

Evidence-based Practice Center Systematic Review Protocol

Project Title: *Comparative Effectiveness of Diagnosis and Treatment of Obstructive Sleep Apnea in Adults*

I. Background and Objectives for the Systematic Review

Sleep apnea is a common disorder that affects all ages. The American College of Chest Physicians (ACCP; 2006) estimates the prevalence of obstructive sleep apnea (OSA) in the United States to be between 5-10 percent and asserts that as many as one in four American adults could benefit from evaluation for OSA.¹ The condition is characterized by periods of disturbed airflow patterns during sleep time, namely reduced airflow (hypopnea) or airflow cessation (apnea). It is postulated that both types of airflow disturbance have similar pathophysiology and bear the same clinical significance.² OSA is by far the most common type of the condition; apneas and hypopneas of central and mixed central and obstructive etiology comprise the other forms.² OSA has been associated with a variety of adverse clinical outcomes, such as mortality secondary to cardiovascular disease,³⁻⁵ decreased quality of life,⁶ cardiac disease and stroke,^{3,7} hypertension,⁸⁻¹⁰ and noninsulin-dependent diabetes and other metabolic abnormalities.^{5,11-14} It also is associated with an increased likelihood for motor vehicle and other accidents.^{15,16}

Diagnosis

The severity of sleep apnea is typically quantified by the number of apneas and hypopneas per hour of sleep, a quantity that has been termed the apnea-hypopnea index (AHI). The symptom of excessive daytime sleepiness is quite variable and is not always present in patients with OSA; thus, in most patients, the condition remains undiagnosed and untreated.⁶

There is a large amount of clinical uncertainty surrounding this condition, including inconsistencies in the definition of the disease. While in-laboratory polysomnography is considered the gold standard in clinical practice to diagnose obstructive sleep apnea, it is not without constraints such as cost, interlaboratory variation in hardware and assessment methods. The standard measurement of AHI (and by extension, the diagnosis of sleep apnea) requires a comprehensive, technologist-attended sleep study with multichannel polysomnography, which is performed in specialized sleep laboratories.^{2,17} Laboratory-based polysomnography records a variety of neurophysiologic and cardiorespiratory signals and is interpreted by trained technologists and sleep physicians after the sleep study has been completed.

However, it is acknowledged that it is not a definitive test to either diagnose or rule out obstructive sleep apnea. In part, this is due to a lack of robust standardized criteria as to the test parameters measured and the thresholds of the parameters used to make the diagnosis.

Since in-laboratory polysomnography is costly, resource intensive, and burdensome to the patient, other diagnostic tools have been developed, including portable tests¹⁷ and questionnaires for pre-screening patients. There are different types of portable monitors, which gather different neurophysiologic and respiratory information and may synthesize the accumulated data differently.¹⁸ Different screening questionnaires exist to pre-screen patients for further testing or treatment. The value of the different tests and of the questionnaires and other screening tools remains unclear. There is also lack of clarity as to whether the tests can be

accurately used to predict the clinical severity of patients' sleep apnea and their likelihood of clinically important sequelae.

Preoperative testing

People with sleep apnea are at increased risk of surgical and anesthesia-related adverse outcomes.¹⁹ Finding patients undergoing surgery with undiagnosed sleep apnea could, in theory, allow optimization of peri-operative care to minimize problems with intubation, extubation, and other respiratory events. At present, though, the need to screen all or selected surgical patients and what method of screening is effective and efficient is unclear.

Treatment

Continuous positive airway pressure (CPAP) is the standard 1st-line therapy for most patients diagnosed with obstructive sleep apnea. Obstructive sleep apnea occurs when the upper airway closes or becomes overly narrow as the muscles in the oropharynx (mouth and throat) relax during sleep. This results in inadequate or stopped breathing, which reduces oxygen in the blood and causes arousal from sleep. The CPAP machine counteracts this sequence of events by delivering compressed air to the oropharynx, splinting the airway (keeping it open with increased air pressure) so that unobstructed breathing becomes possible, reducing and/or preventing apneas and hypopneas.

For many patients, using CPAP results in immediate improvement in sleep and improvement in quality of life largely related to decreased daytime somnolence. However, it has been suggested that approximately one-quarter to one-half of patients with obstructive sleep apnea will either refuse the offer of CPAP therapy, will not tolerate it, fail to use the machine properly, or for other reasons do not comply with CPAP use.²⁰ These patients are essentially untreated and receive little or no benefit from the device.

