reported adverse events following bariatric surgery performed in patients with OSA.[272] Perioperative mortality was 0.21 percent, and 13 percent had bleeding, embolus and/or thrombosis, wound complications, deep infections, pulmonary, and/or other complications.

Weight Loss Diet (Balk [AHRQ] (2011) Appendix D Table 5.25.8)

One study evaluated a liquid, very low energy diet for 30 patients with OSA.[237] After 9 weeks, one patient had transient gout and two had transient elevated liver enzymes. Other reported adverse events included dizziness, dry lips, and constipation.

Drugs (Balk [AHRQ] (2011) Appendix D Table 5.25.9)

Three studies evaluating four drugs used for OSA treatment reported adverse events.[266,267,288] Acetazolamide resulted in the largest number of reported adverse events: any paresthesia in 8/10 patients and intolerable paresthesia in one patient. Protriptyline caused severe dry mouth requiring drug discontinuation in 2/10 patients and —visual upset,‖ urinary symptoms, and altered sexual potency with testicular discomfort in one patient each. Paroxetine use was associated with ejaculation disturbance (15 percent), decreased libido (10 percent), headache (10 percent), and constipation (10 percent). (Other reported adverse events included fatigue, mouth dryness, somnolence, and dizziness with both paroxetine and placebo, and sweating, nervousness, infectious pneumonia and Lyme disease during paroxetine treatment.) During zolpidem use, 1/72 patients (1.4 percent) experienced episodes of sleep walking.

Summary

Each type of OSA treatment carries its own set of potential adverse events. Based on the evidence reported among the eligible (mostly comparative) studies, with only a few exceptions, the only truly serious long-term adverse consequences from OSA treatments occurred among patients having oronasopharyngeal or bariatric surgery. These included perioperative death in up to 1.5 and 1.6 percent of patients undergoing UPPP in two studies. Most studies, however, reported no deaths. Other major postsurgical complications also included infections, 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, these events were found to occur in about 2 to 15 percent of patients (when reported). The largest studies (Kezirian 2004 with 3,130 UPPP surgeries and Stuck 2003 with 422 RFA surgeries) reported no long-term complications (not including perioperative death or cardiovascular complications).

All adverse events related to CPAP treatment were potentially transient and could be alleviated with either cessation of treatment or with adjunct interventions. Approximately 5 to 15 percent of patients reported specific adverse events they considered to be a major problem while using CPAP. These included claustrophobia, oral or nasal dryness, epistaxis, irritation, pain, or excess salivation. No adverse event with potentially long-term consequences was reported in patients receiving CPAP.

Among studies of MAD, four patients in two studies (with 79 patients total) incurred dental crown damage or loosening of teeth. TMJ or jaw pain was reported in about 2 to 4 percent of patients, although no study reported on the long-term consequences of these symptoms. It was also not clear whether the severity or frequency of TMJ symptoms was related to the degree of mandibular advancement or jaw opening.

Adverse events related to a very low energy weight loss diet or to various drugs were treatment specific. None appeared to be an adverse event with long-term consequences.

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

KQ#6. In OSA patients prescribed non-surgical treatments, what are the associations of pre-treatment patient-level characteristics with treatment compliance?

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for each Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 119-121). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

> To address this question, our literature search was restricted to longitudinal studies of at least 100 participants all of whom were prescribed nonsurgical OSA treatments and followed for at least 3 months. Only multivariable analyses of continuous positive airway pressure (CPAP) compliance were included. Because of the small number of potentially eligible mandibular advancement device (MAD) studies, all were included for review. Six studies met criteria. Five evaluated compliance with CPAP,[203,289-292] one compliance with MAD.[293]
>
> **Compliance with CPAP**
>
> Four of the five eligible studies were prospective cohort studies and one was a randomized control trial (RCT) of C-Flex™ versus fixed CPAP (Balk [AHRQ] (2011) Appendix D Table 6.1a-b).[203] The patients in the cohort studies were treated with either fixed CPAP, a variety of CPAP devices, or, in one study,[292] autotitrating CPAP (autoCPAP). The number of patients in the studies ranged from 112 to 1,103, and followup ranged from 3 months to 4 years. The studies were conducted mostly from the mid 1980s through the 1990s (or possibly later based on publication dates in two studies). All patients were enrolled at the beginning of their CPAP therapy. The demographics of the five studies were generally similar: a large majority of men, mean age around 50 years, mean BMI about 30 kg/m^2, and, in four of the studies, a mean AHI between 44 and 50 events/hr (Krieger 1996 apparently included patients with more severe OSA, as their mean AHI was 70 events/hr). Three of the studies (McArdle 1999, Krieger 1996, and Wild 2004) described an active followup program to improve CPAP usage. Hui 2001 described only an initial training session. The lone RCT (Pepin 2009) did not describe the initial ancillary care for CPAP usage (Appendix D Table 6.1a). In general, the studies are applicable to patients initiating CPAP whose AHI is greater than 30 events/hr.
>
> Each study defined compliance differently. Three studies used thresholds of 1, 2, or 3 hours of use per night (or voluntary discontinuation). The RCT used ―objective compliance,‖ which was measured by the device, but was not defined. The smallest study evaluated hours of use per night as a continuous variable.
>
> McArdle 1999, the largest study, provided a well documented, complete, and appropriate analysis, with no obvious selection or ascertainment biases; it was rated quality A. Wild 2004 suffered from some incomplete reporting and was rated quality B. The remaining three studies did not adequately define predictors, outcomes, or statistical analyses used, and were rated quality C (Appendix D Table 6.1b).

In McArdle 1999, 16 percent of patients discontinued CPAP at 1 year and 32 percent at 4 years. Krieger 1996 had somewhat better compliance; 14 percent withdrew from CPAP at a mean of 3.2 years. Pepin 2009 and Hui 2001 both found mean CPAP usage of about 5 hr/night at 3 months. Wild 2004 did not report compliance rates.

The four studies that evaluated baseline AHI as a predictor of compliance with CPAP all found a significant association such that a higher baseline AHI was associated with greater compliance. Krieger 1996 and McArdle 1999 found significant associations between an AHI>15 events/hr and greater compliance at 1-4 years (though the latter study found no significant association with an AHI threshold of 30 events/hr). The other two studies reported that a higher AHI (analyzed as a continuous variable) was associated with greater adherence or more hours of use per night at 1 and 3 months. In a secondary analysis, McArdle 1999 also found that AHI, analyzed as a continuous variable, was significantly associated with compliance across the range of AHI.

Three studies evaluated baseline ESS as a predictor of compliance. McArdle 1999, the quality A and largest study with the longest followup duration (4 yr), found that an ESS score >10 (and as a continuous variable) was associated with greater compliance. Wild 2004 found the same significant association, but Krieger 1996 did not find ESS to be an independent predictor, after adjusting for AHI and age. Only Krieger 1996 found that younger age (as a continuous variable) was associated with greater compliance. McArdle 1999 and Pepin 2009 did not find age to be an independent predictor.

Several potential predictors were evaluated by two studies each. In all cases the studies disagreed as to whether the factors were independent predictors of compliance. Snoring was a predictor in McArdle 1999, but not Hui 2001; lower CPAP pressure a predictor in Wild 2004, but not McArdle 1999; and higher BMI a predictor in Wild 2005, but not McArdle 1999. Pepin 2009 focused primarily on a sleep apnea-specific quality of life scale, and did not report on potential predictors evaluated by the other studies (except age). This study found that at 3 months, higher baseline mean oxygen saturation and greater sleepiness as measured by the Grenoble Sleep Apnea Quality of Life test were associated with greater compliance.

Summary
Across studies, there is a moderate strength of evidence that more severe OSA as measured by higher AHI is associated with greater compliance with CPAP use. There is a moderate strength of evidence 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 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.

Compliance with Mandibular Advancement Devices

Only one retrospective cohort with 144 patients met criteria for studies evaluating predictors of compliance with MAD (Balk [AHRQ] (2011) Appendix D Table 6.2a-b).[293] All patients received a custom-made MAD and received ―standard‖ education concerning its use, including adjustment of the device until it was workable. Patients were predominately male with a mean age of 51 years and a mean baseline AHI of 23 events/hr. Notably, 8 percent of the patients were nonapneic snorers with an AHI <5 events/hr. The study was rated quality C as only univariable analyses were

reported, predictors were poorly defined, and results were not clearly reported (Appendix D Table 6.2a). No explicit definition of compliance was provided. The study is generally applicable to patients initiating use of custom-made MAD.

The study failed to identify potential predictors that were significantly associated with MAD compliance. Variables that were analyzed included age, sex, occupation, ―marital situation,‖ snoring, feeling refreshed after sleep, daytime somnolence, driving problems, ESS, AHI, and CPAP failure or refusal (Appendix D Table 6.2b).

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

KQ#7. What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?

Balk [AHRQ] (2011) Key Findings

Key findings of the Balk [AHRQ] (2011) review for this Key Question are presented below.

