Technology Assessment Program

Long-Term Health Outcomes in Obstructive Sleep Apnea: A Systematic Review of Comparative Studies Evaluating Positive Airway Pressure and the Validity of Breathing Measures as Surrogate Outcomes

Final
Technology Assessment
Project ID: SLPT0919
12/01/2022

Technology Assessment Program

Project ID: SLPT0919

Long-Term Health Outcomes in Obstructive Sleep Apnea: A Systematic Review of Comparative Studies Evaluating Positive Airway Pressure and the Validity of Breathing Measures as Surrogate Outcomes

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 290-2015-00002-I
Task Order No. 75Q80119F32017

Prepared by:
Brown Evidence-based Practice Center
Providence, Rhode Island

Investigators:
Ethan M. Balk, M.D, M.P.H.
Gaelen P. Adam, M.L.I.S., M.P.H.
Wangnan Cao, Ph.D.
Monika Reddy Bhuma, B.D.S, M.P.H.
Shaun Forbes, Ph.D.
Shivani Mehta, M.P.H.
Orestis Panagiotou, M.D., Ph.D.
Carolyn D'Ambrosio, M.D.
Thomas A. Trikalinos, M.D., Ph.D.

This report is based on research conducted by the Brown Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2015-00002-I, Task Order No. 75Q80119F32017). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help healthcare decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of healthcare services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

AHRQ appreciates appropriate acknowledgment and citation of its work. Suggested language for acknowledgment: This work was based on an evidence report, Continuous Positive Airway Pressure Treatment for Obstructive Sleep Apnea, by the Evidence-based Practice Center Program at the Agency for Healthcare Research and Quality (AHRQ).

Individuals using assistive technology may not be able to fully access information in this report. For assistance, contact EPC@ahrq.hhs.gov.

Suggested citation: Balk EM, Adam GP, Cao W, Reddy Bhuma M, Forbes S, Mehta S, Panagiotou O, D'Ambrosio C. Long-Term Health Outcomes in Obstructive Sleep Apnea: A Systematic Review of Comparative Studies Evaluating Positive Airway Pressure and Validity of Breathing Measures as Surrogate Outcomes. Project ID: SLPT0919. (Prepared by the Brown Evidence-based Practice Center under Contract No. 290-2015-00002-I/Task Order No. 75Q80119F32017.) Rockville, MD: Agency for Healthcare Research and Quality. <Month 202#>. Available at: https://www.ahrq.gov/research/findings/ta/index.html.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of healthcare in the United States.

The Centers for Medicare & Medicaid Services requested this report from the EPC Program at AHRQ. AHRQ assigned this report to the following EPC: Brown Evidence-based Practice Center (Contract Number: 290-2015-00002-I, Task Order No. 75Q80119F32017).

The reports and assessments provide organizations with comprehensive, evidence-based information on common medical conditions and new healthcare technologies and strategies. They also identify research gaps in the selected scientific area, identify methodological and scientific weaknesses, suggest research needs, and move the field forward through an unbiased, evidence-based assessment of the available literature. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for healthcare quality improvement projects throughout the Nation. The reports undergo peer review and public comment prior to their release as a final report.

AHRQ expects that the EPC evidence reports and technology assessments, when appropriate, will inform individual health plans, providers, and purchasers as well as the healthcare system as a whole by providing important information to help improve healthcare quality.

If you have comments on this evidence report, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

David Meyers, M.D.
Acting Director
Agency for Healthcare Research and Quality

Arlene S. Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Craig A. Umscheid, M.D., M.S.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Christine Chang, M.D., M.P.H.
Associate Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Elise Berliner, Ph.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Lionel L. Bañez, M.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Peer Reviewers

Long-Term Health Outcomes in Obstructive Sleep Apnea: A Systematic Review of Comparative Studies Evaluating Positive Airway Pressure and the Validity of Breathing Measures as Surrogate Outcomes

Structured Abstract

Background. Obstructive sleep apnea (OSA) is a disorder characterized by periods of airflow cessation (apnea) or reduced airflow (hypopnea) during sleep. The diagnosis and severity of OSA, and response to therapy, have historically been assessed using the apnea-hypopnea index (AHI). However, several definitions of this measure exist, and the utility of the AHI and associated measures as valid surrogate measures of health outcomes has been questioned. OSA is commonly treated with the use of continuous positive airway pressure (CPAP) devices during sleep. The efficacy of CPAP, including for Food and Drug Administration (FDA) clearance/approval, has been based on changes in AHI, but the long-term effect of CPAP on health outcomes and the role of disease severity (as measured by AHI) or sleepiness symptoms on the putative effect of CPAP are unclear.

Methods. We searched Medline, Embase, the Cochrane databases, CINAHL, and ClinicalTrials.gov from January 2010 through March 22, 2021; we screened reference lists of the 2011 Agency for Healthcare Research and Quality (AHRQ) OSA report and other systematic reviews for earlier studies. We included only randomized controlled trials (RCT) and nonrandomized comparative studies (NRCS) of CPAP that adjusted for potential confounders for CPAP evaluation. We also included other comparative studies that reported both changes in potential intermediate or surrogate measures (e.g., AHI) and effects on health outcomes. All studies had to report effects on prespecified long-term (12 months for most outcomes) health outcomes in adults with OSA. We did not evaluate sleepiness, other symptoms, or intermediate outcomes. We excluded observational studies that did not directly compare treatment options.

Results. The 52 identified studies used highly inconsistent criteria to define breathing measures (apneas, hypopneas, and oxygen desaturation). Definitions of respiratory disturbance events (e.g., apneas, hypopneas) and criteria to define or categorize severity of OSA are highly inconsistent across studies, despite frequent claims of using standard national or international definitions. Possible differences in study findings based on heterogeneity of OSA and sleep study measures could not be elucidated. Among 31 studies that directly compared CPAP and no CPAP (29 studies) or sham CPAP (2 studies), 14 were RCTs (12 intention-to-treat [ITT]) and 17 were NRCSs (11 analyses of CPAP users versus nonusers).