When CPAP is refused or not tolerated, a number of 2nd-line treatments are available including, uvulopalatopharyngoplasty (UPPP), radiofrequency ablation, jaw surgery, and bariatric surgery, for eligible candidates. UPPP, radiofrequency ablation, and jaw surgery are surgical techniques to remove or shrink and scar redundant tissue that is causing the obstruction or to otherwise minimize the obstruction. The goal of bariatric surgery is to reduce body weight and fat, which may shrink the oropharyngeal tissue causing the obstruction. However, life-threatening complications have been associated with sleep apnea surgery. Fatalities have been related to upper airway collapse or obstruction secondary to pharmacological sedation and surgical edema.²¹

Other less invasive techniques include oral appliances, which are worn overnight and aim to mechanically splint the oropharynx open; positional therapy, devices to prevent lying supine during sleep, a position that for many patients exacerbates the obstruction; pharyngeal or laryngeal exercises to improve muscle tone; non-surgical weight loss programs; and physical-exercise programs.

Another management approach is to provide interventions that will increase compliance with CPAP use. These include structured education about the value of CPAP and how to use and adjust the CPAP; structured individual follow-up to correct any problems; group support; and relieving nasal congestion or dryness caused by the CPAP machine.

Food and Drug Administration (FDA) Status, Indications, and Warnings

The systematic review will cover the following devices and diagnostic tools: polysomnography, CPAP devices, autotitrating positive airway pressure devices, bilevel positive airway pressure devices, and dental and intraoral devices. We do not plan to review any drugs.

The number of specific devices that have been approved by the FDA for sleep apnea are too numerous to describe in detail here. Briefly, we found 126 CPAP machines made by about 64 companies that have been approved by the FDA between 1976 and November 2009. We found 26 bilevel positive airway pressure machines, all manufactured by Respironics Inc., that were approved by the FDA between 1987 and October 2009. We found 12 additional treatment devices, including tubes, head positioning devices, nasal CPAP, oral appliances, and other devices. These devices are made by 9 different companies and have been approved by the FDA between 1986 and April 2009.

II. The Key Questions

Diagnosis

KQ1: How do different available tests compare to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep?

- a. 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?

KQ2: 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?

KQ3: How does phased testing (screening tests or battery followed by full test) compare to full testing alone?

KQ4: What is the effect of pre-operative screening for sleep apnea on surgical outcomes?

Treatment

KQ5: What is the comparative effect of different treatments for obstructive sleep apnea (OSA) in adults?

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

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

KQ6: In OSA patients prescribed non-surgical treatments, what are the associations of pre-treatment patient-level characteristics with treatment compliance?

KQ7: What is the effect of interventions to improve compliance with device (CPAP, oral appliances, positional therapy) use on clinical and intermediate outcomes?

Public Comments

The large majority of public comments regarding the key questions were either answers given to the key questions by individual commentators or anecdotal evidence related to the key questions. Comments related to potential alterations to the key questions or the eligibility criteria are displayed in Table 1 below.

Table 1. Comments on Draft Key Questions and Evidence-based Practice Center (EPC) Response

Comment	EPC Response
Facility-based polysomnography (PSG) is not a gold standard for diagnosis of sleep apnea	The report will not assume any test is a gold standard. Sensitivity and specificity will be analyzed only for clinical outcomes, not for diagnosis of sleep apnea.
Apnea-hypopnea measurements from portable monitors and facility-based PSG are not interchangeable (especially in the higher end of the AHI spectrum).	The report will not assume that measurements from any test are interchangeable with measurements from other tests. The source of the measurements will be captured and evaluated.
Cost [and cost-effectiveness] is at least one consideration of many in any thorough review of “comparative effectiveness.”	While this may be true for some definitions of comparative effectiveness, it is not a necessary aspect of comparative effectiveness review (CERs). The EPC will not be evaluating cost.
The criteria used to categorize sleep apnea as mild, moderate, and severe are inexact and have varied over time, resulting in uncertainty about what constitutes “obstructive sleep apnea” and how to define severity. The multiplicity of definitions of hypopnea and changes over time make such comparisons impossible.	The EPC will capture definitions of severity of sleep apnea and will review the evidence in light of these comments. The EPC will not specifically review how the definitions have changed over time.
Oximetry can be diagnostic of OSA.	Pulse oximetry will be evaluated among the diagnostic tests.
The symptom of migraines should be considered as a possible predictor of sleep apnea.	Nonstandard symptoms will not explicitly be evaluated.
What is the correlation between the choice of initial diagnostic test to determine the diagnosis of sleep apnea and the ultimate treatment outcomes (including compliance with treatment)?	Key question 3 evaluates phased testing. We will not otherwise review the order of testing.
Night-to-night variation in an important factor to consider when comparing portable monitors to PSG.	We will capture and analyze information on this where it is reported.
Auto-titrating CPAP needs to be considered.	It will be included.
Implied but not directly asked is the value of screeners.	Key question 3 will evaluate different screening tests and protocols.