Note: The text (including the reference numbers cited) indented below is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 121-124). In the Balk [AHRQ] review (2011), references can be found beginning on page 142. All tables from Appendix D of the Balk [AHRQ] (2011) report are included in this WA HTA report starting on page 186 (Appendix G). Tables that describe study characteristics (from the Balk [AHRQ] (2011) Appendix D) are included in this section. These tables are also available in Appendix G of this WA HTA report.

To address this question, we included only prospective comparative studies that enrolled more than 10 subjects per intervention arm and with 2 weeks or more of followup. We accepted any measure of compliance with a device, whether categorical (compliance versus no compliance) or continuous (time spent using device). We restricted the analysis to those interventions whose primary purpose was to improve compliance with treatment. We also included three studies that evaluated different care models (nurse led care versus others) for patients who had continuous positive airway pressure (CPAP) treatments that also reported compliance outcomes.

Eighteen studies met inclusion criteria (Balk [AHRQ] (2011) Appendix D Table 7.1).[174,288,294-309] All studies were RCTs, of parallel or crossover design, that evaluated outcomes of compliance with CPAP use. No trials evaluated measures to improve compliance with oral appliances or positional therapy. Fifteen studies examined a wide variety of interventions whose primary purpose was to improve compliance. For the purpose of this report, we categorized these interventions into four broad groups: 1) nine studies on extra support or education;[174,294,296-301,303] 2) three studies on telemonitoring care;[295,304,305] 3) one study on a behavioral intervention;[302] and 4) two studies on miscellaneous interventions.[288,306] The remaining three studies evaluated different care models (nurse led care versus others) for patients who had CPAP treatments.[307-309] These are reviewed separately.

Interventions To Improve Compliance With CPAP Use

Extra Support or Education

Nine studies evaluated the effects of extra support or education on the outcomes of compliance with CPAP use (Appendix D Tables 7.2 & 7.3).[174,294,296-301,303] The patients in these studies were treated with either fixed CPAP or autotitrating CPAP (autoCPAP). Eight studies enrolled new CPAP users or patients who were newly diagnosed with OSA. The remaining study (Chervin 2007) enrolled mostly (69 percent) people who, at study baseline, were already regular CPAP users. The studies were generally small with sample sizes ranging from 10 to 112 patients followed for 3 weeks to 1 year. Seven studies enrolled patients with similar demographics: mostly men, mean age between 45 and 63 years, mean BMI between 30 and 38 kg/m^2, and mean AHI between 42 and 58 events/hr. Wiese 2005 enrolled a nearly equal mix of men and women with mild OSA (mean AHI 9.3 events/hr). Therefore, these studies are applicable mainly to patients initiating CPAP with an AHI above about 30 events/hr and BMI greater than 30 kg/m^2. Of the nine studies, one was rated quality A, four quality B, and the remaining four quality C. Common quality issues in quality C studies included large dropout rates, different dropout rates between compared groups, and a more complete followup in the active intervention arm than the usual care arm.

Seven studies evaluated compliance as a continuous outcome (hours of use per night). These studies compared a variety of extra support protocols (e.g., telephone calls, videotape, literature) or education programs to usual support/care. Findings were generally inconsistent. Three studies showed that intensive support or literature (designed for patient education) significantly increased hours of CPAP use per night (by an average of 1.1 to 2.7 additional hours) compared with usual care.[174,294,297] However, the other four studies found no significant differences in hours of CPAP use per night between the intervention and control groups.[296,298,300,301]

Three studies reported categorical compliance outcomes using different definitions. Hui 2000 defined compliance with CPAP as at least 4 hours of use per night for more than 70 percent of the nights per week. The study found no significant difference in compliance rates between the augmented support and basic support groups. Smith 2009 defined compliance with CPAP use as 4 or more hours per night on at least 9 of each 14 nights (or at least an 80 percent use rate). This study found that the audio-based intervention packet significantly decreased the rate of shortterm (1 month) noncompliance compared with placebo intervention (11 versus 45 percent, respectively; $P<0.01$). However, there was no significant difference in noncompliance rates between groups at 6 month followup. It should be noted that all dropouts without CPAP use data were counted as nonadherent patients. Wiese 2005 analyzed return to clinic for 1 month followup as a measure of compliance among patients with mild OSA, and found that significantly more patients in the control group did not return to clinic for followup than patients in the group that received an educational videotape about CPAP use (51 versus 27 percent, $P=0.02$). The authors noted that the CPAP usage data from the device were available only for patients who returned to clinic for the followup, thus the usefulness of these data is limited.

Telemonitoring Care

Three studies evaluated the effects of telemonitoring care on the outcomes of compliance with CPAP use.[295,304,305] Telemonitoring care is a computer-based telecommunications system that functions as an at-home monitor, educator, and counselor to improve health-related behaviors. All studies enrolled new CPAP users or patients who were newly diagnosed with OSA. Studies were generally small with sample sizes ranging from 30 to 93 patients who were followed for 30 days to 2 months.

All three studies enrolled patients with similar demographics: mostly men, mean age between 45 and 59 years, mean BMI between 32 and 38 kg/m^2, and mean AHI 42 events/hr. Of the three

studies, one was rated quality B and two were rated quality C. All three studies compared telemonitoring care to usual care, and reported continuous compliance outcome as hours of CPAP use per night. Two studies found that telemonitoring increased hours of CPAP use per night (average 1.3 and 1.5 additional hours; P=0.07 and 0.08, respectively) compared with usual care at 2 month followup.[295,304] The third study did not find a significant difference in hours of CPAP use per night at 30 days between patients who received telemonitoring support and those who received usual care.[305] It should be noted, however, that patients in this study who had difficulties in using telemonitoring support were excluded from the analyses.

Behavioral Interventions

Only Richards 2007 (quality A) evaluated the effect of cognitive behavioral therapy (given to patients and their partners) on compliance outcomes in 96 patients (mean age 58 years old; mean AHI 26 events/hr) who were treated with CPAP.[302] This study found that cognitive behavioral therapy significantly increased hours of CPAP use per night compared with usual care (difference = 2.8 hours; 95 percent CI 1.8, 3.9; P<0.0001). This study also performed logistic regression modeling to explore predictors of CPAP compliance at 28 days, and found that psychological factors were not independent predictors of compliance. In addition, patients in the cognitive behavioral therapy group were 6.9 times more likely to comply with CPAP use (at least 4 hours per night) than the usual care group (95 percent CI 2.8, 18.2).

Miscellaneous Interventions

Bradshaw 2006, in a quality B study, compared the effects of an oral hypnotic agent (zolpidem 10 mg) to placebo or standard care (without a pill) in 72 patients newly using CPAP (mean age 38 years; mean AHI 43 events/hr). The hypnotic was prescribed with the purpose of improving CPAP compliance. The study found no significant differences in hours of CPAP use or categorical CPAP compliance outcomes (using three different definitions) between groups. In a quality B crossover trial, Massie 2003 compared CPAP with nasal pillows (designed to improve the comfort of the CPAP device) to CPAP with a regular nasal mask in 39 patients newly using CPAP (mean age 49 years; mean AHI 47 events/hr). The results showed that there was no significant difference in hours of CPAP use between the two different CPAP nasal appliances.

Summary

Fifteen RCTs examined a wide variety of interventions to improve compliance among mostly new CPAP users. Studies generally had small sample sizes with less than 1 year of followup. Results from these 15 studies were mixed. Compared to 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. Overall, there is a low strength of evidence that some specific adjunct interventions may improve CPAP compliance, but studies are heterogeneous and no general type of intervention (e.g., education) was more promising than others. In addition, no intervention has had its effect on compliance verified.

Studies That Evaluated Different Care Models for Patients Who Had CPAP Treatments

Three RCTs that evaluated different care models (nurse led care versus others) for patients who had CPAP treatments also reported compliance outcomes (Appendix D Tables 7.2 & 7.3).[307-309] Although all three studies compared a nurse-led model of care to usual care (by clinician), the components of both interventions and usual care differed across the studies. These interventions were not designed specifically to improve CPAP compliance and are thus evaluated separately.

A total of 467 patients were analyzed in these studies, which lasted from 3 months to 2 years. Of the three studies, one was rated quality B and two were rated quality C. Common quality issues in quality C studies include differential dropout rates between comparative groups and poor reporting of patient characteristics.

All three studies found no significant differences in CPAP compliance comparing nurse-led models of care to usual care.[307-309]

Summary
Three RCTs did not find improvements in patient compliance with CPAP with nurse-led care compared with usual care models. However, it should be noted that improved CPAP compliance was not a primary goal of the intervention but rather to evaluate whether nurse-led model of care would produce similar health outcomes compared to the usual care models. There is a low strength of evidence that nurse-led care does not improve CPAP compliance.

Subsequently Published Study Key Findings

No subsequently published studies were identified that met inclusion criteria for this Key Question.

Summary – Key Questions # 1 through #7

Balk [AHRQ] (2011) Summary

In the text and tables that follow, the evidence is summarized for each Key Question in the Findings section of the report, and for each category of intervention. In addition, the strength of the evidence for a variety of outcomes is presented. The indented text is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 121-124). References cited below can be found beginning on page 128 of Balk [AHRQ] (2011).

