Evidence-based Practice Center Systematic Review Protocol

Project Title: *Continuous Positive Airway Pressure Treatment for Obstructive Sleep Apnea in Medicare Eligible Patients*

I. Background

Sleep apnea is a common disorder that affects people of all ages. It is characterized by periods of airflow cessation (apnea) or reduced airflow (hypopnea) during sleep. Sleep apnea may be caused by mechanical obstruction of the airways resulting in disturbed airflow patterns, by a central loss of respiratory drive, or a combination of the two (mixed). Obstructive sleep apnea-hypopnea syndrome, more commonly called obstructive sleep apnea (OSA), is the most common type of sleep apnea.[1]

Definition and severity of obstructive sleep apnea

Sleep apnea is primarily diagnosed with sleep tests that measure sleep time (and often other sleep measures) and respiratory events. OSA is distinguished from central sleep apnea by the presence of respiratory effort during episodes of apnea and hypopnea (in contrast with central sleep apnea where respiratory effort is lacking). The severity of OSA is typically quantified by the sum of the number of apneas and hypopneas per hour of sleep, a quantity that has been termed the apnea-hypopnea index (AHI). AHI is often used as part of both diagnosis (and, thus, study inclusion criteria) and as a surrogate measure for health outcomes (also called an intermediate outcome) in studies. The American Academy of Sleep Medicine (AASM) publishes scoring manuals for AHI and other physiological events to characterize OSA. In the United States, AASM is the predominant accrediting institution for sleep laboratories. AASM first published its scoring manual in 1999 (known as the "Chicago" Criteria).[2] They have amended their definitions of breathing events, sleep time, and how these are measured multiple times since their first set of criteria, with major revisions in 2007,[3] and in 2012.[4] Minor revisions have been made almost annually since (the current version is v2.6,[5] released in 2020). Notably, while the AASM has used an evidence-based approach (i.e., making recommendations based on systematically reviewed evidence) to guide their selection and revision of criteria, the large majority of their recommendations (scoring rules) are at the level of consensus of the panel members because of insufficient evidence to support specific criteria.[5] Further complicating the definition of OSA (and evaluations of severity), studies commonly use other international criteria and the application of specific definitions vary even within specific scoring manuals. Examples include whether 90 or 100 percent cessation of airflow is required to define apnea and whether a 3 or 4 percentage point drop in oxygen saturation and/or a 30 or 50 percent reduction in airflow is required to define hypopnea. In addition, respiratory effort related arousals (RERA) from sleep may be allowed as an alternative to desaturation to define hypopnea. When RERAs are measured instead of desaturation, one measures the "respiratory disturbance index" (RDI) in contrast to the AHI.

The variations in how OSA is defined result in variations across studies in which patients are included and how treatments are provided which in turn makes interpretation of studies difficult.

The International Classification of Sleep Disorders (ICSD) has, since 2005, defined OSA as either 1) ≥15 predominantly obstructive respiratory events (apneas, hypopneas, or RERAs) per hour in asymptomatic, otherwise healthy individuals, or 2) ≥5 predominantly obstructive respiratory events per hour in individuals with symptoms (e.g., nonrestorative sleep, waking with gasping, reported breathing interruptions) or certain comorbidities (i.e., hypertension, a mood disorder, cognitive dysfunction, coronary artery disease, stroke, congestive heart failure, atrial fibrillation or type 2 diabetes mellitus).[6, 7] These criteria, though, do not distinguish three populations of adults diagnosed with OSA: 1) those with frequent respiratory disturbances but who do not have symptoms of OSA such as daytime sleepiness, 2) those who have symptoms of OSA such as daytime sleepiness but who may have relatively less frequent respiratory disturbances, and 3) those who have the comorbidities listed above but who also may have relatively less frequent respiratory disturbances. Despite clear differences in the groups of patients (with or without symptoms/comorbidities), each is diagnosed and treated as if they have equivalent conditions.

Treatment of OSA

The most common first-line therapy for OSA is the use of continuous positive airway pressure (CPAP) devices during sleep.[8] The CPAP machine directly relieves the obstruction by counteracting airway narrowing through the delivery of compressed air (under pressure) to the oropharynx, thereby splinting the airway (keeping it open with increased air pressure). The effectiveness of a specific type of CPAP device may vary depending on differences in diagnostic and AHI scoring criteria, severity of disease, comorbidities, and other factors.