Both efficacy and effectiveness need to be considered.	They will be.
--	---------------

Eligibility Criteria

Population(s)

KQ 1 & KQ3

- Adults (>16 yo) with symptoms, findings, history, comorbidities that clinically indicate that they are at increased risk of having sleep apnea
- Exclude:
 - Neuromuscular disease, Down syndrome, Prader-Willi syndrome, major congenital skeletal abnormalities, narcolepsy, narcotic addiction
 - Children (≤ 16 yo)
 - (Studies of) only asymptomatic / healthy / general population / control people
 - (Studies of) only patients who carry a known diagnosis of / with known sleep apnea

KQ 2

- All adults (>16 yo)
- (Ideally) CPAP not being used at baseline / time of screening

KQ 4

- Pre-operative patients, all surgeries, to receive general anesthesia

KQ 5-7

- Diagnosed with obstructive sleep apnea (OSA)
 - Formal testing performed
 - Apnea-hypopnea index ≥ 5
 - With or without symptoms
 - Allow individual study investigator definitions of OSA within these parameters
- Adults (>16 yo)
- Exclude:
 - Central sleep apnea
 - Children (≤ 16 yo)
 - Neuromuscular disease, Down syndrome, Prader-Willi syndrome, major congenital skeletal abnormalities, narcolepsy, narcotic addiction

Interventions

KQ 1, 3, & 4

- Polysomnography (Facility-based, with monitoring)
- Home monitoring devices
 - For polysomnography and home monitoring devices
 - Any combination of 2 or more “channels” (measured factors)
 - Pulse transit time (alone)
 - Peripheral arterial tone (alone)
 - Pulse oximetry (alone)
 - Exclude:
 - Heart rate (alone)
 - Heart rate variability (alone)
 - Actigraphy (alone)
 - Other single channel tests
- Standardized screening and other questionnaires, scales, etc. that include clinical criteria (e.g., signs, symptoms, history, comorbidities)
 - Clinical decision-making tools
 - Exclude:
 - Single or multiple patient characteristic predictors or risk factors (e.g., BMI alone) that are not part of a standardized diagnostic tool
- Include information on the FDA status, indications, and warnings for use of devices covered in the systematic review (in background material / introduction of report)

KQ 2

- Formal testing at baseline
- AHI or ODI (oxygen desaturation index) plus
- Clinical symptoms and history
 - (Focus on Medicare coverage factors)

KQ 5-7

- Positive airway pressure masks (Continuous positive airway pressure [CPAP], Bilevel positive airway pressure machines [BiPAP], Auto-titrating continuous positive airway pressure [APAP], and similar devices)
- Surgery (Not KQ 6, 7)
 - Jaw or mouth
 - Airway (including nasal)
 - Bariatric (for weight loss)
- Oral appliances / Dental devices (mandibular advancement devices, tongue retaining devices, and similar devices)
- Positional therapy (devices to alter sleep positions)
- Physical therapy (Not KQ 6, 7)
 - Pharyngeal or laryngeal exercises, or similar approaches
 - Only formal therapy protocols, not just advice
- Lifestyle modification (Not KQ 6, 7)
 - Non-surgical weight loss programs

- Only formal protocols, not just single-time advice
- Medication (for weight loss)
- Exercise
 - Only formal protocols, not just advice
- Include information on the FDA status, indications, and warnings for use of devices covered in the systematic review (in background material / introduction of report)

Outcomes

KQ 1, 3, & 4

1. Correlation, concordance, or agreement among tests (Not KQ 4)
2. Predictive value (sensitivity, specificity) for clinical outcomes listed under treatment KQs.
3. Change in clinical management
4. Clinical outcomes (listed under treatment KQs), if reported in a randomized controlled trial of diagnostic tests
5. KQ 4:
 - a. Intra-operative events
 - b. Surgical recovery events
 - c. Surgical recovery time
 - d. Post-surgical events
 - e. Length of intensive care or hospital stay
 - f. Intubation failures
 - g. Extubation failures
6. Harms, Adverse events
 - a. Any adverse events