The following table summarizes the main findings that address the seven Key Questions in this systematic review. Of note, where interventions are not discussed (either diagnostic tests or treatments), this does not imply that the interventions were excluded from analysis (unless explicitly stated); instead, no studies of these interventions met eligibility criteria Discussion regarding the report and recommendations for future research follow.

Table 4. 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.[7] 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[8] 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[9] 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. • **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.

[7] Type II monitors are portable devices that record all the same information as PSG (Type I monitors).

Table 4. 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 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 56–85%. 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. • 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**: Thestrength 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.

[8] Type III monitors are portable devices that contain at least two airflow channels or one airflow and one effort channel.
[9] 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.

Table 4. 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	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. • 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.

Table 4. 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	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. • 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).

Table 4. 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	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. • 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 4. 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. • 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.

Table 4. 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. 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. • A total of 19 studies evaluated 21 other interventions including atrial overdrive pacing, 8 different drugs, palatal implants, cropharyngeal 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.

Table 4. 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	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). • 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.

Discussion and Limitations – Key Questions #1 through #7

Balk [AHRQ] (2011) General Discussion and Limitations

In the text that follows, a general discussion of the findings and the limitations of the systematic review is presented. The indented text is excerpted directly from the Balk [AHRQ] systematic review (2011, p. 132-138). References cited below can be found beginning on page 128 of Balk [AHRQ] (2011). All tables from Appendix D of the Balk [AHRQ] (2011) report are included starting on page 186 (Appendix G).

General Discussion

In theory, obstructive sleep apnea (OSA) should be relatively simple to diagnose and treat. Diagnosis involves determining whether the number of apnea and hypopnea events caused by upper airway obstruction during sleep exceed a given threshold, and continuous positive airway pressure (CPAP) is an effective treatment to minimize the apnea-hypopnea index (AHI) and improve symptoms.

Screening and Diagnostic Tests

Polysomnography (PSG), the standard test to diagnose OSA, requires one or more full night stays in a sleep laboratory. This can prove to be a difficult, inconvenient, and resource-intensive procedure requiring separate facilities and a full-time overnight skilled sleep technician. For many patients PSG may not be representative of a typical night's sleep given the foreign setting, lack of nighttime routine, attached equipment, and being under observation. This may be particularly true the first

night the test is given. Further complicating diagnosis, definitions of OSA vary widely, employing thresholds of AHI ranging from 5 to 15 events/hr. The American Academy of Sleep Medicine uses a threshold of 15 events/hr (with or without OSA symptoms) or 5 events/hr with OSA symptoms (unintentional sleep episodes during wakefulness; daytime sleepiness; unrefreshing sleep; fatigue; insomnia; waking up breath-holding, gasping, or choking; or the bed partner describing loud snoring, breathing interruptions, or both during the patient's sleep).[31,32] Variations in how PSG results are read and interpreted are also inevitable, possibly leading to inconsistent diagnosis of OSA across different sleep laboratories. In fact, as discussed in the Introduction, in-laboratory PSG has never been validated, and its true sensitivity and specificity in diagnosing OSA are not well documented.[26] Moreover, the AHI, which is used as the single metric to define OSA for insurance companies and in clinical settings, can vary from night-to-night and does not take into account symptoms, comorbidities, or response to treatment.[30]

Two approaches have been taken to reduce the resources involved in diagnosing (or ruling out) OSA: tests to screen for the likelihood of OSA and portable monitors instead of sleep laboratory PSG. Questionnaires and clinical prediction rules have been developed to screen patients with complaints suggestive of OSA to determine whether full testing is warranted. As addressed by Key Question #1, five questionnaires and 10 validated clinical prediction rules have been compared with PSG to test their accuracy to predict diagnosis with OSA (or severity of OSA). 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. The Berlin questionnaire's accuracy to screen for OSA (based on snoring, tiredness, and blood pressure), the only questionnaire that has been compared with PSG by two sets of investigators is supported by only a low strength of evidence. All other tests, including the commonly used STOP and STOP-Bang questionnaires, have not been adequately tested. To be of clinical value, such tests would need a very high sensitivity (to avoid failures to diagnose) and a sufficiently high specificity to minimize unnecessary testing of patients without OSA. Furthermore, the most clinically useful tests would be those that can be easily performed based on symptoms and signs easily obtainable during a physical examination. Screening tests that require specialized testing, such as pulmonary function tests, are likely of limited clinical value.

The second approach to reduce the resources involved in diagnosing OSA is the use of portable monitors developed for home or outside the laboratory. However, these monitors suffer similar deficiencies in validation. The addition of more recent studies has not substantially changed the conclusions from our Evidence-based Practice Center's 2007 Technology Assessment on *Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome*.[26] Numerous monitors have been evaluated across 93 eligible studies. Most of the tested portable monitors fairly accurately predict OSA, with high positive likelihood ratios and low negative likelihood ratios for various AHI cutoffs in laboratory-based PSG; however, it is unclear whether any of the portable monitors are sufficiently accurate to replace laboratory-based PSG. In general, portable monitors use many fewer ―channels‖ (specific physiologic measures) than typical 16-channel PSG. Different portable monitors use different numbers of channels and different specific channels. In general, across studies, monitors with more channels perform better than monitors with fewer channels. However, a lack of direct comparisons between portable monitors in their performance compared with PSG, and the large number and variety of monitors precludes a clear conclusion regarding which monitors perform best. None of the studies explicitly evaluated the monitors in patients with important comorbidities such as chronic lung disease, congestive heart failure, or with neurological disorders. The applicability of the evidence is unclear for these populations and for patients in whom there is a concern about central sleep apnea. Nevertheless, the evidence does suggest that the measured AHI (or similar measures) from portable monitors are likely to be biased and variable compared with PSG-derived AHI. Though, it is unclear to what degree this is due to the inaccuracies of the portable monitors, where the data loss ranged from 0-23 percent (not including the outlier study with a 78 percent data loss) and to what degree it is due to normal variation in AHI from

night-to-night or differences related to sleeping in different settings (home versus a sleep laboratory). Night-to-night variability may be addressed by repeated measurement over several nights, which may be better addressed by the portable monitors than the PSG due to the increased cost of repeated testing. The studies did not allow us to adequately assess any issues related to night-to-night variation. Additionally, among studies that evaluated a portable device in the sleep lab (simultaneous recording of signals by device and PSG) as well as home (nonsimultaneous recording of signals by device and PSG), the range of mean bias, sensitivity and specificity reported were not different for the two settings. More to the point, however, the only truly clinically valuable assessments of portable monitors would be tests of their predictive ability for clinical outcomes or response to treatment. No such studies have been conducted.

With the assumption that screening tests and/or portable monitors are of clinical value to triage patients for the treatment of OSA, further testing for OSA, or further testing for other conditions, the question arises as to what the algorithm for diagnosing these patients should be. Ideally, any proposed algorithms of phased testing should be compared to alternative algorithms, including full testing of all patients with symptoms suggestive of OSA. However, to date, no such studies have been conducted.

An important caveat to the evidence on the diagnostic and screening tests is that almost all the studies were performed at academic or research centers. It is not clear how the results generated in these settings, under research conditions, should be generalized to acceptance, use, and accuracy when used among the general population.

A related topic of interest to is the value of preoperative screening for OSA. As discussed in the Introduction, patients with OSA are at an increased risk of perioperative pulmonary and cardiovascular complications, but a large proportion of patients with OSA remain undiagnosed. Therefore, it has been suggested that all (or selected) patients undergoing anesthesia or surgery should routinely be screened or tested for OSA; however, the strength of evidence addressing this question is insufficient. One retrospective study of patients undergoing bariatric surgery found that a cohort of patients who had routine PSG had better perioperative outcomes than a cohort of patients who had PSG only if they were considered to be at an increased risk for OSA.

There are also no adequate studies that compared phased testing (simple tests followed by more intensive tests in selected patients) to full evaluation (PSG).

Apnea-Hypopnea Index as a Predictor of Clinical Outcomes

As described under Key Question #4, there is a high strength of evidence that higher baseline AHI is a strong and independent predictor of all-cause mortality over several years of followup. The association was strongest among people with severe OSA (AHI over 30 events/hr). The strength of evidence for the association between baseline AHI and other long-term clinical outcomes is either low (for incident diabetes mellitus) or insufficient (for other examined outcomes). These findings would seem to imply that individuals with OSA (and particularly severe OSA) should be treated aggressively to reduce their AHI. Unfortunately, as discussed below, there are almost no trial data to support that treatment of OSA and reduction of AHI improves clinical outcomes. Thus, the clinical value of these associations remains theoretical.