As of 2008, the Centers for Medicare and Medicaid Services (CMS) covers an initial 12 week trial of "CPAP in adult patients with OSA if either of the following criteria is met: (1) AHI or RDI ≥15, or (2) AHI or RDI ≥5 and ≤14 with documented symptoms of excessive daytime sleepiness, impaired cognition, mood disorders or insomnia, or documented hypertension, ischemic heart disease, or history of stroke".[9] Of note, in contrast with how apneas and hypopneas are defined in at least some studies, CMS requires that "Apnea is defined as a [100%] cessation of airflow…[and h]ypopnea is defined as an abnormal respiratory event… with at least a 30% reduction in …airflow... with at least a 4% oxygen desaturation." The 2001 CMS Coverage Decision Memorandum for CPAP, which had substantively the same criteria as current CMS policy, noted that their criteria were derived from the inclusion criteria of studies of CPAP devices and that there was not any direct evidence to support the use of the criteria.[10]

Non-CPAP treatments for OSA include dental and mandibular devices to improve oral airway obstruction, along with a range of surgical treatments, including implanted structural supports to reduce obstruction. Other nonsurgical interventions used to treat OSA include devices to alter sleep position (positional therapy), physical therapy to improve oropharyngeal muscle tone, complementary and alternative medicine techniques, pharmacological agents (including ventilatory stimulants or rapid eye movement sleep suppressants), and nerve stimulation. For specific groups of patients, other interventions include atrial overdrive pacing for patients with nocturnal bradycardia, weight loss interventions (including bariatric surgery), and various surgical interventions exist that aim to alter the anatomy of the air passages to alleviate

postulated obstructive mechanisms in various patients.[11] These specialized interventions are not first line treatments, are not a direct comparator to CPAP for the majority of incident patients, and thus are not a focus of this review.

AHI as a surrogate or intermediate outcome

While AHI and related measures are used to diagnose patients with OSA and evaluate its severity, they are essentially laboratory measures. . From a patient's point of view, health outcomes caused by OSA are more important. These include cardiovascular events, quality of life, changes in cognitive function, and symptoms including sleepiness (as measured by a sleep questionnaire) and sequelae such as motor vehicle accidents, among other outcomes. Because AHI is commonly used to evaluate the mechanical effectiveness of CPAP (and other treatments)—i.e., whether it is reducing episodes of apnea and hypopnea—and because CPAP (when used properly) immediately affects AHI, it is the most commonly reported outcome and clinical outcomes are more rarely reported.[11] Studies have demonstrated that CPAP improves AHI as defined in those studies and other surrogate or intermediate measures of OSA severity and measures of sleepiness,[11] but questions remain about the effectiveness of CPAP to reduce or improve clinical outcomes (e.g., cardiovascular events, stroke, mortality).

A large randomized trial of long-term CPAP use (SAVE) in people with coronary or cerebrovascular disease was recently published. Despite improvements in AHI as defined in that study, the study found no improvement of cardiovascular, kidney, and weight outcomes,[12-14] raising questions of whether change in AHI as defined in that study is a valid intermediate or surrogate outcome for the patient-centered clinical outcomes and whether, or in whom, CPAP may be a clinically effective treatment modality.

Impact of Comorbid Conditions

It is important for clinical decision-making to understand whether the effect of CPAP differs based on the presence of comorbidities such as obesity.is an important risk factor for OSA. Higher body mass index (BMI) is strongly associated with increased risk of OSA.[15] Notably, many of the health outcomes associated with OSA (and for which treatment with CPAP is aimed to reduce) are also strongly associated with obesity, including metabolic syndrome/prediabetes, type 2 diabetes, hypertension, cardiovascular disease, stroke, and all-cause mortality, among others.[16]

Purpose of the Review

CMS nominated the topic to the Agency for Healthcare Research and Quality (AHRQ) for a Technology Assessment (TA) in order to critically appraise evidence on improvement of health outcomes with CPAP treatment and for the validity of criteria used as surrogate outcomes, e.g., AHI.

II. Contextual and Key Questions

Contextual Questions

CQ 1: What measures related to apneas and hypopneas (e.g., apnea indices, hypopnea indices, and apnea-hypopnea indices with various measurements) or other measures (e.g., time spent with oxygen saturation below 90% or other cutoffs, electrophysiologic signal analysis metrics such as time and frequency domain analyses of heart beats) are used in contemporary research and clinical settings? How have standard definitions of these measures changed over time and what is the explanation for such changes?

CQ 2: What are commonly used sleep questionnaires and how have they been validated?

CQ 3: What treatment modalities for OSA are currently being marketed in the US? What OSA treatments (experimental or approved) are currently being investigated in ongoing trials for patients as an alternative to CPAP?

CQ 4: What are the variable features of marketed CPAP devices?

CQ 5: What are the patient-centered health outcome goals and symptom relief goals of CPAP devices?