KQ 2, 5

7. Sleep / wakefulness clinical outcomes
 - a. Quality of life
 - i. Disease specific (e.g., Functional Outcomes of Sleep Questionnaire [FOSQ], Calgary questionnaire)
 - ii. General (e.g., Short Form survey instrument-36 [SF-36], EuroQoL EQ-5D, others but with priority given to preference-based utility scores)
 - b. Sleepiness / somnolence measures
 - i. Subjective (e.g., Epworth sleepiness scales); only validated measures
 - ii. Objective (e.g., Multiple Sleep Latency Test, Maintenance of Wakefulness Test); only validated measures
 - c. General symptom scales (e.g., diagnostic screening questionnaires); only validated measures
 - d. Psychological function scales; only validated measures
 - e. Cognitive or Executive function scales; only validated measures
 - f. Physical function scales; only validated measures
 - g. Accidents ascribed to somnolence (e.g., motor vehicle, home accidents)



- h. Sleep quality; only formal scales or questionnaires
 - i. Work days lost (and equivalent)
8. Comorbidities clinical outcomes
 - a. Mortality
 - b. Cardiovascular events (Congestive heart failure symptoms, stroke, arrhythmia events, coronary artery disease events, hypertension (diagnosis, resolution, or reduction in medications))
 - c. Non-insulin dependent diabetes (diagnosis, resolution, start or end treatment)
 - d. Depression event (diagnosis, recurrence, etc.)
9. Intermediate or Surrogate outcomes
 - a. Sleep study measures
 - i. Apnea hypopnea index
 - ii. Number of arousals
 - iii. Time in deeper sleep stages
 - b. Hemoglobin A1c
 - c. Blood pressure
10. Adherence / Compliance
 - a. Categorical (adhering or not adhering)
 - b. Time (hours) using device per time period
11. Harms, Adverse events
 - a. Any adverse events

KQ 6-7

- Adherence / Compliance
 - a. Categorical (adhering or not adhering)
 - b. Time (hours) using device per time period

Timing & Study Design

KQ 1, 3, & 4

- Cross-sectional or longitudinal of any duration
- Prospective
- Determine minimum N (sample size) after list of potentially eligible studies are available

KQ 2

- Prospective
- $N \geq 1000$ (may be flexible about this)
- Multivariate analysis (probably)

KQ 5-7

- Minimum duration:
 - Sleep clinical outcomes: 4 weeks

- Accidents, work days lost: 1 year
 - Comorbidities clinical outcomes: 1 year
 - Intermediate outcomes
 - Sleep study measures: No minimum duration except that not within the original sleep study; must be from a subsequent evaluation
 - Hemoglobin A1c and blood pressure: 4 weeks
 - Compliance: 4 weeks (KQ 6 & 7: no minimum duration)
 - Adverse events: No minimum duration
 - Post-surgery: minimum 6 weeks (except adverse events)
- Prospective
 - Except surgical studies, where retrospective studies (may be) allowed
- Comparative (≥ 2 treatments or treatment vs. placebo/control)
 - Except surgical studies, where cohorts are allowed
 - Plan to treat surgical cohorts within non-randomized comparative studies as separate “studies”/cohorts
 - (Not KQ 6)
- Determine minimum N (sample size) after list of potentially eligible studies are available
 - Will use separate minimum N’s for different topics, including different categories of surgery¹ (KQ 6: $N \geq 100$, tentatively)
- KQ 6: Multivariate analysis (possibly)

Setting

KQ 1-3

- Any setting (primary or specialty care, in-facility or home; inpatient or outpatient)

KQ 4

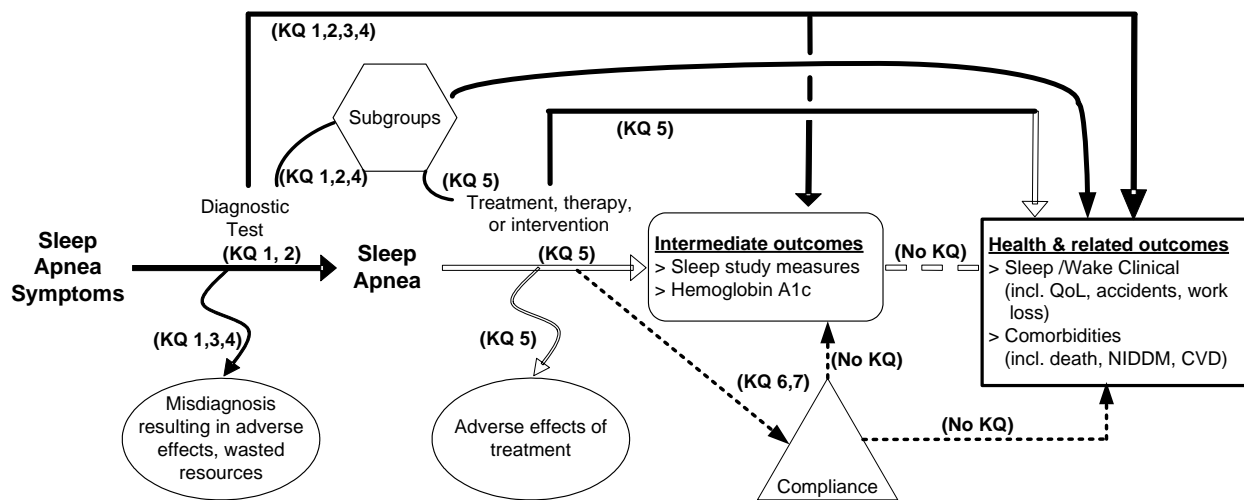
- Any pre-operative setting (hospital or ambulatory, any surgery)