With one exception, RCTs did not find statistically significant effects for any cardiovascular (CV) outcome. RCTs did not provide evidence that CPAP affects the risk of all-cause mortality (summary effect size [ES] 0.89, 95% confidence interval [CI] 0.66 to 1.21), CV mortality (summary ES 0.99, 95% CI 0.64 to 1.53), stroke (summary ES 0.99, 95% CI 0.73 to 1.35), myocardial infarction (summary ES 1.05, 95% CI 0.78 to 1.41), or composite CV outcomes (ES range 0.42 to 1.10 across studies, all statistically nonsignificant); all with low strength of evidence (SoE). The three RCTs that aimed to be powered for composite CV events failed to

show a significant or clinically meaningful effect. The NRCSs, overall, found significant adjusted associations between CPAP use and all-cause mortality. Combining the RCTs and NRCSs, the summary ES for all-cause mortality was ES 0.61 (95% CI 0.49 to 0.76), supporting a low SoE of an association between CPAP use and lower risk of death. Data from the NRCSs did not change other conclusions. Both RCTs and NRCSs provide insufficient evidence regarding the effect of CPAP on the risk of transient ischemic attack, angina, coronary artery revascularization, congestive heart failure, and atrial fibrillation.

RCTs also did not provide evidence that CPAP affects the risk of driving accidents or the risk of incident diabetes (both low SoE), or that CPAP results in clinically significant changes in depression or anxiety scores, executive cognitive function measures, or nonspecific quality of life measures (all low SoE). RCTs provide insufficient evidence regarding the effect of CPAP on incident hypertension, functional status measures, male or female sexual function, or days of work missed. Data from the NRCSs did not change these conclusions.

Eligible studies provided insufficient evidence regarding possible differences in the effect of CPAP based on patient characteristics (such as disease severity or comorbidities), different diagnostic criteria, or whether RCTs were analyzed as ITT or "as-treated".

Eligible studies provided insufficient evidence about adverse events due to CPAP use. Many studies did not collect such data. Adverse events reported in the U.S. Food and Drug Administration (FDA) database mostly related to inadequate humidification, user errors, or device malfunction. No deaths were attributed to CPAP use.

Review of the FDA database found 163 CPAP devices used to treat adults with sleep apnea. The large majority of FDA 510(k) Premarket Notification records cite other previously approved CPAP devices to support claims of equivalence. The available data did not reference clinical studies to support the device manufacturers' claims.

Review of the National Institutes of Health's RePORTER revealed no germane funded trials. Review of ClinicalTrials.gov revealed a single large RCT with a mortality endpoint, but no updating of the site since 2015 and five additional small trials (2 addressing events in patients with paroxysmal atrial fibrillation or hypertensive pulmonary edema; 3 measuring AHI, cognition, or kidney function).

RCTs comparing CPAP and mandibular advancement devices found no differences in depression or anxiety symptoms (low SoE). There was insufficient evidence for other outcomes. RCTs comparing fixed and autoCPAP found no differences in functional status; other long-term outcomes were not reported. No eligible studies evaluated comparisons of other types of CPAP.

No studies have evaluated the validity of changes in intermediate or surrogate measures (such as change in AHI during a clinical trial) as predictors of long-term health outcomes. No studies reported surrogacy or mediation analyses, nor did any compare the concordance of changes in different sleep study and symptom measures with health outcomes. Across the 15 studies that reported both changes in intermediate or surrogate measures and effects on long-term health

outcomes, data were too sparse to allow adequate cross-study evaluation of concordance between change in any specific measure and health outcome or sleep questionnaires.

Conclusions. Studies are highly inconsistent as to how they define breathing measures during sleep studies and OSA itself. Insufficient evidence exists to assess the validity of change in AHI as a surrogate or intermediate measure for long-term health outcomes. Until such validation has been conducted, it cannot be assumed that changes (e.g., improvements) in intermediate or surrogate outcomes are correlated with long-term health outcomes.

RCTs do not provide evidence that CPAP prescription affects long-term, clinically important outcomes. Specifically, with low SoE, RCTs do not demonstrate that CPAP affects all-cause mortality, various CV outcomes, clinically important changes in psychosocial measures, or other clinical events. NRCSs reported associations between CPAP use and reduced risk of all-cause death. NRCS results did not differ from RCTs for other outcomes. We have limited confidence that the summary estimates are close to any true effect.

Comparative studies did not adequately address whether the effect of CPAP varies based on disease severity (e.g., as assessed by AHI), symptoms (e.g., as assessed by sleepiness scales), other patient characteristics, different features or modes or CPAP, or different criteria or definitions of sleep measures or OSA diagnosis.

Additional well-conducted comparative studies are needed to better assess the potential effects of CPAP on long-term outcomes for patients with OSA, whether any particular group of patients may benefit to a greater or lesser degree from CPAP treatment, and whether of changes in AHI (and/or other breathing measures) are valid intermediate or surrogate measures of health outcomes. Associations identified in comparative studies could serve as the basis for more rigorous trials.

Contents

Tables

Figures

Appendixes
Appendix A. Methods
Appendix B. Rejected Studies
Appendix C. Study Descriptions
Appendix D. Results