Treatment of Obstructive Sleep Apnea

As previously noted, there is a moderate strength of evidence that fixed CPAP is an effective treatment to minimize AHI and improve sleepiness symptoms. While the relevant trials are conclusive regarding the effects of CPAP on AHI and sleepiness measures, among over 40 trials of patients treated with CPAP or no treatment, none have reported long-term clinical outcomes. However, compliance with CPAP treatment is poor. Among the large cohort studies (with multivariable analyses for predictors of noncompliance), one reported 16 and 32 percent of

patients discontinued CPAP at 1 and 4 years, respectively;[289] 14 percent of patients had CPAP withdrawn for noncompliance at a mean of 3.2 years in a second study;[290] and patients used CPAP for an average of only about 5 hours per night at 3 months in two studies.[203,292] 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). There are numerous reasons why patients do not tolerate CPAP, including the difficulty of dealing with the equipment and its noisiness, mask discomfort, claustrophobia, oral and nasal irritation and dryness, and others. Among reviewed studies, up to 15 percent of patients described adverse events they considered to be major (though essentially none of the adverse events resulted in long-term consequences).

Because patients frequently do not tolerate fixed CPAP, many alternative treatments have been proposed and used, including alternative CPAP devices, devices to splint the oropharynx open during sleep, surgeries to minimize airway obstructions, and numerous other less commonly used or researched interventions. As discussed in the Introduction, several alternative CPAP devices have been designed to vary the pressure during the patient's inspiratory cycle or to titrate the pressure to a minimum level necessary to maintain airway patency. Other modifications include alternate masks, nasal pads, and humidification. The primary goal of these modifications is to improve comfort and thereby increase compliance with the treatment. The large majority of the trials that have compared different CPAP machines have compared autotitrating CPAP (autoCPAP) with fixed CPAP. There is a moderate strength of evidence that there are no clinical differences between autoCPAP and fixed CPAP, though, again, none of these trials evaluated clinical outcomes. Meta-analysis revealed patients using autoCPAP used the machines only 11 minutes longer on average than patients using fixed CPAP. A low strength of evidence suggests no statistically significant differences between the proprietary C-Flex™ and fixed CPAP. The strength of evidence is insufficient for other comparisons. Overall, the evidence does not support the use of one device for all patients. The decision as to which CPAP device to use may depend on numerous factors such as patient preference, specific reasons for noncompliance, cost, and others.

For patients who do not tolerate CPAP, or who refuse CPAP, or for whom CPAP is determined to be inappropriate, an alternative is the use of oral devices. A wide range of these devices have been designed with the goal of splinting open the oropharynx to prevent obstruction during sleep. The most common are mandibular advancement devices (MAD), which are generally worn intraorally and force the mandible to protrude forward several millimeters. Ten trials provide a moderate strength of evidence that MAD use is an effective treatment for OSA and results in a significant reduction in AHI and a clinical improvement in ESS score (though there is no trial evidence on clinical outcomes). By indirect comparison across trials, CPAP is more effective than MAD (e.g., meta-analysis summary net AHI reduction with CPAP is -33 events/hr; with MAD -12 events/hr). This was confirmed by trials that directly compared the two types of devices (CPAP versus MAD, net AHI reduction -8 events/hr). There is a moderate strength of evidence that CPAP is superior to MAD in the relief of sleep apnea signs and symptoms. However, given the issues with CPAP compliance, the decision as to whether to use CPAP or MAD will likely depend on patient preference. The adverse events associated with MAD are different than CPAP and stem from wearing an intraoral device. Most are self-limited, though tooth damage has been reported. Given the wide range of specific oral devices and the small number of trials that have made comparisons among them, the strength of evidence regarding their relative merit remains insufficient.

The third major alternative for OSA treatment includes surgical interventions to alleviate airway obstruction. The Key Questions addressed and the studies reviewed for the present report do not address what the indications for surgery may be, nor which surgery may be the most appropriate. As is often the case for surgical interventions, very few randomized trials have been conducted on surgical treatments for OSA. It is reasonable to assume that patients who choose surgery are

fundamentally different from those who are either not offered surgery or who choose a different (or no) treatment. These differences are borne out by the differences in baseline characteristics (including mean AHI and ESS, age, obesity, and others) found in the nonrandomized comparative studies reviewed herein that compare surgery to other treatments. Thus, indirect comparisons of surgical studies and CPAP or MAD studies are inappropriate and even though it was agreed upon to include retrospective comparisons of surgery and CPAP or MAD in this report, the value of these studies is highly suspect. Given these issues, the strength of evidence to determine the relative value of surgery as compared with no treatment, to CPAP, to MAD, or to alternative types of surgery remains insufficient.

Other interventions that have been tested in randomized trials (that met eligibility criteria for this review) include: weight loss programs, 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. There is a low strength of evidence (from three trials) that some intensive weight loss programs may be an effective treatment for OSA in obese patients; however, the strength of evidence is insufficient to determine the effects of any other of these alternative potential treatments for OSA.

Notably, little evidence exists across interventions supporting any OSA treatment as improving quality of life or neurocognitive function. Although trials did report improvements in these outcomes for CPAP, MAD, and surgical intervention, overall findings were inconsistent. Too few studies evaluated functional outcomes (such as driving skills) to formulate a conclusion.

Potential Different Treatment Effects in Different Patient Subgroups

For all the treatment comparisons, it is of particular interest which subgroups of patients may benefit most from which specific treatments. Unfortunately, the trials reviewed are almost completely silent on this issue. Very few trials reported subgroup analyses based on baseline or patient factors, such as OSA severity or demographics. For most comparisons, there were too few studies or the interventions examined were too heterogeneous (e.g., different types of MAD) to analyze potential differences across studies based on patient characteristics.

There were, however, a large number of trials of CPAP versus control (no treatment, placebo treatment, or sham CPAP). Subgroup meta-analysis of the trials based on their implied definitions of OSA (study eligibility criteria using different minimum AHI thresholds) failed to demonstrate any clear or consistent relationships between strictness of OSA definition and effectiveness of CPAP (to reduce AHI or ESS). Though statistical heterogeneity existed across trials, we were unable to find any patient- or intervention-level factors to explain the heterogeneity. Differences also existed based on whether the study design was parallel or crossover, and whether the control was no treatment or sham CPAP; however, there are no clinical implications of these findings. A large number of trials, conducted in a wide variety of settings, with a wide range of eligibility criteria, all found statistically and clinically significant improvements in AHI with CPAP. Based on this consistency (despite statistical heterogeneity), it is our conclusion that CPAP is effective (to minimize AHI) in all patients with OSA. The relative effectiveness in different populations may then be a moot point. The one exception might be patients with mild OSA (with AHI <15 events/hr). By definition, people with a low AHI cannot have as large an improvement in their AHI as people with severe OSA. A trial examining longterm clinical outcomes is necessary to make a definitive evaluation in this population.

The other intervention comparison for which cross-study evaluation of effectiveness in different populations may be possible is autoCPAP versus fixed CPAP, for which 21 trials qualified for review. No differences could be discerned based on patient characteristics except that the relative

improvement in ESS conferred by autoCPAP was larger in studies restricted to patients with AHI >20 or >30 events/hr, compared with those that included patients with less severe disease (or did not define a minimum AHI threshold).

Interventions to Improve Treatment Compliance

Given the difficulties with treatment compliance, an important question remains on how to improve usage of the interventions. Trials addressing this issue have investigated only interventions to improve CPAP compliance. Eighteen trials have each investigated unique interventions. These can be categorized as intensive education, telemonitoring, nursing care models, cognitive behavioral therapy, and miscellaneous interventions. Although, overall there is a low strength of evidence that some specific adjunct interventions may improve CPAP compliance, the strength of evidence is insufficient regarding any specific intervention. No trials have investigated interventions to improve compliance with any other devices.

Limitations

The present systematic review is subject to several important limitations. The most critical is the failure of the extant research to evaluate long-term clinical outcomes. Secondly, and in a similar vein, is the meagerness of evidence with respect to several Key Questions. Almost no study of diagnostic tests or treatments attempted to assess how results may vary in different subgroups of patients.

In general, intervention trials were of quality B or C, with few quality A studies. Followup durations tended to be very short, on the order of weeks to a few months, and are clearly insufficient for the appraisal of the treatment of a life-long disease whose clinical sequelae may take decades to develop. Study dropout rates were also frequently very high, particularly given the short duration of followup. In some studies, up to 40 percent of participants were lost to followup within weeks. The ability to meaningfully interpret the findings from these studies is clearly diminished. Other frequent methodological problems with studies included incomplete reporting and/or inadequate analyses. In particular, relatively few studies provided the net differences between interventions (in parallel design studies) or the difference between final values with appropriate adjustments for correlation (in crossover studies) with their confidence intervals and P values. Thus, for the large majority of studies we had to estimate the confidence intervals of the differences between interventions. Due to incomplete reporting or analyses, we also frequently had to estimate whether there was a statistically significant difference between interventions.

Patient compliance is an important outcome within this review, being a major outcome for treatment studies (particularly those that compare different devices) and the focus of two additional Key Questions. However, many studies measured self-reported compliance (either hours of use per night or nights of use per week). This raises the concern of inaccurate reporting, although there is no obvious reason to suspect biased reporting in favor of any specific device. In addition, a variety of definitions of compliance were used, complicating interpretation of results across studies.

Publication is a possible major concern for the validity of this review. The large majority of intervention studies (particularly those of diagnostic monitors and of mechanical treatment devices) were sponsored directly or indirectly by the manufacturers of the devices. Unfortunately, due to the magnitude of this review and limited time and resources to perform the review, we were not able to attempt a grey literature search to seek unpublished studies.