Key Questions

KQ 1: What is the efficacy, effectiveness, comparative effectiveness, and harms of CPAP devices to improve *clinically significant outcomes*?

> **KQ 1a:** How are respiratory disturbance events (apnea, hypopnea, arousal) defined in each study? What are the diagnostic criteria for OSA (or criteria to treat with CPAP) in each study? How do the diagnostic criteria relate to time of AASM criteria? Do treatment effects of CPAP differ by the specific diagnostic criteria used within or across studies?

> **KQ 1b:** What is the within-study concordance in CPAP trials among apnea and hypopnea indices (e.g., AHI), sleep questionnaires (e.g., Epworth Sleepiness Scale), and clinically significant outcomes?*

> **KQ 1c**: Do the clinical effects or harms of specific CPAP devices differ by patient subgroups, duration of followup, or particular CPAP features?

> **KQ 1d**: Summarize the methodological issues in the existing studies.

KQ 2: What is the evidence that apnea and hypopnea-based measures of sleep-disordered breathing (e.g., apneic indices, hypopnea indices, and apnea-hypopnea indices) used in current practice and research are valid surrogate or intermediate measures for clinically significant outcomes?

KQ 2a: Summarize the methodological issues in the existing studies. What is the ideal study design for establishing the validity of a surrogate or intermediate measure?

* Note that the association between changes in apnea and hypopnea indices and clinical outcomes across a broader set of studies is primarily addressed in KQ 2.

Systematic Review Study Eligibility Criteria

Eligibility Criteria Relevant to <u>Both KQs</u>
Population
- Adults (>18 years)
- *Exclude* studies with any pregnant women
- *Exclude* studies in which any participants are reported to have, at baseline, central sleep apnea (from any cause including prior stroke, severe heart failure, among others), obesity hypoventilation syndrome (Pickwickian syndrome), neuromuscular disease, Parkinson disease, Down syndrome, Prader-Willi syndrome, major congenital skeletal abnormalities, narcolepsy, narcotic addiction, Alzheimer disease, epilepsy and or with mild cognitive impairment

Intervention/Comparator
- *Exclude* studies of surgical interventions for OSA or bariatric surgery

Outcomes
- Hard clinical outcomes
    - o Major clinical outcomes
        - Death
        - Cardiovascular and cerebrovascular events or incident diagnosis
        - Motor vehicle accidents
        - Composite outcomes that include only major clinical outcomes (e.g., major adverse cardiovascular events defined as including all-cause mortality)
    - o Other patient-centered and/or clinically significant outcomes
        - Other cardiovascular outcomes
            - Objective measures of cardiovascular severity (categorized, not continuous measures such as intima media thickness)
            - Incident hypertension (or regression to normotension)
            - Arrhythmias
            - Incident arrhythmias (or resolution of arrhythmias)
            - Clinically significant ventricular arrhythmias
            - Atrial fibrillation
        - New-onset diabetes mellitus or prediabetes (or regression to normoglycemia)
        - Mental health conditions, including depression, anxiety, and substance use disorder: incident diagnosis or resolution

- Cognitive function: clinical diagnosis (e.g., of dementia) or validated executive function measures
- Quality of life and functional outcomes (validated measures)
- Sexual function: clinical diagnosis (e.g., diagnosis of erectile dysfunction or anorgasmia) or their resolution
- Sequelae of sleep deprivation (e.g., trauma, missed work or school)
- Other clinically significant outcomes reported in studies or as found for CQ 5
- Exclude
  - Blood pressure
  - Asymptomatic arrhythmias or laboratory measures (e.g., captured by electrophysiologic testing [heart rate variability, QTc interval, etc.])
  - Glycemia measures (e.g., hemoglobin A1c, fasting blood glucose)
  - Instruments to measure severity of OSA (including AHI and sleepiness)
- Minimum duration for associations with death, incident cardiovascular events, hypertension, or diabetes is 1 year
- Minimum duration for all other outcomes is 6 months

Mediators of treatment effect (or association) (E.g., factors to be evaluated in subgroup analyses) Note that these are not eligibility criteria, but are factors that will evaluated to potentially explain different findings across studies; e.g., by subgroup analysis, regression, or other methods to evaluate heterogeneity of treatment effect)
- Body weight/obesity/neck circumference, etc.
- Weight change (loss or gain)
- Prior cardiovascular, cerebrovascular, or other major clinical disease/condition
- Sex/gender
- Race/ethnicity
- Severity of OSA (as defined by study)
- Other mediators as reported in primary studies

Setting
- Outpatient only (except for sleep laboratory setting for measurement of sleep and breathing measures)
- *Exclude* acute care hospital settings (including perioperative)