Evidence Summary

Main Points

- **Contextual information**
  - Numerous measures from sleep studies are used in clinical practice and research settings to characterize abnormal breathing during sleep and to diagnose and grade sleep apnea. These include apnea-hypopnea index (AHI), oxygen desaturation index (ODI), respiratory disturbance index (RDI), respiratory effort related arousals (RERA), and others. Collectively, and individually, these measures do not adequately describe breathing disturbances during sleep. They also do not correlate well with signs and symptoms of obstructive sleep apnea (OSA). Definitions of the measures have evolved over time.
  - Various questionnaires are used to screen for OSA, most commonly the STOP and STOP-BANG. The Epworth Sleepiness Scale (ESS) is the most common tool to measure the severity of subjective sleepiness. The partially validated Functional Outcomes Sleep Questionnaire (FOSQ) is an OSA-specific functional status tool that is commonly used, particularly in research settings, to assess the health status of people diagnosed with OSA.
  - Continuous positive airway pressure (CPAP) is commonly considered the first-line treatment for OSA. Other treatments employed include positional therapy, weight loss, behavioral modifications, oral appliances (e.g., mandibular advancement devices), and, for selected patients, orofacial surgery. Newer treatments include nasal expiratory positive airway pressure (EPAP) devices, oral pressure therapy, and implantable hypoglossal nerve stimulation devices.
    - The U.S. Food and Drug Administration (FDA) database includes 163 CPAP devices used to treat adults with sleep apnea. The large majority of FDA 510(k) Premarket Notification records cite other previously approved CPAP devices to support claims of equivalence. Almost all devices ultimately refer back to four CPAP devices. Notably, though, the available data did not reference clinical studies or unpublished data submitted to FDA that may have supported the device manufacturers' claims.
  - CPAP devices include numerous variable features. These mostly relate to tunable breathing settings, automation, humidity control, monitoring and documentation of use, communication, connectivity and software solutions, and sundry accessories.
  - CPAP treatment has short- and long-term therapeutic goals. The primary short-term goals of CPAP therapy include alleviating symptoms of fatigue and somnolence; improving activities of daily living, health-related quality of life (QoL), and cognitive function; resolving snoring; and reducing the risk of work or motor vehicle accidents. The primary long-term goals of CPAP include reducing the risks of a range of chronic disease outcomes and premature death. Long-term OSA has been associated with cardio- and cerebrovascular disease, hypertension, impaired cognitive function (and Alzheimer disease), type 2 diabetes, anxiety and depression symptoms, and impaired sexual function in both men and women. Under the assumption that CPAP will prevent or alleviate the clinical sequelae of OSA, CPAP has been used as an intervention for these conditions.
- **Definitions of breathing measures used across studies**
  - Studies are highly inconsistent in the criteria used to define sleep study measures (apnea, hypopnea, oxygen desaturation, and respiratory effort), to define OSA, and to categorize

severity of OSA (e.g., based on an AHI threshold). This was the case even among studies stating that definitions are based on the same standard criteria.

- o Most studies (60%) did not fully and explicitly report the definitions of sleep study measures used.
- o Definitions of respiratory disturbance events (e.g., apneas, hypopneas) and criteria to define or categorize severity of OSA are highly inconsistent across studies, despite frequent claims of using standard national or international definitions. However, possible differences in study findings based on heterogeneity of OSA and sleep study measures could not be elucidated. Across studies, the effect of CPAP on health outcomes tended to be statistically similar, precluding attempts to explain differences, and studies did not compare effects of CPAP using different sleep measure definitions or criteria.

- **Effect of CPAP versus no CPAP on long-term health outcomes**
  - o Comparative studies (randomized controlled trials [RCTs] and adjusted nonrandomized comparative studies [NRCSs]) provide, at best, low strength of evidence (SoE) on the effect of CPAP on long-term health outcomes. We did not evaluate sleepiness, other symptoms, or other important outcomes, such as intermediate outcomes (including blood pressure and glucose). We did not evaluate other study designs, including comparisons between patients with OSA who use CPAP and healthy controls or pre-post studies.
  - o 31 studies (14 RCTs and 17 NRCSs) compared CPAP with no CPAP, of which 12 RCTs primarily reported intention-to-treat (ITT) analyses, 6 NRCSs provided "as-prescribed" analyses (of all patients prescribed vs. not prescribed CPAP), and 13 (2 RCTs and 11 NRCSs) provided analyses of CPAP adherent patients versus nonadherent or nonusers . Two RCTs used sham CPAP as the comparator.
  - o Few comparative studies were explicitly powered for long-term health outcomes: 3 RCTs were designed to be powered for composite cardiovascular (CV) outcomes (SAVE, RICCADSA, ISAACC); however, RICCADSA was ultimately underpowered. One RCT (APPLES) was powered for neurocognitive outcomes. No eligible study was explicitly powered for any other outcome. Patients enrolled in RCTs had variable, but generally poor adherence; where reported, 35 to 64 percent of patients used their devices for at least 4 hours per night (on at least 70% of nights). The NRCSs that reported "as-prescribed" analyses mostly did not have data related to CPAP adherence. The NRCSs that reported analyses of CPAP users versus nonusers mostly compared adherent users (>4 hours/night) to those who used their devices less.
  - o Five RCTs did not provide evidence that CPAP reduces the risk of all-cause death (low SoE). As with similar conclusions below, this conclusion does not imply that CPAP has been proven to be ineffective to reduce all-cause death. Rather they are insufficient to provide a definitive conclusion. The RCTs included patients with moderate or severe OSA (AHI ≥15) who were mostly at increased risk of CV death, but without excessive daytime sleepiness. The RCTs were at moderate risk of bias. The studies had relatively short duration of followup (2 to about 5 years) and provided a nonsignificant summary estimate (effect size [ES] 0.89, 95% CI 0.66 to 1.21).
    - Inclusion of 6 adjusted NRCSs (with between 3 and 11 years of followup) provide some evidence that CPAP might be associated with decreased risk of all-cause mortality in longer-term followup among all adults with moderate to severe OSA, with or without sleepiness symptoms. The overall summary ES (combining RCTs and NRCSs) was 0.61 (95% CI 0.49 to 0.76). We have limited confidence in this

conclusion (low SoE) and additional evidence is needed for a more definitive conclusion.

- Both within RCTs and across RCTs and NRCSs, ITT analyses (of RCTs) or "as-prescribed" analyses (of patients prescribed CPAP in NRCSs) yielded similar summary ESs as analyses of adherent versus nonusers.
- Comparative studies (RCTs and NRCSs) provide insufficient evidence to support whether effect sizes may differ in different subgroups of patients, including by severity of OSA (as assessed by AHI or similar measures), severity of symptoms (e.g., as measured by ESS), age, other risk factors or comorbidities.

o Four RCTs do not provide evidence that CPAP reduces the risk of CV mortality (low SoE). The addition of one NRCS does not change the conclusion; the overall summary ES was imprecise (0.79, 95% CI 0.50 to 1.26; low SoE).

o Five RCTs do not provide evidence that CPAP affects the risk of stroke (summary ES 0.99, 95% CI 0.73 to 1.35; low SoE) or of acute myocardial infarction (summary ES 1.05, 95% CI 0.78 to 1.41; low SoE). Both summary ESs are imprecise. One NRCS did not alter the conclusion regarding stroke.

o Four RCTs do not provide evidence that CPAP affects the risk of atrial fibrillation. The summary ES was imprecise (0.94, 95% CI 0.58 to 1.51; low SoE).