Nevertheless, this concern is tempered by a number of factors. Most of our conclusions were that the strength of evidence is either low or inadequate for interventions (minimizing the concern about publication bias among the current literature). The effects of CPAP and MAD on sleep

measures were generally large enough that conclusions about the effectiveness of these devices would be unlikely to change with the addition of unpublished trials. However, the reliance of the field on industry funding may partially explain the general lack of long-term trials with clinical outcomes, since these studies are not required by the Food and Drug Administration for device marketing.

Key Question #8. What is the evidence of cost implications and cost-effectiveness of sleep apnea diagnosis and treatment?

Findings

One systematic review and five individual studies met inclusion criteria (Balk [AHRQ] (2011) to address this key question. These studies assessed cost implications and cost-effectiveness of OSA from the perspective of either diagnosis or treatment, and in some cases, aspects of both. The types of analyses conducted varied and included:

- Cost-utility analysis
    - Measuring healthy years (e.g. quality adjusted life years, health years equivalent)
    - 1 systematic review and 2 individual studies
- Cost-analysis
    - No measure of benefits
    - 2 individual studies
- Cost minimization
    - Benefits found to be equivalent
    - 1 individual study

These studies were not comprehensive in covering all of the areas of interest to this report topic. Not all diagnostic tools, interventions, or outcomes were studied. The findings will be grouped when possible by the common issues addressed. Details of these studies can be found in the evidence tables of Appendix H and I.

Cost Implications; Social, economic, and healthcare utilization consequences of OSA

Two studies addressed these issues, both using a cost analysis approach.

One study (Jennum 2011) examined the direct and indirect costs of OSA using data from the Danish National Patient Registry. Every OSA patient (N=19,438) was randomly matched with four citizens by age, sex and socioeconomic status. Direct and indirect costs were obtained from two other governmental agencies. Findings noted lower employment rates for patients up to 8 years prior to diagnosis with further decline after diagnosis. Direct and indirect costs for patients compared to controls were €5257 vs €1396 ($p<0.0001$), corresponding to an annual mean excess cost per patient of €3860.

Another study (Tarasiuk 2008) used a case-control methodology to assess the effect of OSA on morbidity and health care utilization of middle-aged and older adults. Conducted in two sleep disorder centers in Israel, healthcare utilization from government databases provided cost information for the 2 years prior to PSG diagnosis of OSA for patients and age, sex, geographically, and primary care physician-matched controls. The 2-year total costs were 2.02 and 1.81 times as high in elderly and middle-aged patients with OSA, respectively, as in their controls (P<.001). Healthcare costs were 1.93 times as high in elderly subjects with OSA as in middle-aged subjects with OSA (P <.001). Elderly and middle-aged subjects with OSA had more consultations with medical specialists than controls (p<.01 and p<.001 respectively for age strata). Costs for drugs for elderly and middle-aged patients with OSA were 1.73 and 1.91 times as high, respectively, as for controls (P<.001). The main drug category prescribed to elderly and middle-aged patients with OSA were for CVD and respiratory conditions. In summary, elderly and middle-aged patients with OSA consumed approximately two times as much in the way of healthcare resources as their paired controls.

Cost Effectiveness

Sleep study alternatives to laboratory PSG

Three studies addressed this issue, each with a different type of analysis.
One study (Deutsch 2006) provided a cost-utility analysis of the cost-effectiveness of evaluations that employ either:

- split-night polysomnography (PSG); or
- unattended home partial sleep monitoring (UHPSM), and home CPAP titration using an auto-titrating device (CPAP autotitration).

These diagnostic modalities are compared with a conventional approach using full-night PSG.

- Split-night PSG is an attended, in-laboratory, overnight procedure during which sleep and breathing variables are recorded during the first 2 hours of the sleep period, and, if criteria are met, CPAP titration is performed during the remainder of the night. In theory, split-night PSG may provide diagnostic effectiveness comparable with that of full-night PSG for patients who demonstrate frequent obstructive events in the early sleep period. However, concerns have been raised regarding both inadequate diagnostic sampling and time for CPAP autotitration.
- UHPSM is not attended by a technician and does not include EEG tracings, raising uncertainty about when tracings are actually recorded during sleep. But it is less accurate than full-night PSG and split-night PSG and is susceptible to data loss when probe leads shift in the unattended study.

A decision tree model was developed using a hypothetical cohort of persons aged 30 to 64 years of whom 85% were men. All had symptoms highly suggestive of OSAS, specifically, excessive daytime somnolence, persistent snoring, and witnessed apneas during sleep. An OSAS pretest probability of 82% was selected to be consistent with published studies for the model. Diagnostic criteria for OSA were used for all modalities from established values in published

studies. In this symptomatic group, all who met these OSA criteria were considered for CPAP therapy.
The cost analysis was performed from the perspective of third party payers, and only direct healthcare costs were considered. Health outcomes were expressed as (QALYs), the product of the utility and life expectancy for the health state. A diagnostic strategy was considered dominant over another if the total costs were lower and QALYs were the same or higher. Strategies that were more costly and more effective, in terms of QALYs, were assessed according to the incremental cost-effectiveness ratio (cost per QALY gained). Trade-offs of overall costs versus effectiveness were identified between the modalities. The home-studies strategy was less costly and less effective than split-night PSG and full-night PSG, as was split-night PSG compared with full-night PSG. The costs for additional QALYs incurred by full-night PSG and split-night PSG over home studies, and even by full-night PSG over split-night PSG, compare favorably with cost-utility estimates for a variety of widely accepted healthcare interventions.

The third study is a cost minimization study (Masa 2011) which compares home respiratory polygraphy (HRP) to lab PSG to diagnose sleep apnoea-hypopnea syndrome (OSAHS). This analysis provides cost estimates of these two modalities at the same level of diagnostic efficacy. Performed at eight centers in Spain, this blinded randomized crossover study in a highly symptomatic patient cohort, developed receiver operator curves (ROCs) calculated for various PSG AHI cut-off scores. To estimate the cost of HRP for diagnostic efficacy equal to PSG the most unfavorable polysomnographic cut-off point (AHI≥15) was chosen and an HRP AHI cut-off point to effectively exclude (<10) and to effectively confirm (≥25) OSAHS, with indeterminate scores (cases needing further PSG assessment) for 28% of subjects. The cost of achieving an HRP efficacy equal to that of PSG was calculated at half the cost of PSG, making HRP a significantly lower cost alternative for diagnosis in a population with an intermediate or high probability of moderate to severe OSAHS.

<u>Economic evaluations of various OSA treatment options</u>
One systematic review (SR) provided an evaluation of CPAP treatment of OSA using a cost-utility analysis (McDaid 2009). Another individual study evaluated oral appliances (OAs) used to treat OSA, also using a cost-utility analysis (Sadatsafavi 2009).

Note: Consideration of this key question in regards to treating OSA should be undertaken with the caveat that the evidence about the long term improvements in clinical outcomes of these treatments is indirect. Therefore, the cost-effectiveness of an as yet unproven treatment that lacks high quality evidence of effectiveness is somewhat speculative.

A cost–utility analysis uses the consequences of alternatives that are measured in 'health state preferences', which are given a weighting score. In this type of analysis, different consequences are valued in comparison with each other, and the outcomes (e.g., life-years gained) are adjusted by the weighting assigned. In this way, an attempt is made to value the quality of life associated with the outcome so that life-years gained become quality adjusted life-years gained (QALYs).

The SR (McDaid 2009) used four published and one CPAP manufacturer's (ResMed) full economic evaluations to inform a de novo economic model (the York Economic Model) to evaluate cost-effectiveness of CPAP for treatment of OSAHS. In this model, a cost–utility analysis was undertaken that compared CPAP with use of dental devices and conservative management over a lifetime time horizon. The costs of the use of these resources were reported related to 2005. The health effects of OSAHS, and the impacts of alternative treatments, were expressed in terms of QALYs. Due to the paucity of HRQoL data, estimates using other data were required. Three clinical endpoints were related to QALYs:

- Epworth Sleepiness Scale (ESS)
- Blood pressure
- Road traffic accidents (RTAs)

Additional model features included:

- Yearly cycles were chosen for the model.
- For expressing HRQoL utilities, the EQ-5D and SF-6D were used.
- Annual discount rate of 3.5%. Costs relating to the financial year 2005 were reported.
- Adults>16 yrs.
- Diagnosis of OSAHS by appropriate tool (e.g. AHI or arterial oxygen desaturation index and the ESS).
- The model was run separately by age and sex, given the availability of age- and sex-specific mortality data and CVE risks.
- The base-case analysis based on a male aged 50.

The base-case analysis compared the costs and QALYs of CPAP versus dental devices versus conservative management in a male aged 50 years. The model characterizes the patient's prognosis over his or her lifetime in terms of four health states: (1) OSAHS; (2) OSAHS post coronary heart disease (CHD); (3) OSAHS post stroke; and (4) death.