KQ 5-7

- Any setting

¹The surgical categories currently include: UPPP, UPPP+tongue surgery, tongue surgery, genioglossohyoid surgery, maxillomandibular surgery, mandibular surgery, rapid maxillary expansion/splint.

III. Analytic Framework



Legend: CVD, cardiovascular disease; KQ, key question; NIDDM, non-insulin dependent diabetes mellitus; QoL, quality of life.

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

We plan to use the eligibility criteria for populations, interventions, comparators, outcomes, timing, and settings as enumerated above. We will discuss with the Technical Expert Panel further criteria regarding study design and publication date range. We do not plan to add data from the grey literature. We do not expect to contact authors for additional data.

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies To Answer the Key Questions.

Appendix 1 at the end of this document has our proposed literature search strategy. This search will be conducted in Medline and the Cochrane Central Register of Controlled Trials. Hand searches will not be done. The Tufts EPC is developing a computerized screening program. We will use this program to assist with screening. For program testing and training purposes, we will manually screen 1000 abstracts, each twice. Among the abstracts that are rejected by the program, we will review a sufficient number until we are confident that it has accurately rejected them. Remaining abstracts will be manually screened based on the eligibility criteria. Full-text copies will be retrieved for all potentially relevant articles. These will be rescreened for eligibility. The reasons for excluding these articles will be tabulated. We will ask the technical experts and others to inform us of any potentially missing articles. All suggested articles will be screened for eligibility using the same criteria as for the original articles. If necessary, we will revise the literature search to find articles similar to those missed. When the draft report has been

submitted, we will run an updated literature search (using the same search strategy) and will add any additional articles we find to the final report.

C. Data Abstraction and Data Management

Each study will be extracted by one experienced methodologist. The extraction will be reviewed and confirmed by at least one other methodologist. Extracted data will be recorded on standard forms in Microsoft Word. The basic elements and design of the forms will be the same as multiple forms we have used for other comparative effectiveness reviews, technology assessments, evidence reports, and other systematic reviews. Prior to use, the form will be customized to capture all the relevant elements for the key questions. We will use separate forms for questions related to diagnosis (KQ1–3) and treatment (KQ4–6). We will test the forms on several select studies and revise the forms as necessary before data extraction is fully performed for all articles.

We will extract basic demographic data such as age, sex, and race and any and all factors that may have a role in effect modification of the intervention-outcome association. These will, at a minimum, include baseline weight, AHI, symptoms, sleepiness measures, bed partner, airway, and other physical characteristics.

D. Assessment of Methodological Quality of Individual Studies

We will use methodology for evaluating study quality that is standard within the Tufts EPC and is recommended in chapter 6 of the [Methods Reference Guide](#) for Effectiveness and Comparative Effectiveness Reviews. Briefly, we will rate each study as being of good, fair, or poor quality based on their adherence to well-accepted standard methodologies and adequate reporting. The grading will be outcome-specific such that a given study that reports its primary outcome well but did an incomplete analysis of a secondary outcome would be graded of different quality for the two outcomes. Studies of different study designs will be graded within the context of their study design. Therefore, randomized controlled trials will be graded good, fair, or poor, and observational studies will separately be graded good, fair, or poor. However, we expect to limit any included retrospective studies to fair or poor.