o Seven RCTs do not provide evidence that CPAP affects the risk of various composite CV outcomes. Effect sizes were mostly imprecise and ranged from 0.42 to 1.10 across studies, all statistically nonsignificant (low SoE). Six of the RCTs were evaluated as ITT analyses of all users, in which adherence ranged from 35 to 64 percent. Additional evidence from five adjusted NRCSs did not change this conclusion. The range of ESs among the NRCSs was similar to that of the RCTs (0.38 to 0.88; 2 of which were statistically significant). The NRCSs all evaluated adherent CPAP users.

o There is insufficient evidence regarding the effect of CPAP on other CV outcomes, including transient ischemic attacks, angina-related outcomes, coronary artery revascularization, or congestive heart failure outcomes. For each outcome, there were few studies and/or effect estimates were highly imprecise.

o Three RCTs do not provide evidence that CPAP affects the risk of incident type 2 diabetes; the summary ES was imprecise (1.02, 95% CI 0.69 to 1.51; low SoE). An additional adjusted NRCS, though statistically significant, does not alter the conclusion (overall summary ES 0.88, 95% CI 0.57 to 1.35).

o Comparative studies do not provide evidence that CPAP results in clinically meaningful changes in depression symptoms (4 RCTs), anxiety symptoms (4 RCTs), executive cognitive function (4 RCTs and 1 NRCS), or QoL (10 RCTs and 1 NRCS); although the small differences were statistically significant (all low SoE).

o There is insufficient evidence regarding the effect of CPAP on other outcomes, including hypertension, functional status, sexual function, and days of work missed. For each outcome, there were sparse studies, effect estimates were highly imprecise, and/or studies reported highly inconsistent results. We did not evaluate sleepiness or intermediate outcomes.

- Comparative studies do not provide evidence regarding whether CPAP use is more (or less) effective in any specific subgroup. For all health outcomes, studies provide insufficient evidence to suggest whether the effect of CPAP on all-cause mortality varies by baseline AHI, age, body weight, or pulmonary function. Within and across studies,

associations did not clearly or consistently differ based on diagnostic criteria for OSA, definitions of AHI and other sleep study measures, or whether analyses were based on CPAP prescription (e.g., ITT analyses) or CPAP adherence.

- **Effect of CPAP versus other active treatments on long-term health outcomes**
  - Two RCTs provide low SoE that changes in depression and anxiety symptoms do not significantly differ between patients receiving either CPAP or mandibular advancement device (MAD).
  - Two RCTs evaluated QoL and two other RCTs evaluated functional status. The RCTs provide low SoE that changes in QoL and functional status do not significantly differ between patients receiving either CPAP or MAD.
  - The comparative effects on sexual function are insufficient; other long-term health outcomes have not been reported.
  - Three RCTs provide low SoE that changes in functional status scores do not significantly differ between patients prescribed either fixed CPAP or auto-adjusting CPAP. Other long-term health outcomes have not been reported.
- **Adverse events related to CPAP**
  - Eligible comparative studies (RCTs and adjusted NRCSs) that reported long-term outcomes provided insufficient evidence regarding adverse events. Many studies did not report adverse event data.
  - Adverse events reported in the FDA database related to oral and dental health, respiratory system, otolaryngology, odors, allergies and rashes, burns, eye health, aspiration, aerophagia, and miscellaneous other adverse events. Postulated reasons for adverse events mostly related to inadequate humidification, user errors, or device malfunction. No deaths were attributed to CPAP device use.
- **Heterogeneity of treatment effect**
  - Eligible studies rarely evaluated potential differences in effect across subgroups of participants. Both within and across studies, there is insufficient evidence to evaluate any potential heterogeneity of putative treatment effects.
- **Validity of changes in breathing or sleepiness measures as intermediate or surrogate measures for long-term outcomes**
  - No study directly evaluated the validity of changes in breathing or sleepiness measures as intermediate or surrogate measures for long-term health outcomes. No study explicitly evaluated surrogacy or mediation analyses of the measures.
  - Studies did not evaluate whether changes in different breathing or sleepiness measures correlated with long-term health outcomes.
  - Too few studies reported on any given pair of change in breathing or sleepiness measures and long-term health outcomes to allow adequate cross-study evaluation of concordance.
- **Research gaps**
  - Additional well-conducted and analyzed long-term comparative studies are needed to evaluate the effect of CPAP on health outcomes. There has been reluctance to conduct rigorous, adequately-powered, long-term RCTs because CPAP is effective at improving some breathing measures (e.g., AHI), which have been perceived, but not shown, to be intermediate markers of clinical outcomes; CPAP improves some symptoms (e.g., snoring); and associations with health benefits in NRCSs and single group studies have been reported after CPAP treatment. However, clinical equipoise regarding long-term health benefits of CPAP remains. State-of-the-art causally explicit analyses of observational data may provide

some insights, but RCTs will likely be required to provide definitive data on many of the outstanding questions on OSA management and CPAP effectiveness.

- o Studies should be powered to evaluate potential differential effects of CPAP based on such factors as baseline AHI (or other validated surrogate metrics), patient symptoms, and other patient characteristics. All studies need to be sufficiently well-reported to allow readers to fully understand patient eligibility (e.g., how OSA was defined, how sleep and breathing measures were measured, and what thresholds and other criteria were applied to each measure). Individual participant-level meta-analysis of existing studies could generate hypotheses to be tested in subsequent rigorous trials.
- o In a search of the National Institutes of Health's RePORTER, we found no studies that would have been eligible for this review. In ClinicalTrials.gov, we found 6 potentially eligible RCT protocols. The record for the largest listed trial, a mortality study in cardiac patients who had just undergone cardiac bypass surgery or percutaneous coronary procedures, has not been updated since 2015. In 2 of the 5 small studies (<200 patients), the cardiac endpoints were very selective: paroxysmal atrial fibrillation or acute hypertensive pulmonary edema. In the remaining 3, the endpoints assessed were AHI, cognition via the Montreal Cognitive Assessment, or estimated glomerular filtration. Review of published manuscripts will determine the utility of any of these studies.
- o Studies are needed to assess the validity of AHI, other breathing measures, and sleepiness scores as intermediate or surrogate measures for long-term health outcomes. Ideally, studies should assess and compare multiple putative intermediate or surrogate measures. Existing studies or databases may have sufficient data to allow exploratory analyses that can subsequently be further evaluated in rigorous trials.