Additional Eligibility Criteria Specific to KQ 1

Populations
- As listed above, for both KQs

Intervention (CPAP)
- CPAP for treatment (not diagnosis or staging) of OSA
  - At least 1 month of prescribed (planned) treatment

- *Exclude* intervention designed only to improve CPAP compliance/adherence (i.e., not an intervention of CPAP, *per se*)
- *Exclude* evaluations of accessories only (e.g., nasal cannulas, head straps, humidifiers)
- *Exclude* evaluation of CPAP titration methods, *per se*, including specific parameters or modes (e.g., starting pressures)
- *Exclude* evaluations of other features meant to improve comfort or adherence
- *Exclude* other non-CPAP interventions (e.g., different times of monitoring, scoring), including noninvasive ventilation

Comparators
- No CPAP
- Non-CPAP active interventions for OSA (e.g., mandibular advancement device)
  - *Exclude* bariatric surgery (as a comparator treatment)
  - *Exclude* surgical OSA procedure (as a comparator treatment)
- Other CPAP modality or protocol (e.g., autoCPAP vs. bilevel CPAP)

Exclude comparisons with different accessories, titration methods, features to improve comfort or adherence, other non-CPAP interventions (e.g., different times of monitoring, scoring), including noninvasive

Outcomes
- As listed above, for both KQs
- Sleep and breathing measures (e.g., AHI) and validated sleep questionnaires (e.g., Epworth Sleepiness Scale) (only for the purpose of addressing KQ 1b, not as outcomes of interest)
- Adverse events related to CPAP use

Mediators of treatment effect (E.g., subgroup analyses; see note above about mediators)
- As listed above, for both KQs
- New or prior OSA diagnosis
- Treatment naïve versus failed prior treatment
- First versus second or more use of CPAP
- Treatment (CPAP) compliance
- Treatment (CPAP) discontinuation

Design
- Randomized controlled trials (RCT)
  - Consider whether study met power calculation for the outcome(s) of interest (including adverse events)
- Nonrandomized comparative studies (NRCS)
  - Restrict to studies that use modeling or other analytic methods to minimize confounding bias (due to inherent differences between people who receive one or the other intervention)
  - *Exclude* case-control design
  - *Exclude* "pre-post" design (comparison of before and after CPAP treatment in a single group of participants)

- Longitudinal
  - *Exclude* cross-sectional

Additional Eligibility Criteria <u>Specific to KQ 2</u>

For KQ 2, we will include studies that measure a change in the intermediate/surrogate measure (e.g., AHI) over a period of time and report on outcomes of interest. We will include studies that provide formal evaluation of validity of the intermediate/surrogate measure for the clinical outcome and other studies that report sufficient data to analyze a potential association between the change in the measure and the clinical outcome.

Population
- Adults
  - Do not require a diagnosis of OSA (for evaluations of associations of measures)
  - *Exclude* populations as described under "Eligibility Criteria relevant to Both KQs"

Intermediate/Surrogate measures (variables of interest evaluated regarding their association with clinical outcomes)
- Sleep and breathing measures
  - Indices based on apneas or hypopneas (e.g., AHI, RDI) or other respiratory events such as RERAs, oxygen desaturations
- *Exclude* evaluations of isolated neurophysiologic parameters of sleep (e.g., respiratory effort, heart rate, air flow, pulse oximetry alone) and cardiac electrophysiology indices (e.g., heart rate variability)

Outcomes
- As listed above, for both KQs
- Each study must report both one or more intermediate/surrogate measures (i.e., sleep and breathing measures) *and* one or more outcomes of interest

Additional mediators of association (e.g., analyzed by subgroup analyses)
- As listed above, for both KQs
- Definition of sleep and breathing measure

Study Design
- Longitudinal studies informing on person-level associations of sleep and breathing measure(s) with outcome(s)
  - Patient-level association between *change* in measure and *change or incident* outcome (i.e., evaluation of association reported within study)
  - *Exclude* cross-sectional studies
- Comparative or noncomparative (single group) studies
- N ≥30 analyzed for a given association between intermediate/surrogate measure and outcome

III. Analytic Framework

To guide the development of the Key Questions for the diagnosis and treatment of OSA, we developed an analytic framework (**Figure 1**) that maps the specific linkages associating the populations and subgroups of interest, the intervention, and outcomes of interest, both intermediate and clinically significant. Specifically, this analytic framework depicts the chain of logic that evidence must support to link the interventions to improved health outcomes. The figure lays out which Key Questions address each aspect of the framework