E. Data Synthesis

All included studies will be summarized in narrative form and in summary tables that tabulate the important features of the study populations, design, intervention, outcomes, and results. For questions regarding comparisons of diagnostic tests (KQ1–2), we will consider using Bland-Altman plots, which graph the differences in measurements against their average. This approach is recommended for analyses in which neither test can be considered a reference (gold) standard, as will be the case with the sleep apnea diagnostic studies. Analyses of sensitivity and specificity can be inappropriate. For KQ3, KQ4, and KQ6—which evaluate the effect of an intervention on intermediate and clinical outcomes—we will consider performing meta-analyses when at least 3 unique studies are deemed to be sufficiently similar in population and to have the same comparison of interventions and the same outcomes. We expect to require input from domain experts to assess whether studies are too clinically heterogeneous for a meta-analysis to be appropriate. We will perform only random effects model meta-analyses. For KQ5, we will

search for studies that directly analyze the question of whether any pretreatment patient characteristics are associated with treatment failure (or success); these will be described and discussed in narrative form. We will not attempt any meta-regression of these studies. If data are available, we will consider sub-group meta-analyses based on the findings of these studies.

F. Grading the Evidence for Each Key Question

We will follow Chapter 11 in the [Methods Reference Guide](#) for Effectiveness and Comparative Effectiveness Reviews to grade the strength of the bodies of evidence for each key question.

We will define the risk of bias (low, medium, or high) based on the study design and the methodological quality of those studies.

We will determine the consistency of the data as having or not having inconsistency (or not applicable if only one study). We do not plan to use rigid counts of studies (e.g., 4 of 5 agree, therefore there is consistency), but instead we will evaluate the direction, magnitude, and statistical significance of all studies and make a determination. We will describe our logic where studies are not unanimous.

We will assess the directness (direct or indirect) of the evidence. Indirect evidence will mean that either the populations are not applicable to the general population of adults with obstructive sleep apnea or that the comparison of interest was not made in individual trials (e.g., that A vs. B can only be assessed by evaluating studies of A vs. placebo and B vs. placebo). Since we will be assessing primarily clinical outcomes, we do not expect to consider whether an outcome is intermediate or surrogate in our determination of directness.

We will assess the precision (precise or imprecise) of the evidence based on the degree of certainty surrounding an effect estimate. A precise estimate is an estimate that would allow a clinically useful conclusion. An imprecise estimate is one for which the confidence interval is wide enough to include clinically distinct conclusions (e.g., both clinically important superiority and inferiority (i.e., the direction of effect is unknown), a circumstance that will preclude a conclusion.

We will follow the [Methods Reference Guide](#) for Effectiveness and Comparative Effectiveness Reviews and use four strength-of-evidence levels: high, moderate, low, and insufficient. We will assign these levels of evidence based on our level of confidence that the evidence reflects the true effect for the major comparisons of interest.

V. References

1. Hiestand DM, Britz P, Goldman M, et al. Prevalence of symptoms and risk of sleep apnea in the US population: Results from the national sleep foundation sleep in America 2005 poll. *Chest* 2006;130:780-6.
2. Polysomnography Task Force, American Sleep Disorders Association Standards of Practice Committee. Practice parameters for the indications for polysomnography and related procedures. *Sleep* 1997;20:406-22.
3. Marin JM, Carrizo SJ, Vicente E, et al. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet* 2005;365:1046-53.
4. Yaggi HK, Concato J, Kernan WN, et al. Obstructive sleep apnea as a risk factor for stroke and death. *N Engl J Med* 2005;353:2034-41.
5. Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thoracic Soc* 2008;5:136-43.
6. Young T, Palta M, Dempsey J, et al. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993;328:1230-5.
7. Shahar E, Whitney CW, Redline S, et al. Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study. *Am J Respir Crit Care Med* 2001;163:19-25.
8. Nieto FJ, Young TB, Lind BK, et al, for the Sleep Heart Health Study. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. *JAMA* 2000;283:1829-36.
9. Peppard PE, Young T, Palta M, et al. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000;342:1378-84.
10. Young T, Peppard P, Palta M, et al. Population-based study of sleep-disordered breathing as a risk factor for hypertension. *Arch Intern Med* 1997;157:1746-52.
11. Babu AR, Herdegen J, Fogelfeld L, et al. Type 2 diabetes, glycemic control, and continuous positive airway pressure in obstructive sleep apnea. *Arch Intern Med* 2005;165:447-52.
12. Gruber A, Horwood F, Sithole J, et al. Obstructive sleep apnoea is independently associated with the metabolic syndrome but not insulin resistance state. *Cardiovasc Diabetol* 2006;5:22.
13. Caples SM, Gami AS, Somers VK. Obstructive sleep apnea. *Ann Intern Med* 2005;142:187-97.
14. Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med* 2002;165:1217-39.
15. Howard ME, Desai AV, Grunstein RR, et al. Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. *Am J Respir Crit Care Med* 2004;170:1014-21.
16. Terán-Santos J, Jimenez-Gomez A, Cordero-Guevara J, for the Cooperative Group Burgos-Santander. The association between sleep apnea and the risk of traffic accidents. *N Engl J Med* 1999;340:847-51.