Background and Purpose

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. OSA is the most common type of sleep apnea. In OSA, the volume of airflow is diminished despite appropriate ventilatory effort. It differs from central sleep apnea, in which there is a reduced drive to breathe, reduced ventilatory capacity, or abnormal ventilatory pattern. Sleep apnea is primarily diagnosed with sleep tests that measure sleep time (and often other sleep measures), respiratory events, and respiratory effort. The diagnosis and severity of OSA are typically assessed with the AHI, a measure of the sum of the numbers of apneas and hypopneas per hour of sleep. AHI is often used as part of both diagnosis (and, thus, study inclusion criteria) and as an intermediate or surrogate measure for health outcomes in studies. However, variations in definitions of apneas, hypopneas, and OSA may result in variations across studies, in which patients are included, and in how treatments are provided. This may hinder the interpretation of the studies.

The most common first-line therapy for OSA is the use of CPAP devices during sleep. The CPAP device directly relieves the obstruction by delivering compressed air (under pressure) to the oropharynx, thereby keeping the airway open.

The Centers for Medicare & Medicaid Services (CMS) nominated the topic to the Agency for Healthcare Research and Quality (AHRQ) in order to evaluate the evidence on improvement of long-term clinical health outcomes with CPAP treatment, as well as the validity of criteria used as surrogate outcomes (e.g., AHI). Per CMS's General Methodological Principles of Study Design, "CMS places greater emphasis on health outcomes actually experienced by patients,

such as quality of life, functional status, duration of disability, morbidity and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses."[1]

This report aims to inform sleep medicine clinicians, sleep technologists, other care providers, sleep apnea researchers, policymakers, and other decision makers about: the definitions of breathing measures used in research and clinical practice; the validity of changes in sleep study and sleepiness measures; and the effect of various treatment modalities on long-term health outcomes based RCTs and adjusted NRCSs (with an emphasis on CPAP). The review provides information to address various background Contextual Questions on diagnostic measures, treatment modalities, and CPAP specifically. It also describes the variability across comparative research studies (both RCTs and adjusted NRCSs) in definitions of breathing measures (e.g., apneas, hypopneas) and criteria to diagnose OSA.

Methods

We employed methods consistent with those outlined in the Agency for Healthcare Research and Quality Evidence-based Practice Center Program Methods Guidance (https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview). Our searches covered studies published through March 22, 2021.

To address the comparative effectiveness of OSA treatments, we conducted a focused systematic review restricted to RCTs and NRCSs adjusted for potential confounders in propensity score analyses or via regression. We included only studies with long-term health outcomes (≥12 months) or measures of QoL, cognitive function, or mental health (≥6 months). We did not include comparisons between patients with OSA who use CPAP and healthy controls, or other indirect comparisons of CPAP versus no CPAP. We did not evaluate sleepiness, other symptoms, or intermediate outcomes (such as AHI, blood pressure, or glucose). To evaluate the validity of surrogate and intermediate measures, we also included other comparative studies with relevant data. Our conclusions are restricted to the findings from the eligible studies.

Results

Definitions of breathing measures: Across 52 eligible studies, reporting and choice of criteria to define sleep study breathing measures and OSA were highly inconsistent. The majority of studies did not explicitly report full criteria or definitions. For example, only 40 percent of studies fully explicitly reported apnea and hypopnea definitions. Examples of inconsistent definitions include apnea defined as between 75 and 100 percent airflow cessation, and hypopnea defined as between 25 and 50 percent decreased airflow. Most studies citing published criteria to define sleep study measures (26/32) cited some version of the American Academy of Sleep Medicine (AASM) criteria. However, there was no discernable consistency in choice of a threshold and citation of a specific AASM version. Of interest was whether the different definitions of sleep measures used had an impact on study findings regarding clinical effect of CPAP. However, as described below, there were no discernable differences across studies, so we could not assess the impact of the variable definitions.

---

[1] Centers for Medicare & Medicaid Services. Appendix A: General Methodological Principles of Study Design. https://www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id101.pdf.

CPAP versus no CPAP: We found 14 RCTs (total n = 7,449) and 17 NRCSs with adjustment for potential confounding (n = 34,462; 25,389 in one study that reported only on all-cause mortality, 9,073 in other NRCSs) that reported long-term health outcomes comparing CPAP and no CPAP in participants with OSA. Most RCTs were of moderate risk of bias (RoB) primarily related to lack of patient and clinician blinding; outcome assessors were usually blinded. Several RCTs were at high RoB due to such issues as high dropout or crossover rates. All NRCSs included multivariable adjustments for potential confounders between groups (CPAP vs. no CPAP) and outcomes; three NRCSs used propensity score matching. The adjusted NRCSs were mostly at high RoB due to concerns about poor descriptions of patient eligibility and/or CPAP treatment, possible selective outcome reporting, and/or comparisons of CPAP use (as opposed to prescription). Among the 14 RCTs, 12 reported ITT analyses. Among the 17 NRCSs, 11 evaluated adherent CPAP users, which may not have been able to fully account for possible biases related to self-selection regarding CPAP use and adherence. However, the ITT analyses generally had to contend with poor adherence (35% to 64%).

Five RCTs did not yield a statistically significant summary effect size for CPAP on all-cause mortality (ES 0.89, 95% CI 0.66 to 1.21). Six NRCSs reporting on all-cause mortality had generally stronger association effect sizes, which were mostly statistically significant. The event rates (percent who died) in the NRCSs tended to be higher than in the RCTs, particularly among two studies of older adults and one study with particularly long-term followup (11 years). The higher event rates in the NRCSs, and their generally larger sample sizes, increased the likelihood of finding statistically significant effect sizes. Combining the RCTs and NRCSs yielded a statistically significant effect size favoring CPAP (summary ES 0.61, 95% CI 0.49 to 0.96). One RCT (SAVE) found a stronger, but still statistically nonsignificant effect in their adherent CPAP user analysis compared with the ITT analysis. This RCT also found no association between degree of CPAP adherence and CV outcomes. A second RCT (ISAACC) found near-identical effects in their adherent CPAP user and ITT analyses. Across studies, the summary effect size of ITT-analyzed RCTs and "as-prescribed" analyzed NRCSs (ES 0.66, 95% CI 0.53 to 0.81) was very similar to the summary effect size of "as-treated" RCT and adherent NRCS analyses (P = 0.41).

