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

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. 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).

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 conte

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 Both KQs

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
  - 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 cardiac mortality)
  - Other patient-centered and/or clinically significant outcomes
    - Other cardiovascular outcomes
      - Objective measures of cardiovascular severity (categorical, 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
-

- 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 1, 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 Specific to KQ 2

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

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 CQ2, 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, 23] 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]

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 screened.

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 fond inttif4.9(t)-16(-1(act)-6(-1(act.)4(de)4(d i)-3m(Fma)3(or)3[(R)-3(C)-3(T)2(s)-1(w))-2(e)4(w)2(i)-2(l)

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 out1 exvsu5(k)2(s)5( )]TJ 0 Tc 0 Tw ( )Tj EMC /LBody<</MCID 17>>BDC q 1 0 0 1 38.16 [illegible]

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.