The findings of the York model in the hypothetical cohort of men aged 50 with specified CV risk factors:

- CPAP was associated with both higher costs and higher QALYs in comparison with treatment with dental devices or conservative management.
- The incremental cost-effectiveness of CPAP compared with dental devices is estimated to be £ 4000 per QALY. CPAP might therefore be considered cost-effective at a cost-effectiveness threshold per QALY of £ 20,000.
- The effect of CPAP on ESS has an ICER below a cost-effectiveness threshold of £ 20,000 for moderate and severe OSAHS. Note: It was not possible to estimate the differential effect of baseline severity of OSAHS on CVD and RTA risks (i.e., these cost-effectiveness results by severity include only treatment effects on ESS).

- The implications of the treatment effect of CPAP on blood pressure need to be estimated. The Framingham risk equations provide a link between risk factors such as blood pressure and the incidence of fatal and non-fatal CVEs. The relative risk reduction for CVE implied by the difference in SBP with CPAP compared with usual care is estimated to be relatively low using the Framingham risk equations (RR ≈ 0.98 for mean reduction in SBP of 1.06 mmHg).

Another individual study (Sadatsafavi 2009) provided a cost-utility analysis of the cost-effectiveness of oral appliances in the treatment of obstructive sleep apnoea–hypopnoea OSAH. A model was designed to simulate the costs and benefits of treatment of OSAH with OA or CPAP based on their effects on quality of life, motor vehicle crashes, and cardio-vascular effects.

This model used weighted age strata wherein the weights assigned to each age group reflect the demographic characteristics of patients with moderate to severe OSAH in a centre in the UK, which are comparable to that in the USA. A Markov model was created to simulate the natural course of moderate/severe OSAH and the impact of different strategies (no treatment, OA and CPAP) on disease outcomes over a 5-year period. Cardio-vascular and cerebro-vascular (CV) events were also modeled. Among many assumptions made in the model, an important feature was using the impact of CPAP and OA on the AHI was as a surrogate for their effectiveness on reducing other events due to OSAH. The assumption was made that a reduction in the risk of all events for CPAP and OA is proportional to their effect on reducing AHI.

The primary outcome of this study was the incremental cost-effectiveness ratio (ICER) expressed in terms of costs per one quality-adjusted life year (QALY) gained, 5 years after treatment. Assessing the relative efficacy of CPAP versus OA, the AHI was decreased by 18.3 (95% CI 14.1–22.0) and by 9.7 (95% CI 6.4–12.8) in these groups respectively. The relative efficacy of OA versus CPAP based on their ability to reduce AHI was estimated to be 0.53 (95% CI 0.38–0.77). CPAP reduced the ESS score by 3.91 (95% CI 2.29–8.10), versus OA which reduced ESS by 2.20 (95% CI 0.69–6.84).

In this model, OA and CPAP are both highly cost-effective treatments for OSAH when compared to no treatment, with CPAP being the best option. These results corroborate the current recommendations on the use of CPAP as the primary treatment for moderate/severe OSAH, with OA the preferred treatment in patients unwilling or unable to use CPAP.

In the USA, it is currently recommended that treatments that result in less than $50,000 costs per one additional QALY be adopted (Hirth 2000). Based on these standards, the ICER of $2,984 for OA and $13,698 for CPAP versus no treatment are highly favorable.

Summary

Over the past ten years, one systematic review and seven individual studies met inclusion criteria to address this Key Question. Four different methodologies provided these economic evaluations: (1) cost-utility analysis; (2) cost-analysis; and, (3) cost minimization analysis.

In assessing sleep study alternatives to laboratory PSGs, the costs for additional QALYs incurred by full-night PSG and split-night PSG over home studies, and even by full-night PSG over split-night PSG compared favorably with cost-utility estimates for a variety of widely accepted healthcare interventions. Furthermore, the home-studies strategy was less costly and less effective than split-night PSG and full-night PSG, as was split-night PSG compared with full-night PSG (Deutsch 2006). Another study noted that the cost of achieving an HRP efficacy equal to that of PSG was calculated at half the cost of PSG. This makes HRP a significantly lower cost alternative for diagnosis in a population with an intermediate or high probability of moderate to severe OSAHS (Masa 2011).

Social, economic, and healthcare utilization consequences of OSA were analyzed. In a case-control study using Danish governmental databases, patients with OSA were noted to have lower employment rates up to 8 years prior to diagnosis with further decline after diagnosis. Direct and indirect costs for patients compared to controls were €5257 vs €1396 ($p<0.0001$), corresponding to an annual mean excess cost per patient of €3860 (Jennum 2011). In another case-controlled study in Israel, elderly and middle-aged patients with OSA consumed approximately two times as much in the way of healthcare resources as their paired controls (Tarasiuk 2008).

Economic evaluations of various OSA treatment options, specifically comparing CPAP and OAs, were presented using economic models. In the York model, informed by the SR ((McDaid 2009), CPAP was associated with both higher costs and higher QALYs in comparison with treatment with dental devices or conservative management. The incremental cost-effectiveness of CPAP compared with dental devices is estimated to be £4000 per QALY. CPAP might therefore be considered cost-effective at a cost-effectiveness threshold per QALY of £ 20,000. In another model, OA and CPAP are both highly cost-effective treatments for OSAH when compared to no treatment, with CPAP being the best option. These results corroborate the current recommendations on the use of CPAP as the primary treatment for moderate/severe OSAH, with OA the preferred treatment in patients unwilling or unable to use CPAP. This study (Sadatsafavi 2009) the ICER of $2,984 for OA and $13,698 for CPAP versus no treatment are highly favorable.

Obstructive sleep apnea is a cause of significant morbidity and mortality, and is thus an important public health issue. In addition, the diagnosis and treatment of OSA have societal cost implications, making cost-effectiveness a concern in both of these aspects.

Limitations

- The studies identified for this Key Question were not comprehensive in covering all of the areas of interest to this report topic. Not all diagnostic tools, interventions, or outcomes were studied.

- These studies focused on moderate to high OSA severity of disease as little evidence exists which assesses mild cases. This is a severe limitation of both diagnosis and treatment options to be recommended for less symptomatic patients.
- Some studies did not report the discount rate used in cost analyses.
- The long term outcomes related to diagnosis and treatment of OSA have not been well studied and therefore are often not reported:
    - As a result, surrogate or intermediate outcomes are reported;
    - Frequently, these intermediate outcomes are extrapolated relative to the long term outcomes;
    - One example, of many, would be the use of the ESS as an intermediate outcome. A "connection" linking ESS to RTAs with all of their attendant down-stream consequences are made;
    - Another example would be blood pressure being "linked" by the Framingham risk equations to CVD or even OSA mortality (McDaid 2009).

Guidelines

Summary of Guidelines and Quality Assessment

The search for clinical practice guidelines identified 13 guidelines that were published within the past five years and pertained to obstructive sleep apnea: NICE (2007, 2008), American Academy of Sleep Medicine (AASM) (2006a, 2006b, 2007a, 2007b, 2008, 2009, 2010), American Society of Anesthesiologists (ASA 2006), University of Texas at Austin School of Nursing (USTN 2006), American Society of Plastic Surgeons (ASPS 2009) and European Federation of Neurological Societies (EFNS 2007). The included guidelines are summarized below and described in more detail in Appendix K. Appendix L describes each guideline's quality assessment rating. Appendix M includes the guideline quality assessment tool used for performing these guideline assessments.

Guidelines addressing screening for and diagnosis of obstructive sleep apnea (Table 1) agree that polysomnography is recommended for diagnostic purposes (AASM 2009; EFNS 2007; UTSN 2006). Portable monitoring for diagnosis of OSA is addressed by one fair quality guideline (AASM 2007a) that recommends for portable monitoring. Another fair quality guideline recommends against using auto-titrating CPAP ("APAP") for diagnostic purposes (AASM 2007b). Additionally, one poor quality guideline (UTSN 2006) recommends use of the Epworth Sleepiness Scale to aid in screening for OSA. This guideline was rated as poor quality primarily because the methods used to formulate the recommendations were not described.

Table 1. Guidelines addressing screening and diagnosis of obstructive sleep apnea

	UTSN 2006	AASM 2009	EFNS 2007	AASM 2007a	AASM 2007b
Screening	Epworth Sleepiness Scale: Yes				
Polysomnography	Yes	Yes	Yes		
Other diagnostic tools				Portable monitoring: Yes	APAP: No

Guidelines addressing treatment of OSA (Table 2) agree with their recommendation for CPAP as an OSA treatment (AASM 2006b, AASM 2008, EFNS 2007, NICE 2008). One high-quality guideline (NICE 2008) further specifies that CPAP is recommended for patients with moderate or severe OSA, and recommended for patients with mild OSA in cases where other treatments have failed or there are quality-of-life issues. Additionally, one high quality guideline (NICE 2007) recommends against soft palate implants for treatment of OSA, while one fair quality guideline (AASM 2010) recommends palatal implants in patients with mild OSA who cannot tolerate or are unwilling to adhere to CPAP therapy, or those in whom oral appliances have failed. One good quality guideline and one fair quality guideline (AASM 2009; AASM 2006a) both recommend bariatric surgery as an adjunct to weight loss for treatment of OSA in obese patients. One good quality guideline and two fair quality guidelines recommend tracheostomy only in certain cases, typically after other approaches have failed (AASM 2006b; AASM 2009; AASM 2010). A fair quality guideline (AASM 2010) addresses several surgical treatments for OSA, and recommends, in certain cases, multi-level stepwise surgery (MLS), and radiofrequency ablation (RFA) as potential options. The same guideline recommends against uvulopalatopharyngoplasty (UPPP) used alone, and against laser assisted uvuloplasty. The same guideline recommends against use of various pharmaceutical treatments, including selective serotonergic uptake inhibitors (SSRIs), protriptyline, methylxanthine derivatives, and estrogen therapy (estrogen preparations with or without progesterone), but does recommend Modafinil for treatment of residual excessive daytime sleepiness despite successful PAP treatment. Maxilo-mandibular advancement (MMA) is recommended by one good quality guideline (AASM 2009) and one fair quality guideline (AASM 2010).