Four RCTs yielded an imprecise estimate of the effect of CPAP on CV mortality (0.99, 95% CI 0.64 to 1.53). When combined with a single propensity-score adjusted NRCS, the estimate shifted, but was similarly imprecise (0.79, 95% CI 0.50 to 1.26). Five RCTs provided an imprecise estimate of effect of CPAP on risk of stroke (odds ratio [OR] 0.99, 95% CI 0.73 to 1.35). Five RCTs yielded an imprecise estimate of the effect of CPAP on risk of myocardial infarction (OR 1.05, 95% CI 0.78 to 1.41).

Seven RCTs and five NRCSs reported on multiple unique, but overlapping, composite CV outcomes. The RCTs provided mostly imprecise estimates of effect (with effect sizes ranging from 0.18 to 1.10), but the NRCSs found strong associations (not causality) between CPAP use and reduced risk of (variable) composite CV outcomes (with approximate effect sizes ranging from 0.37 to 0.83). Adherence in the ITT RCTs ranged from 35 to 64 percent. Across studies, the summary effect sizes of RCT ITT analyses (or "as-prescribed" NRCS analyses) and "as-treated" or adherent analyses were not significantly different, although studies that reported both analyses all found nominally stronger effect sizes among adherent CPAP users. The one RCT that reported a within-study analysis reported no association between degree of CPAP adherence and CV outcomes.

Two RCTs reported imprecise effects of CPAP on traffic or home accidents. Three RCTs

found no significant difference in risk of incident diabetes; the addition of an adjusted NRCS does not alter the conclusion. For each analysis, four studies found small, but clinically nonsignificant improvements with CPAP in depression and anxiety symptom scores, and in measures of executive cognitive function. Ten studies also found small, clinically nonsignificant differences in various measures of QoL and functional status. However, the measures have mostly not been validated in adults with OSA and the QoL measures might not qualify as health-related QoL measures. The evidence for other outcomes was sparse, imprecise, and/or inconsistent, including that for transient ischemic attack, angina, coronary artery revascularization, congestive heart failure, atrial fibrillation, incident hypertension, sexual function, and days of work missed.

Subgroup and cross-study comparisons did not elucidate potential modifiers of effect of CPAP. Eligible studies did not evaluate whether different CPAP features or settings had an impact on outcomes. No study reported within-study correlations among outcomes (e.g., effect on AHI and effect on all-cause mortality). For most outcomes, effect sizes were somewhat stronger in CPAP adherent analyses than in intention-to-treat analyses (of all who were prescribed CPAP), but almost universally the differences between analyses were not statistically significant (both within and across studies), and the differences were generally small. No definitive explanation was provided for the lack of difference between analyses with and without nonadherent CPAP users.

CPAP versus other active treatments: Long-term health outcomes comparing CPAP with other active treatments were sparse. Two RCTs found imprecise estimates of differences in depression and anxiety symptom scores among those using CPAP or MAD. Four studies found no differences in effects on QoL or functional status. One RCT, found an imprecise estimate of differences in sexual function between CPAP and MAD. Three RCTs found no significant differences in QoL or functional status between patients using autoCPAP or fixed CPAP suggesting low SoE of no difference in effect.

Intermediate and surrogate measures: No study directly evaluated the validity of *changes* in AHI, ESS, or other breathing or sleepiness measures as predictors of long-term health outcomes. Also, no study explicitly reported analyses of endpoint surrogacy or mediation. We did not assess the validity of single measurements of breathing or sleepiness measures (e.g., measured pretreatment) as predictors of outcomes or treatment effect. None of the health outcomes (e.g., stroke) was reported by a sufficient number of studies that also reported change in breathing or sleepiness measures to allow cross-study evaluation of concordance. In the relatively few instances where correlations could be analyzed across studies, statistically *non*significant correlations were seen between changes in AHI and changes in Short Form (SF)-36 scores and functional status scores; between changes in ODI and changes in functional status scores; and between changes in ESS and depression and anxiety symptom scores, SF-36 scores, and functional status scores. No assessment could be made regarding the relative strength of correlations of AHI versus ESS with health outcomes.

Adverse events: Eligible studies provided insufficient evidence about adverse events due to CPAP use. Many studies did not collect such data. Adverse events reported in the FDA database mostly related to inadequate humidification, user errors, or device malfunction. No deaths were attributed to CPAP use.

Limitations

Our conclusions about the effect of CPAP on health outcomes are restricted to the findings

about specific outcomes from the eligible studies; namely, studies conducting direct comparisons between intervention options (RCTs and adjusted NRCSs only, including analyses of adherent and nonadherent CPAP users) that reported long-term clinical events and certain test scores (such as QoL). We did not evaluate sleepiness, other symptoms, or intermediate outcomes (including AHI, blood pressure, or glucose).

An inherent limitation of the literature is the great variability in, and the often poor descriptions of, how breathing and sleep measures were defined and, thus, exactly how OSA was diagnosed (or the decision to treat with CPAP was made). The lack of clarity about which patients were enrolled in the studies limits our ability to accurately determine which particular patients the studies are most applicable to. Trials provided, at best, low SoE regarding the effect of CPAP on long-term clinically important outcomes. SoE was downgraded primarily due to issues related to study risk of bias and imprecise effect estimates. Few RCTs were powered for health outcomes of interest, including only three powered for composite CV outcomes (one of which ended up being underpowered) and one for neurocognitive outcomes. Subgroup analyses to determine which patients may most (or least) benefit from CPAP were rare and inconclusive. Studies rarely reported analyses evaluating association of adherence with CPAP use and long-term health outcomes. Despite our restriction to multivariable-adjusted NRCSs, NRCSs may be subject to inherent biases related to selection of CPAP or no CPAP (and with adherence). The validity of evaluated QoL, cognitive function, and sexual function measures is unclear for adults with OSA, and the reported QoL measures may not qualify as health related. Lack of blinding may bias findings to favor CPAP for outcomes other than clinical events. Inferences on the validity of breathing or sleepiness measures as surrogate or intermediate endpoints are limited by the fact that none of the clinical event outcomes (e.g., stroke) was reported by a sufficient number of studies that also reported a change in breathing measure, which would allow for cross-study evaluation of concordance.