Figure 1. Analytic Framework
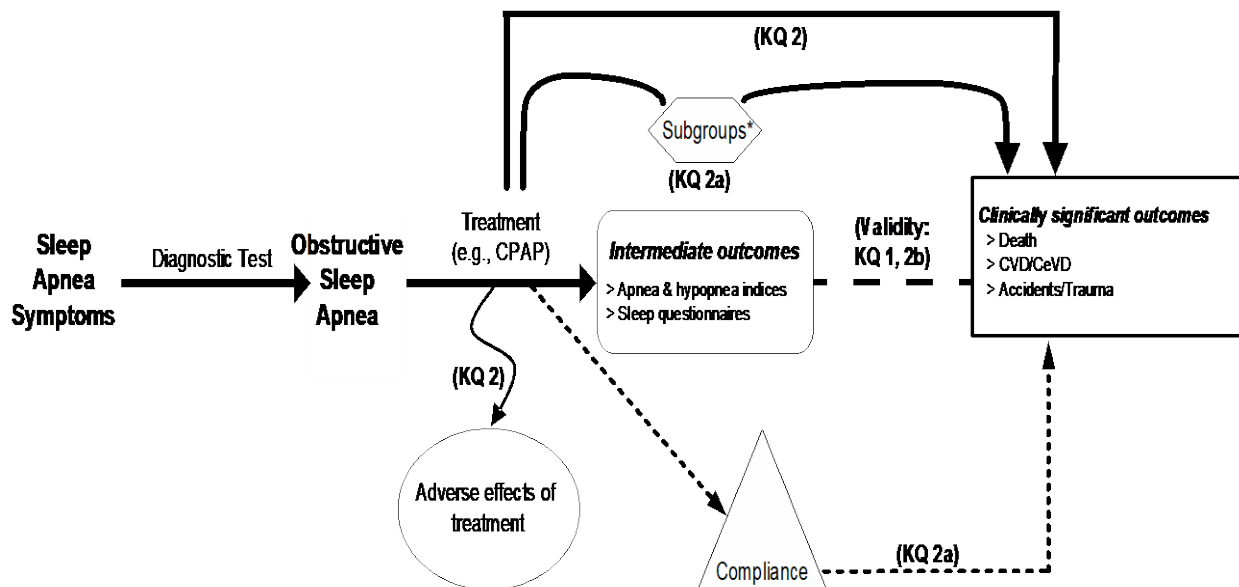
CeVD = cerebrovascular disease; CPAP = continuous positive airway pressure treatment; CeVD = cerebrovascular disease event; CVD = cardiovascular disease event; KQ = key question.

* List of subgroups (etc.) of interest, to include:
- Obesity and weight groups
- Comorbidities (e.g., cardiovascular disease, diabetes)
- Sex/gender
- Race/ethnicity
- Diagnostic and scoring criteria
- Severity of OSA
- Others to be determined *post hoc* based on reported subgroup analyses

IV. Methods
Contextual Questions and Key Question 2a

To address the CQs, we will follow standard methods used by AHRQ, which follows the general guidance of the U.S. Preventive Services Task Force. [17] We will conduct general nonsystematic searches in PubMed and of the internet for existing systematic reviews (SRs), guidelines,

editorials, narrative reviews, policy statements, and other potentially relevant information sources. During abstract screening (for the SR), we will also identify any potentially relevant studies that are opportunistically found.

For CQ 1, we will review guidelines, scoring manuals, narrative reviews, and the studies included in the SR. For CQ 1, we will also tabulate changes in ICSD and AASM criteria and definitions.

For CQ 2, we will review existing systematic reviews and guidelines.

For CQs 3, 4, and 5, we will search the FDA website, pulmonary society and OSA organization websites, and manufacturer websites for marketed CPAP devices and their features. We will search clinicaltrials.org for ongoing trials.

To address KQ 2a regarding the ideal study design to establish validity of a surrogate or intermediate measure, we will describe major alternative ways of thinking about surrogate and intermediate outcomes, including the Prentice framework,[18] causal mediation analysis,[19] and principal stratification analysis.[20]

Systematic Review

Literature Search: We will search MEDLINE (via PubMed), Embase, Cochrane databases, CINAHL, ClinicalTrials.gov, and Epistemonikos for primary studies, existing SRs, and published guidelines.

We will also search the ECRI guidelines Trust[21] for relevant guidelines published in the last 5 years and the FDA medical device databases.[22] To ensure availability for future researchers , we will create the Evidence Map in the Systematic Review Data Repository (SRDR, https://srdr.ahrq.gov/). We will also clearly identify where published literature is unavailable. Separate, overlapping searches will be conducted for each KQ. For KQ 2 (CPAP efficacy), we will search all listed databases. Duplicate citations will be removed prior to screening. *De novo* searches will be restricted to 2010 or later. To capture literature published prior to 2010, we will rescreen for eligibility all studies that were included in our previous systematic reviews on OSA diagnosis and treatment.[11, 23-25] These prior systematic reviews included, at a minimum, studies that meet the eligibility criteria for the current review. (Of note, the prior systematic reviews will be used only to identify studies; all eligible studies will be re-extracted and re-analyzed anew.)