Table 2. Guidelines addressing treatment of obstructive sleep apnea

	NICE 2007	NICE 2008	AASM 2010	AASM 2006a	AASM 2008	AASM 2006b	EFNS 2007	AASM 2009
CPAP		Moderate or severe OSA: Yes Mild OSA: Certain cases			Yes	Yes	Yes	
Other surgical procedures	Soft palate implants: No		Certain procedures in	Bariatric surgery: Certain				Certain procedures in certain

	NICE 2007	NICE 2008	AASM 2010	AASM 2006a	AASM 2008	AASM 2006b	EFNS 2007	AASM 2009
			certain cases (see guideline)	cases				cases (see guideline)
Other non-surgical treatment				Weight loss: Yes Modafinil: in certain cases Various other Rx: No (see guideline)				

Two guidelines addressed pre-operative issues relating to patients with obstructive sleep apnea (Table 3). One good quality guideline recommends criteria for screening patients for OSA, and recommends pre-operative initiation of PAP for patients with OSA (ASA 2006). Additionally, one fair quality guideline (ASPS 2009) recommends pre-operative use of CPAP for certain patients.

Table 3. Guidelines addressing pre-operative care of patients with obstructive sleep apnea

	ASA 2006	ASPS 2009
Pre-op screening	Yes	No recommendation described
Pre-op CPAP	Yes	Yes, in certain cases

Policy Considerations

As directed by the Washington HTA program, the policies for Medicare, Blue Cross Blue Shield, Aetna, GroupHealth and state insurance mandates were reviewed. Summaries of identified policies are provided below. Please see Appendices N and O for more detail.

Medicare National Coverage Determination

In 2009, a CMS National Coverage Determination (NCD) was issued to cover the following sleep tests in the specified settings for the purpose of diagnosing OSA. Other diagnostic tests were not determined to be sufficient support for CPAP coverage.

1. Type I PSG in a attended sleep lab facility, Type II or Type III sleep testing devices performed under supervision or at home;
2. Type IV sleep testing devices measuring three or more channels, including airflow, covered in or out of facility; and
3. Sleep testing devices measuring three or more channels including actigraphy, oximetry, and peripheral arterial tone covered in or out of facility.

In 2008, CMS issued a NCD for CPAP therapy for OSA in adults. The use of CPAP is considered reasonable and necessary when all of the following conditions are met:

1. 12 week 'trial' period to determine benefit. This period is covered if apnea-hypopnea index (API) or respiratory disturbance index (RDI) is greater than or equal to 15 events per hour or if between 5 and 14 events with additional symptoms including excessive daytime sleepiness or impaired cognition, or documented hypertension, ischemic heart disease, or history of stroke;
2. Providers must provide education with patient prior to use of CPAP machine to ensure proper use. Caregivers may be educated instead if they will be consistently operating the device; and
3. Positive diagnosis through polysomnogram (PSG) or Home Sleep Test (HST).

Local Coverage Determinations including Washington

L171 *LCD for Positive Airway Pressure (PAP) Devices for the Treatment of Obstructive Sleep Apnea* (Alaska, Oregon, Washington – Region X). Single-level continuous positive airway pressure devices are covered when a clinical evaluation is completed, an approved sleep test has resulted in a positive diagnosis, and the patient has received instruction from the supplier in proper use and care of the equipment. For individuals where CPAP machines are ineffective, a bi-level respiratory assist device can be used. Continued coverage requires that between the 31st and 91st day a clinical evaluation must be completed to determine therapeutic benefit. Replacement of the PAP device beyond the 5 year reasonable useful lifetime requires an re-evaluation but does not require a new sleep test. For patients who received a PAP device prior to enrollment in Medicare and are seeking coverage for a new device, documentation of a qualified sleep test and evaluation following Medicare enrollment are necessary.

L28606 *LCD for Oral Appliances for Obstructive Sleep Apnea* (Alaska, Oregon, Washington – Region X). Custom mandibular advancement oral appliances are covered following the same requirements for PAP devices and the patient is not able to tolerate PAP or the physician determines the use of PAP is contraindicated. Following the sleep test, the physician will order the device to be provided and billed for by a dentist. Prefabricated appliances are not covered due to insufficient evidence.

L30731 *LCD for Surgical Treatment of Obstructive Sleep Apnea* (LCD for 40 states – includes Washington). Three surgical options are covered for treatment of OSA when a diagnosis has been made, CPAP or other non-invasive treatments are not tolerated, and patients have been informed of the benefits and risks of surgery. The additional criteria are necessary for coverage of these procedures:

1. Uvulopalatopharyngoplasty (UPPP)
    - evidence of retropalatal and/or retrolingual obstruction as the cause of OSA;
2. Mandibular maxillary osteotomy
    - Evidence of retrolingual obstruction or previous failure of UPPP; and

3. Tracheostomy
    - when other treatments have failed or would not be effective Anatomic abnormalities of the upper airway when contributing to OSA.

Laser-assisted uvulopalatoplasty (LAUP), Somnoplasty, palatal implants, and submucosal ablation of the tongue base are not covered.

Washington Medicaid

Washington State's Health and Recovery Services Administration (HRSA) (2008) covers the rental and purchase of medically necessary CPAP equipment and related accessories when all of the following criteria are met:

1. The results of a prior polysomnogram performed in a HRSA-approved sleep center indicate the client has sleep apnea;
2. The client's physician determines the patient's sleep apnea is chronic;
3. CPAP is the least costly and most effective treatment modality; and
4. The item requested is not included in any other reimbursement methodology.

Washington HRSA covers the rental of CPAP equipment for a maximum of two months. Once client compliance and effectiveness is documented, HRSA will purchase the equipment.

Private Payers

Diagnosis for coverage of OSA by private payers requires either a recorded AHI greater than or equal to 15 events per hour or greater than or equal to five events per hour with documented excessive daytime sleepiness, impaired cognition, mood disorders or insomnia, hypertension, ischemic heart disease or history of stroke with exceptions noted below. A summary of coverage for diagnostic tests and treatment options can be found in Appendix O.

Aetna

Aetna (2011) considers the diagnosis and treatment of OSA in adults aged 18 and older medically necessary. In addition to the common criteria, a diagnosis requires at least 15 AHI events per hour with a minimum of 30 AHI events total. When the criteria are at least five events, there must be at least 10 total events and a diagnosis can be made if there are 20 or more episodes of oxygen saturation in a night. Diagnosis testing is covered when using a Type I, II, III, or specified Type IV device.

The following treatment options are covered when the given conditions are met:

1. CPAP
    - OSA diagnosis
    - Straps, masks, cushions and other accessories and supplies are also covered;
2. Alternative PAP machines

- Demonstrated intolerance to CPAP;

3. Custom-fitted oral appliances
    - OSA diagnosis; and
4. Surgical interventions
    - When other interventions fail the following procedures are approved:
        i. Uvulopalatopharyngoplasty (UPPP)
        ii. Uvulectomy and laser assisted uvuloplasty (LAUP);
        iii. Tracheostomy; and
        iv. Mandibular-maxillary advancement (MMA) surgery.

Treatments not covered due to insufficient evidence include somnoplasty and coblation; tongue base reduction surgery; partial glossectomy; the Repose system; cautery-assisted palatal stiffening operation (CAPSO); and, Pillar palatal implant system.

Continuing coverage may require repeated sleep studies up to twice per year to determine whether PAP treatment continues to be effective, whether pressure treatment settings need to be changed, or to determine whether continued treatment is necessary. Repeated sleep studies may also be necessary to assess treatment response after surgical procedures or treatment with oral appliances.

Blue Cross Blue Shield (BCBS)
Diagnostic criteria for BCBS (2010) for individuals with at least five AHI events per night also includes obesity, diabetes and glucose intolerance, unexplained choking or gasping during sleep, recurrent wakings during sleep, and unrefreshing sleep with daytime fatigue.

The only treatment policy was located under the surgery section and there were no matches for medical therapy coverage. The medical therapy treatment options mentioned are

1. CPAP;
2. Maximal medical treatment of any underlying disease;
3. Adjustment in sleep position; and
4. Avoidance of alcohol and sedative drugs.