Implications and Conclusions

Studies are highly inconsistent in how they define breathing measures during sleep studies and OSA itself. It is, thus, difficult to interpret the specific eligibility criteria across studies in terms of severity of OSA (as determined by AHI and similar measures). There is insufficient evidence to determine how study results may vary based on the different breathing measures and OSA criteria. Studies have not directly evaluated the validity of *changes* in breathing or sleepiness measures (e.g., AHI or subjective sleepiness as measured by ESS) as intermediate or surrogate measures of long-term health outcomes. Analysis of study-level correlations between changes in potential intermediate or surrogate measures and health outcomes were based on too few studies to allow conclusions. Thus, there is not adequate evidence to support the contention that changes in AHI or ESS scores translate to improvements in long-term health outcomes.

RCTs of long-term outcomes do not provide evidence that CPAP use affects the risk of all-cause mortality, stroke, myocardial infarction, composite CV outcomes, driving accidents, and incident diabetes (all with low SoE). These conclusions do not imply that CPAP has been proven to be ineffective to reduce these health outcomes. The conclusions were in large part based on imprecise, nonsignificant effect sizes. RCTs also do not provide evidence that CPAP yields clinically meaningful changes in depression and anxiety symptoms, cognitive function, or QoL (all with low SoE). It is unclear whether the failure to find an effect of CPAP treatment on long-term health outcomes is related to a lack of power, insufficient followup duration, or due to an actual lack of effect of CPAP. Notably, only two of the eligible RCTs were adequately powered

for a composite cardiovascular outcome, and none were powered for other cardiovascular or mortality outcomes.

Adjusted NRCSs were generally concordant with the RCTs in terms of the direction of associations with long-term health outcomes. In the case of all-cause mortality, while the RCTs did not find a significant effect, the eligible NRCSs found stronger, statistically significant associations between CPAP use and lower mortality. Thus, the comparative studies overall (RCTs and NRCSs together) may suggest that CPAP use is associated with reduced risk of all-cause mortality. This conclusion is based on only low SoE, which suggests that we have limited confidence in this conclusion and that additional evidence is needed for a more definitive conclusion. Based on the patients who were included in the eligible studies, this conclusion may be most applicable to older adults and may pertain most to longer-term followup. NRCS evidence did not change conclusions about other outcomes.

RCTs and adjusted NRCSs that provide direct comparisons between CPAP use and nonuse in patients with OSA provide unclear conclusions about the effect of CPAP on most long-term health outcomes. Studies largely provide insufficient evidence from sparse studies and/or yield highly imprecise estimates. It is unclear whether the lack of significant differences between adherent user and intention-to-treat analyses is due to a lack of power to indicate a difference in effect or a possible real lack of difference. Notably, the within-study comparisons were all *post hoc* analyses and the between-study comparisons should be considered only hypothesis-generating. If there is, in fact, no difference in effect between adherent and nonadherent CPAP users, this may suggest that either any benefit seen with CPAP use is actually not due to CPAP itself, but to some other behavior or action by CPAP users (for example, increased communication with the sleep clinic) or that even the "low dose" of CPAP achieved by nonadherent users is effective. However, such explanations are just conjecture and would need to be explored critically. There is also concern that many of the NRCSs, particularly those that conducted analyses of adherent participants, may be subject to inherent, incompletely adjusted-for, biases related to self-selection of who chooses (or is chosen for) use of CPAP, who is adherent with using the device, and different reasons for poor adherence (including both remediable reasons, such as behavioral choices, and reasons that are possibly not remediable, such as inability to achieve adequate fit).There is a large degree of inconsistency across studies in how breathing and sleep measures, and thus OSA, are defined. It is, therefore, difficult to interpret the specific eligibility criteria across studies in terms of severity of OSA (as determined by AHI and similar measures). No studies have analyzed whether *change* in AHI, ESS, or other sleep and breathing measures are valid intermediate or surrogate measures for long-term health outcomes.

Additional direct comparative studies are needed before we have a clear understanding of the potential effects of CPAP on long-term outcomes for patients with OSA. Based on RCTs and adjusted NRCSs, there is inadequate evidence to support whether any particular group of patients may benefit to a greater or lesser degree from CPAP treatment to reduce adverse health outcomes.

Two small RCTs provide low SoE that any changes in depression and anxiety symptoms are not affected by the choice of CPAP versus MAD. Four small RCTs provide low SoE that measures of QoL and functional status are not affected by the choice of CPAP versus MAD. Three small RCTs provide low SoE that a measure of functional status is not affected by the choice of autoCPAP versus fixed CPAP.

For the various outcomes in studies comparing CPAP to placebo or to an active control, the low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding that the findings are stable or correct. It would not be unexpected for future evidence to alter these conclusions.

The focus of our review should not be interpreted to imply that we consider other evidence (e.g., comparisons between treated OSA and health controls, pre-post studies), short-term studies, or other outcomes (including sleepiness, other symptoms, and intermediate outcomes) to be unimportant for patients and clinicians.

Introduction

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 by a combination of the two (mixed). The first of these, obstructive sleep apnea-hypopnea syndrome, which is 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 respiratory events during sleep with sleep determined by electroencephalography or surrogate measures such as actigraphy. OSA is distinguished from central sleep apnea by the presence of respiratory effort during episodes of apnea and hypopnea (in central sleep apnea, respiratory effort is lacking). In OSA, the volume of airflow is diminished despite appropriate ventilatory effort. It differs from central sleep apnea, in which there is a reduced drive to breathe, reduced ventilatory capacity, or abnormal ventilatory pattern.