Literature search strategies include filters to remove nonhuman studies and articles that are not primary studies, systematic reviews, or clinical practice guidelines. The searches will include MeSH or Emtree terms, along with free-text words, related to OSA and CPAP. Search strategies will be peer reviewed by an independent medical librarian. The planned MEDLINE search strategies is included in Appendix A; the strategies for other databases will be adapted from this strategy.

The reference lists of relevant existing SRs (including our previous systematic reviews on OSA diagnosis and treatment.[11, 23-25]) will be screened for additional eligible studies. Additional articles suggested to us from any source, including peer and public review, will be screened applying identical eligibility criteria.

Searches will be updated during the public posting period.

Citations from all electronic databases will be entered into Abstrackr software (http://abstrackr.cebm.brown.edu/) to enable abstract screening. The team will conduct one or more rounds of pilot screening, during which all members of the team will screen the same 100 abstracts and discuss conflicts, with the goals of training the team in the nuances of the eligibility criteria and refining the criteria as needed. Thereafter, we will screen all remaining abstracts in duplicate. The Abstrackr software has machine learning capabilities that predict the likelihood of relevance of each citation. Daily, the list of unscreened abstracts will be sorted so that most potentially-relevant articles are presented first. This process will make screening more efficient and will enable us to capture the large majority of relevant articles relatively early in the abstract screening process. We will consider the possibility of stopping screening early if the likelihood of the remaining unscreened papers being relevant is very low (e.g., if the maximum prediction score of the unscreened citations is <0.40). Once Abstrackr's predictions indicate that there are no relevant papers remaining among the yet unscreened ones, we will stop screening if there are no eligible citations identified in a consecutive sample of 370 consecutive citations (sample size chosen because the upper 97.5% confidence interval bound for a proportion of 0/370 is less than 1%).

Potentially relevant citations will be retrieved in full text and rescreened.

Data Extraction and Data Management: Eligible studies will be data extracted into the Systematic Review Data Repository-Plus (SRDR+) software. Each article will be extracted by one researcher and extracted data will be confirmed by a second, independent researcher. Individual studies with multiple publications will be extracted as a single study (with a single entry in SRDR+). For each study, we will extract publication identifying data, study design features (including funding source), population characteristics, intervention and comparator (or measure) names and descriptions, relevant outcomes and their definitions, results, and information necessary for risk of bias, generalizability, and strength of evidence assessments. For KQ 1, results analyses may be extracted into a separate database (e.g., a spreadsheet); once completed, these files will be uploaded into SRDR+.

Assessment of Methodological Risk of Bias of Individual Studies: We will evaluate each study for risk of bias by assessing the risk of individual bias domains and integrating them in an overall risk of bias assessment. At a high level, study Risk of Bias assessments involve comparing a study with an *ideal study* that would have the same purpose (e.g., with an idealized RCT, when the purpose is treatment effect estimation) and judging the importance of major deviations between the study at hand and the ideal study. We will structure these assessments as follows:

For RCTs, we will use the Cochrane Risk of Bias 2.0 tool,[26] focusing on issues related to randomization and allocation concealment methodology; patient, caregiver, and outcome assessor blinding; loss to followup (omissions from analyses); adequacy of descriptions of study participants, interventions, and outcomes; and other issues. Questions related to outcome assessor blinding, loss to followup, and reporting adequacy will be assessed for each outcome.

For CPAP harms, we will assess specific elements from the McMaster tool for assessing quality of harms assessment and reporting in study reports (McHarm) pertaining to prespecification, definitions, adjudication, and completeness of reporting of harms.[27, 28]

For studies addressing KQ 2 we will assess whether they have used a formal mediation analysis according to one of three well-known analysis frameworks, namely, the Prentice framework,[18] causal mediation analysis,[19] and principal stratification analysis.[20] Briefly, at a high level, all three frameworks examine the following logic from different angles:

- To establish that, say a specific measure of AHI, is a valid intermediate endpoint for a specific clinical endpoint (e.g., strokes at 1 year) one has to show that a manipulation of AHI levels (e.g., by using CPAP) corresponds to a change in the clinical outcome.
- Depending on the type of analysis, a goal may be to estimate how much of the *total* effect of the intervention on the outcome is
  - an *indirect effect* (i.e., "is mediated or explained" by the change in the intermediate outcome) versus
  - a *direct effect* (i.e., is not mediated by changes in the intermediate outcome).
- The validity of an intermediate outcome would have to be examined for every intervention/clinical outcome pair. For example, if there exists a mediation analysis that supports the concept that AHI is as a valid intermediate for the effect of CPAP on strokes, it cannot be extrapolated to support that AHI is a valid intermediate outcome for the effect of CPAP on myocardial infarctions or workplace accidents; or for the effects if mandibular advancement devices on strokes.