When these medical therapies fail, including an adequate trial of CPAP, the following surgeries are covered

1. Hyoid suspension;
2. Mandibular-maxillary advancement (MMA); and
3. Uvulopalatopharyngoplasty (UPPP).

Procedures not covered due to insufficient evidence include uvulectomy or laser-assisted uvuloplasty (LAUP); Pillar palatal implant system; cautery-assisted palatal stiffening operation (CAPSO); the Repose system; partial glossectomy; and tongue base reduction surgery.

GroupHealth

For any coverage determination for OSA by GroupHealth (2010, 2011a, 2011b, 2011c, 2011d), criteria for diagnosis also includes a Sleep Apnea Clinical Score (SACS) of more than 15 and a completed baseline Standford Sleepiness Score and three night autotitration PAP and reported response. The following treatments are covered when given conditions are met:

1. CPAP
    - OSA diagnosis
2. Mandibular-maxillary advancement (MMA) surgery:
    - Type I test used for OSA diagnosis;
    - CPAP therapy has failed or is not tolerated;
    - Completed counseling from a physician of potential benefits and risks of surgery; and
    - Evidence of retrolingual and/or retropalatal obstruction as a cause of OSA or previous failure of UPPP.

Procedures not covered due to insufficient evidence include the Pillar palatal implant system; uvulectomy or laser-assisted uvulopalatoplasty (LAUP), somnoplasty, the Repose system, cautery-assisted palatal stiffening operation (CAPSO); and adaptive servo-ventilation therapy (ASV).

Overall Summary

General conclusions

Obstructive sleep apnea is a public health problem with a significant burden of morbidity and mortality. Accurately diagnosing and effectively treating OSA can improve symptoms of sleep disordered breathing and its consequences.

Diagnosing OSA, by detailing the obstructive episodes, has been done using: facility-based PSG; portable monitors in a laboratory or home environments; several questionnaire designs; and, with clinical prediction modeling:

- Compared to the current diagnostic standard, the PSG, the strength of evidence that Type II, III, and IV monitors can accurately diagnosis OSA is low to moderate with wide bias in estimating the actual AHI.
- There is a low strength of evidence that the Berlin questionnaire is moderately accurate to screen for OSA. Other questionnaires could not be evaluated due to insufficient

evidence with the exception of 1 study suggesting the STOP-Bang instrument may accurately screen for OSA.

- There is a low strength of evidence supporting the usefulness of clinical prediction modeling in OSA diagnosis.

There was insufficient evidence to determine the utility of phased testing (i.e., using a screening test result to determine the next test to be performed in a series) as compared to PSG.

The utility of preoperative screening for OSA could not be determined due to insufficient evidence.

Using AHI greater than 30 events per hour was found to be an independent predictor of all-cause mortality, with a high strength of evidence. A higher AHI was also associated with incident diabetes, based on a low strength of evidence. The association of diabetes and OSA may be confounded by obesity, which may result in both conditions. There was insufficient evidence to determine an association of AHI with other clinical outcomes (e.g., cardiovascular mortality and hypertension).

Moderate strength of evidence was found for the treatment of OSA with CPAP, though there was insufficient evidence to determine which patients CPAP might benefit the most.

The various different CPAP devices, delivery methods, and regimens had insufficient evidence regarding comparisons.

Mandibular advancement devices (MAD) had moderate strength of evidence supporting their use as an effective treatment for OSA. However, as with CPAP, there was insufficient evidence to indicate which patients might benefit from their use.

There was moderate strength evidence, that the use of CPAP is superior to MAD, with regard to improved sleep study measures, though no clinical outcomes were studied.

There was insufficient evidence to compare CPAP to the various oral devices, other than MAD.

Surgical interventions for the treatment of OSA had insufficient evidence with which to evaluate their relative efficacy. When each modality was compared to CPAP, both surgical interventions, and MAD had insufficient evidence to determine their relative merits. Of the other treatments for OSA that were considered, only intensive weight loss programs were an effective treatment, in obese patients with OSA, with a low strength of evidence. The remainder of the other management modalities (e.g., atrial overdrive pacing, medications, palatal implants, oropharyngeal exercises, tongue-retaining devices with positional alarms either in isolation or in combination, bariatric surgery, acupuncture, and auricular plaster) had insufficient evidence to determine the effects for treatment for OSA.

There was moderate strength of evidence that CPAP compliance was greater in those patients with more severe OSA. There was insufficient evidence regarding potential predictors of MAD compliance.

The strength of evidence was low for indentifying any specific intervention which may improve CPAP compliance. No intervention type (e.g., education, telemonitoring), was more promising than any of the others.

The diagnosis and treatment of OSA have societal cost implications, making cost-effectiveness a concern in both of these aspects. Full-night PSGs are more cost-effective than split-night PSG, followed by home sleep studies. However, when an intermediate or high probability of moderate to severe OSAHS exists, home sleep studies are an equally effective, lower cost alternative to lab PSGs. Social, economic, and healthcare utilization consequences of OSA are significantly higher than in those subjects without OSA. Elderly and middle-aged patients with OSA consumed approximately two times as much in the way of healthcare resources as their paired controls. Various OSA treatment options, specifically comparing CPAP and OAs, were presented using economic models. OA and CPAP are both highly cost-effective treatments for OSAH when compared to no treatment, with CPAP being the best option.

Guidelines addressed OSA diagnosis, pre-operative screening, and treatment. Three guidelines make recommendations for use of polysomnography in OSA diagnosis (AASM 2009; EFNS 2007; UTSN 2006). One guideline (AASM 2007a) recommends use of unattended portable monitors for OSA diagnosis, and one guideline recommends against auto titrated CPAP (APAP) for diagnostic use (AASM 2007b). Four guidelines recommend CPAP for treatment of OSA (AASM 2006b; AASM 2008; EFNS 2007; NICE 2008). A recommendation against soft palate implants was issued by one high-quality guideline (NICE 2007). Implants were recommended, in certain circumstances, by a fair-quality guideline (AASM 2010). Other surgical and non-surgical treatments were addressed by single guidelines. Recommendations for interventions such as multi-level stepwise surgery and radiofrequency ablation were recommended by single guidelines. Tracheostomy was recommended, in certain situations, by three guidelines (AASM 2006b; AASM 2009; AASM 2010). Maxilo-mandibular advancement was recommended by two guidelines (AASM 2009; AASM 2010). Bariatric surgery as an adjunct to weight loss for obese patients was recommended by two guidelines (AASM 2006a; AASM 2009). Modafinil was recommended by one guideline (AASM 2006a) for treatment of residual daytime sleepiness despite successful PAP treatment. This guideline recommended against various pharmaceuticals, such as serotonergic uptake inhibitors (SSRIs), protriptyline, methylxanthine derivatives and estrogen therapy for treatment of OSA. Pre-operative screening for OSA and CPAP initiation in certain cases were recommended by two guidelines (ASA 2006; ASPS 2009).

Federal, state and private payer policies are fairly consistent in their coverage of the diagnosis and treatment of OSA. Requirements for devices used to test for OSA are common across payers. To diagnose OSA, Medicare and Aetna require a Type I PSG in a facility, or a Type II, III, or IV sleep test (with three or more channels) in a facility or at home. In contrast, Washington Medicaid requires a PSG at a HRSA-approved sleep center for diagnosis. Criteria for diagnosis

were similar across payers with small variations in the noticeable symptoms that would lead to a positive diagnosis. Most payers cover CPAP as the first line treatment option, often followed by the use of a custom-fitted oral device, and varying forms of surgical treatment. Medicare National Coverage Decisions provide coverage for CPAP and specify which sleep tests are covered for diagnostic purposes. A number of Local Coverage Decisions that include Washington indicate coverage for CPAP, custom-fitted oral devices. When those treatments fail and additional indications are present, three surgical options are covered. Lifestyle counseling is included inconsistently as an initial part of treatment.

Limitations of the evidence

- Almost all studies were conducted in academic or research centers, raising concerns as to the generalizability and applicability of the findings to the general population
- There was a lack of trials which evaluated long-term clinical outcomes (i.e., as compared to sleep study parameter outcomes)
    - For example, using a high AHI as a predictor of severe OSA, would imply that these patients should be aggressively treated to reduce their AHI
    - However, there are almost no trial data to support that treatment of OSA, and the reduction of AHI, improves clinical outcomes
    - *Therefore, the clinical value of these associations remains theoretical*
- Several Key Questions related to assessing how results may vary in different subgroups of patients
    - These questions could not be addressed because very few trials reported subgroup analyses
- Study follow-up durations tended to be short, and dropout rates were often high
- Many studies had incomplete reporting and inadequate analyses, which required the authors of this systematic review to estimate pertinent results
- Publication bias was a concern, as many trials of devices were funded by industry
- The economic studies reported were not comprehensive in that, not all diagnostic tools, interventions, or outcomes were studied.
- Most economic studies focused on moderate to high OSA severity of disease as little evidence exists which assesses mild cases. This is a severe limitation of both diagnosis and treatment options to be recommended for less symptomatic patients.