17. Flemons WW, Littner MR, Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature: an evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest* 2003;124:1543-79.
18. Ferber R, Millman R, Coppola M, et al. Portable recording in the assessment of obstructive sleep apnea: ASDA standards of practice. *Sleep* 1994;17:378-92.
19. Benumof JL. Obstructive sleep apnea in the adult obese patient: implications for airway management. *Anesthesiol Clin North America* 2002;20:789-811.
20. Veasey SC, Guilleminault C, Strohl KP, et al. Medical therapy for obstructive sleep apnea: a review by the Medical Therapy for Obstructive Sleep Apnea Task Force of the Standards of Practice Committee of the American Academy of Sleep Medicine. *Sleep* 2006;29:1036-44.
21. Won CH, Li KK, Guilleminault C. Surgical treatment of obstructive sleep apnea: upper airway and maxillomandibular surgery. *Proc Am Thoracic Soc* 2008;5:193-199.
22. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. 1st ed. Rockville, MD: Agency for Healthcare Research and Quality. Draft posted October 2007. Available at: http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf.

VI. Definition of Terms

All terms requiring definition have been addressed in the background and objectives.

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

NOTE: The following protocol elements are standard procedures for all protocols.

VIII. Review of Key Questions

For Comparative Effectiveness reviews (CERs) the key questions were posted for public comment and finalized after review of the comments. For other systematic reviews, key questions submitted by partners are reviewed and refined as needed by the EPC and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed.

IX. Technical Expert Panel (TEP)

A TEP panel is selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological

approaches do not necessarily represent the views of individual technical and content experts. The TEP provides information to the EPC to identify literature search strategies, review the draft report and recommend approaches to specific issues as requested by the EPC. The TEP does not do analysis of any kind nor contribute to the writing of the report.

X. Peer Review

Approximately five experts in the field will be asked to peer review the draft report and provide comments. The peer reviewer may represent stakeholder groups such as professional or advocacy organizations with knowledge of the topic. On some specific reports such as reports requested by the Office of Medical Applications of Research, National Institutes of Health there may be other rules that apply regarding participation in the peer review process. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

It is our policy not to release the names of the Peer reviewers or TEP panel members until the report is published so that they can maintain their objectivity during the review process.

Appendix 1. Proposed Search Strategy

#	Searches	Results
	Sleep Apnea	
1	exp Sleep Apnea Syndromes/ or exp Sleep Apnea, Obstructive/	19267
2	exp Airway Resistance/	13219
3	exp snoring/	2922
4	Upper airway resistance syndrome.mp.	229
5	Respiratory disturbance.mp.	1151
6	obstructive sleep apn?ea.mp.	11746
7	or/1-6	35183
	Comparative Studies	
8	randomized controlled trial.pt.	561439
9	controlled clinical trial.pt.	161324
10	randomized controlled trials/	67116
11	Random Allocation/	88433
12	Double-blind Method/	194574
13	Single-Blind Method/	22535
14	clinical trial.pt.	748992
15	Clinical Trials.mp. or exp Clinical Trials/	282752
16	(clinic\$ adj25 trial\$.tw.	234940
17	((singl\$ or doubl\$ or trebl\$ or tripl\$) adj (mask\$ or blind\$)).tw.	223366
18	Placebos/	48648
19	placebo\$.tw.	236373
20	random\$.tw.	781376
21	trial\$.tw.	607720
22	(randomized control trial or clinical control trial).sd.	227662
23	(latin adj square).tw.	3635
24	Comparative Study.tw. or Comparative Study.pt.	1644250
25	exp Evaluation studies/	149831
26	Follow-Up Studies/	437652
27	Prospective Studies/	331595
28	(control\$ or prospectiv\$ or volunteer\$).tw.	2591469
29	Cross-Over Studies/	45716
30	Or/8-29	5078479
	Treatments	
31	exp Positive-Pressure Respiration/ or exp Continuous Positive Airway Pressure/	18518
32	exp Intermittent Positive-Pressure Ventilation/ or exp Ventilators, Mechanical/ or exp Masks/	16688
33	general surgery/ or neurosurgery/ or otolaryngology/ or surgery, plastic/ or thoracic surgery/	80977
34	Surgical Procedures, Operative/	46806
35	oral appliances.mp.	286
36	exp Physical Therapy Modalities/ or exp Exercise Therapy/	107007
37	positional therapy.mp.	41
38	exp Weight Loss/	22268
39	exp Exercise/ or exp Exercise Therapy/	78677