Data Synthesis: Each study included in the review will be described in summary and evidence tables presenting study design features, study participant characteristics, descriptions of interventions, outcome results, and risk of bias/methodological quality.

For Key Question 1, we will compare CPAP interventions to their comparators, for their effects, primarily with odds ratios (ORs) or hazard ratios (HRs) of event rates, "net differences" (between-intervention comparison of within-intervention changes or difference-in-difference measures) of continuous outcomes with both pre- and post-intervention data, and differences (between interventions) in continuous outcome data post-intervention (if pre-intervention data are not reported). We will also report or calculate number needed to treat/harm. We will compare results with available minimal clinical important differences. We will examine the feasibility of making multiple comparison adjustments based on information provided in the protocol (e.g., related to number of evaluated measurement time points).

For Key Question 2, we will focus on assessments of surrogacy or causal mediation for the intermediate outcome with respect to each clinically significant outcome for each combination of population and treatment status. We will summarize evidence for each combination. If at least 4 studies exist that use the same analytical approach we will meta-analyze study-specific estimates using standard approaches.

Where there are sufficient studies reporting sufficiently similar results, we plan to meta-analyze these comparisons. We expect to summarize harms data semiquantitatively (i.e., without meta-analysis).

For both Key Questions, we will evaluate potential mediators of effect or association that may explain differences in effects or associations found across or within studies. We will extract, summarize, and analyze within-study analyses (such as subgroup or regression analyses) and across-study analyses as data allow. As feasible, we will quantitatively combine and analyze subgroups across studies, but will avoid ecological fallacy (e.g., comparing studies of, on average, younger and older patients). We will summarize reported analyses of potential mediators of effects (e.g., subgroup analyses), but will critique these in terms of their analytic validity (e.g., whether the analysis was planned *a priori*, whether analytic methods to compare subgroups were appropriate) and whether the subgroup definitions are appropriate or valid (e.g., how OSA severity was categorized).

For all analyses, we will conduct sensitivity analyses, as feasible, based on features such as study publication status (e.g., peer reviewed vs. non-peer reviewed), fidelity to the eligibility criteria (e.g., excluding and including studies with small numbers of participants with mixed sleep apnea); similarity of studies regarding features such as patient characteristics, CPAP type, definitions (and validity) of measures (e.g., AHI), definitions of outcomes; duration of followup; mediators (e.g., compliance rates), and study design (e.g., whether or not outcome assessors were blinded to intervention).

Grading the Strength of Evidence for Major Comparisons and Outcomes: Following AHRQ Methods guidance, we will evaluate the strength of evidence (SoE) addressing each major comparison or evaluation for each KQ. We expect that evaluation of the strength of evidence will include:

- Adequacy of reporting of criteria for OSA diagnosis
- Ascertainment of clinically significant outcomes
- Validity of surrogate or intermediate outcomes.
- Clinical effects of CPAP
- Harms of CPAP

Assessing Applicability: For each Key Question (or specific subquestion), we will assess the applicability of the included studies to people eligible for Medicare coverage (based on age or co-existing disability). Applicability will be assessed primarily based on the studies' eligibility criteria and their included participants, specifically related to such factors as severity of disease, prior history, age, sex, and race/ethnicity. These will be qualitatively compared with typical distributions of these factors among Medicare beneficiaries. Other factors may include the age and geographic location of the study.

References
1. Practice parameters for the indications for polysomnography and related procedures. Polysomnography Task Force, American Sleep Disorders Association Standards of Practice Committee. Sleep. 1997 Jun;20(6):406-22. PMID: 9302725.
2. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep. 1999 Aug 1;22(5):667-89. PMID: 10450601.

3. Iber C, Ancoli-Israel S, Chesson AL, et al. AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specification, Version 1. American Academy of Sleep Medicine, Westchester, IL. 2007.

4. Berry RB, Brooks R, Gamaldo CE, et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0. Darien, IL: American Academy of Sleep Medicine. 2012.

5. Berry RB, Quan SF, Abreu AR, et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.6, www.aasmnet.org, American Academy of Sleep Medicine, Darien, IL. 2020.

6. Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea?: The Rational Clinical Examination systematic review. Jama. 2013 Aug 21;310(7):731-41. doi: 10.1001/jama.2013.276185. PMID: 23989984.

7. American Academy of Sleep Medicine. International classification of sleep disorders, 3rd ed. Darien, IL: American Academy of Sleep Medicine. 2014.

8. Loube DI, Gay PC, Strohl KP, et al. Indications for positive airway pressure treatment of adult obstructive sleep apnea patients: a consensus statement. Chest. 1999 Mar;115(3):863-6. doi: 10.1378/chest.115.3.863. PMID: 10084504.

9. CMS. National Coverage Determination (NCD) for Continuous Positive Airway Pressure (CPAP) Therapy For Obstructive Sleep Apnea (OSA) (240.4). https://www.cms.gov/medicare-coverage-database/details/ncd-details.aspx?ncdid=226&ver=3. Accessed June 9, 2020. 2008.

10. CMS. Decision Memo for Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA) (CAG-00093N). https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=19&fromdb=true. Accesssed June 9, 2020. 2001.

11. Balk EM, Moorthy D, Obadan NO, et al. AHRQ Comparative Effectiveness Reviews. Diagnosis and Treatment of Obstructive Sleep Apnea in Adults. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.

12. McEvoy RD, Antic NA, Heeley E, et al. CPAP for Prevention of Cardiovascular Events in Obstructive Sleep Apnea. N Engl J Med. 2016 Sep 8;375(10):919-31. doi: 10.1056/NEJMoa1606599. PMID: 27571048.

13. Ou Q, Chen B, Loffler KA, et al. The Effects of Long-term CPAP on Weight Change in Patients With Comorbid OSA and Cardiovascular Disease: Data From the SAVE Trial. Chest. 2019 Apr;155(4):720-9. doi: 10.1016/j.chest.2018.08.1082. PMID: 30268694.

14. Loffler KA, Heeley E, Freed R, et al. Effect of Obstructive Sleep Apnea Treatment on Renal Function in Patients with Cardiovascular Disease. Am J Respir Crit Care Med. 2017 Dec 1;196(11):1456-62. doi: 10.1164/rccm.201703-0603OC. PMID: 28743190.

15. Seicean S, Kirchner HL, Gottlieb DJ, et al. Sleep-disordered breathing and impaired glucose metabolism in normal-weight and overweight/obese individuals: the Sleep Heart Health Study. Diabetes Care. 2008 May;31(5):1001-6. doi: 10.2337/dc07-2003. PMID: 18268072.

16. Guh DP, Zhang W, Bansback N, et al. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. BMC Public Health. 2009 Mar 25;9:88. doi: 10.1186/1471-2458-9-88. PMID: 19320986.

17. U.S. Preventive Services Task Force. Procedure Manual. https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes. 2015.

18. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989 Apr;8(4):431-40. doi: 10.1002/sim.4780080407. PMID: 2727467.
19. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. Annu Rev Public Health. 2016;37:17-32. doi: 10.1146/annurev-publhealth-032315-021402. PMID: 26653405.
20. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002 Mar;58(1):21-9. doi: 10.1111/j.0006-341x.2002.00021.x. PMID: 11890317.
21. ECRI Guidelines Trust. https://guidelines.ecri.org. 2019.
22. U.S. Food and Drug Administration. Medical Device Databases. https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/medical-device-databases. 2020.
23. Trikalinos TA, Lau J. AHRQ Technology Assessments. Obstructive Sleep Apnea-Hypopnea Syndrome: Modeling Different Diagnostic Strategies. Rockville (MD): Agency for Healthcare Research and Quality (US); 2007.
24. Trikalinos TA, Ip S, Raman G, et al. AHRQ Technology Assessments. Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome. Rockville (MD): Agency for Healthcare Research and Quality (US); 2007.
25. Ip S, D'Ambrosio C, Patel K, et al. Auto-titrating versus fixed continuous positive airway pressure for the treatment of obstructive sleep apnea: a systematic review with meta-analyses. Syst Rev. 2012 Mar 8;1:20. doi: 10.1186/2046-4053-1-20. PMID: 22587875.
26. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Bmj. 2019 Aug 28;366:l4898. doi: 10.1136/bmj.l4898. PMID: 31462531.
27. Chou R, Aronson N, Atkins D, et al. Assessing harms when comparing medical interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]. Agency for Healthcare Research and Quality (US); 2008.
28. Santaguida P, Raina P. The Development of the McHarm Quality Assessment Scale for adverse events: Delphi Consensus on important criteria for evaluating harms. 2008.