Source: www.effectivehealthcare.ahrq.gov

Published Online: March 15, 2010



40	exp Therapeutics/	2836406
41	exp Anesthesia/ or Pre-operative screening/ or Anesthetic agents/	164281
42	Sleep Apnea, Obstructive/th	2408
43	*tonsillectomy/	4707
44	or/31-43	3119682
	Sleep Apnea Diagnostic Terms	
45	exp Polysomnography/	11867
46	exp Oximetry/	10559
47	exp Monitoring, Physiologic/	110172
48	pulse transit time.mp.	246
49	exp Monitoring, Ambulatory/	18720
50	peripheral Arterial Tonometry.mp.	62
51	exp Questionnaires/	224876
52	exp Diagnostic Tests, Routine/	5348
53	exp "Laboratory Techniques and Procedures"/	1586381
54	(Epworth or Stanford or Berlin or Pittsburgh or scale).af.	470093
55	(friedman or surgical or staging).mp.	887171
56	STOP-Bang.af.	2
57	Sleep Apnea, Obstructive/di	2379
58	or/45-57	3098114
	General Diagnostic Tests	
59	exp "sensitivity and specificity"/	320624
60	exp Predictive Value of Tests/	104513
61	exp ROC CURVE/	17069
62	exp Mass Screening/	102939
63	exp diagnosis/	5190511
64	exp REPRODUCIBILITY OF RESULTS/	203415
65	exp false negative reactions/ or false positive reactions/	30462
66	predictive value.tw.	44461
67	(sensitivity or specificity).tw.	582712
68	accuracy.tw.	162713
69	screen\$.tw.	350059
70	diagno\$.tw.	1333212
71	roc.tw.	12925
72	reproducib\$.tw.	90416
73	(false positive or false negative).tw.	39967
74	likelihood ratio.tw.	4450
75	accuracy.tw.	162713
76	di.fs.	1684695
77	or/59-76	6907870
	Group 1: Comparative studies on sleep apnea and treatment	
78	7 and 30 and 44	6008
79	limit 78 to english language [Limit not valid in CCTR; records were retained]	5354
80	limit 79 to humans [Limit not valid in CCTR; records were retained]	4883
81	79 and humans.sh.	4876
82	80 or 81	4883
83	remove duplicates from 82	3751

Source: www.effectivehealthcare.ahrq.gov

Published Online: March 15, 2010

	Group 2: All but comparative studies on sleep apnea and treatment (with the exclusion of selected study designs)	
84	7 and 44	11488
85	84 not 83	7737
86	limit 85 to english language [Limit not valid in CCTR; records were retained]	5933
87	limit 86 to humans [Limit not valid in CCTR; records were retained]	4943
88	86 and humans.sh.	4942
89	87 or 88	4943
90	remove duplicates from 89	4596
91	limit 90 to (addresses or bibliography or biography or case reports or comment or congresses or consensus development conference or dictionary or directory or festschrift or in vitro or interactive tutorial or interview or lectures or legal cases or legislation or news or newspaper article or overall or patient education handout or periodical index or portraits or "scientific integrity review" or twin study) [Limit not valid in CCTR; records were retained]	913
92	90 not 91	3683
	Group 3: Comparative studies on sleep apnea and sleep apnea diagnosis	
93	7 and 30 and 58	6424
94	limit 93 to english language [Limit not valid in CCTR; records were retained]	5775
95	limit 94 to humans [Limit not valid in CCTR; records were retained]	5492
96	94 and humans.sh.	5453
97	95 or 96	5492
98	remove duplicates from 97	4295
99	98 not (83 or 92)	2477
	Group 4: General diagnostic studies on sleep apnea and sleep apnea diagnosis (excludes Gp 3 above)	
100	7 and 58 and 77	11451
101	limit 100 to english language [Limit not valid in CCTR; records were retained]	9709
102	limit 101 to humans [Limit not valid in CCTR; records were retained]	9209
103	101 and humans.sh.	9199
104	102 or 103	9209
105	104 not (83 or 92 or 99)	3623
106	remove duplicates from 105	3296
	Group 5: Systematic Reviews and Practice Guidelines on Sleep Apnea	
107	limit 7 to (guideline or meta analysis or practice guideline)	151
108	7 and Cochrane Database of Systematic Reviews.jn.	25
109	107 or 108	170
110	remove duplicates from 109	129
111	110 not (83 or 92 or 99 or 106)	41