The severity of OSA is typically quantified by the apnea-hypopnea index (AHI), the sum of the number of apneas and hypopneas per hour of sleep measured during a sleep study. AHI is often used as part of both diagnosis (and, thus, study inclusion criteria) and as an intermediate or surrogate measure for health outcomes in studies. Other commonly used sleep study measures include oxygen desaturation index (ODI), respiratory disturbance index (RDI), and respiratory effort related arousals (RERA). 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 guidance on scoring 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, with their first scoring manual in 2007[3] and a major revision 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, at times, used an evidence-based approach (i.e., making recommendations based on systematically reviewed evidence) to guide their selection and revision of criteria, the majority of their recommendations (i.e., scoring rules) are based on the 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 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, RERA from sleep may be allowed as an alternative to desaturation to define a hypopnea. When RERAs are measured instead of desaturation, one measures the RDI in contrast to the AHI.

The variations in definitions of OSA result in subtle variations across studies in which patients are included. Notably, unusual for medical disease diagnostic criteria, diagnosis of OSA in part depends on presence (or absence) of comorbidities. 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] Furthermore, these criteria do not distinguish patients with OSA based on symptomatology. For example, despite clear differences in OSA characteristics among groups, people 1) with frequent respiratory disturbances who do not have symptoms of OSA, such as daytime sleepiness, 2) who have symptoms of OSA, such as daytime sleepiness, but have relatively less frequent respiratory disturbances, and 3) who have the comorbidities listed above but have relatively less frequent respiratory disturbances are all diagnosed and treated as if they have equivalent conditions.

Treatment of Obstructive Sleep Apnea

The most common first-line therapy for OSA is prescription of continuous positive airway pressure (CPAP) devices, or variants of CPAP, including auto-adjusting autoCPAP and bilevel devices (BPAP), for use during sleep. The CPAP device 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). Conceptually, the effectiveness of CPAP use may vary depending on a host of factors, including differences in diagnostic criteria (which will affect which patients are treated), differences in scoring AHI and other sleep study measures (which will affect which patients are treated, how the severity of disease is assessed, and the degree to which treatment is deemed to be working), comorbidities (which may affect both who is treated and the likelihood that treatment will benefit the patient), 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."[8] CMS subsequently covers CPAP for those whose OSA improved as a result of CPAP during this 12-week period. 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.[9]

Non-CPAP treatments that are prescribed in clinical practice 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. In addition, clinicians routinely attempt to treat potential underlying causes of OSA, particularly obesity, with recommendations for weight loss (including bariatric surgery) and exercise. For selected patients, various surgical

interventions are used with the goal of altering the anatomy of the air passages to alleviate postulated obstructive mechanisms. The nonspecific interventions (e.g., weight loss) and specialized interventions (e.g., surgery) are not a focus of this review.

Apnea-Hypopnea Index as a Surrogate or Intermediate Outcome

While AHI and related measures are used in clinical practice to diagnose patients with OSA and evaluate its severity, they are essentially laboratory measures. From the patient's perspective, health outcomes caused by OSA are more important. Health outcomes of concern include cardiovascular (CV) events, health-related quality of life (QoL), changes in cognitive function, and symptoms. OSA commonly results in daytime sleepiness, which can have important sequelae, such as motor vehicle accidents, reduced productivity, and mental health consequences. Because AHI is commonly used to evaluate the mechanical effectiveness of CPAP and because CPAP (when used properly) immediately affects AHI, AHI is the most commonly reported outcome. Clinical health outcomes are more rarely reported.[10, 11] As described in the 2011 Agency for Healthcare Research and Quality (AHRQ) Comparative Effectiveness Review on OSA conducted by the same authors as this review,[11] studies have demonstrated that CPAP improves AHI, as defined in those studies, other surrogate and intermediate measures of OSA severity, and measures of sleepiness. Nevertheless, questions remain about the effectiveness of CPAP to reduce or improve long-term health outcomes (e.g., CV events, stroke, mortality).

A large randomized trial of long-term CPAP use (Sleep Apnea cardioVascular Endpoints [SAVE]) in people with coronary or cerebrovascular disease was recently published.[12-14] Despite improvements in AHI, as defined in the study, it found no improvement of CV, kidney, and weight outcomes. Since CPAP effectively lowers AHI in clinical practice and across studies, the SAVE trial raises questions about whether change in AHI is a valid intermediate or surrogate outcome for these patient-centered health outcomes, as well as whether, and in whom, CPAP may be a clinically effective treatment modality.

Purpose of the Review

This report aims to inform sleep medicine clinicians, sleep technologists, other care providers, sleep apnea researchers, policymakers, and other decision makers about: the definitions of breathing measures used in research and clinical practice; the validity of changes in sleep study and sleepiness measures; the effect of various treatment modalities on long-term health outcomes based RCTs and adjusted NRCSs (with an emphasis on CPAP); and reports of adverse events due to CPAP use. The review also provides information to address various background Contextual Questions on diagnostic measures, treatment modalities, and CPAP specifically, including a summary of the Food and Drug Administration (FDA) 510(k) Premarket Notification records for CPAP devices. It also describes the variability across comparative research studies (both RCTs and adjusted NRCSs) in definitions of breathing measures (e.g., apneas, hypopneas) and criteria to diagnose OSA.

Methods

Review Approach

The Brown Evidence-based Practice Center conducted this systematic review (SR) based on the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews (available at https://effectivehealthcare.ahrq.gov/topics/cer-methods-guide/overview).[15] This SR also reports in accordance with the Preferred Items for Reporting in Systematic Reviews and Meta-Analyses (PRISMA),[16] A Measurement Tool to Assess Systematic Reviews (AMSTAR 2),[17] and any relevant extension statements.

A more detailed version of the SR methodology used can be found in Appendix A. Other appendixes include a list of rejected studies (Appendix B), summary descriptions of the included studies (Appendix C), and results details (Appendix D).

The topic of this report and preliminary Key and Contextual Questions (KQ and CQ) arose through a process involving the nominator (CMS) and AHRQ staff. The protocol was uploaded to the AHRQ's Effective Health Care (EHC) Program website and to the Federal Register for public comment from July 14 to August 14, 2020. The final protocol was posted on the EHC Website at https://www.ahrq.gov/research/findings/ta/index.html on June 16, 2020. On July 17, 2020, PROSPERO published the protocol with registration number CRD42020192725.

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 obstructive sleep apnea (OSA) are currently being marketed in the U.S.? What OSA treatments (experimental or approved) are currently being investigated in ongoing trials for patients as an alternative to continuous positive airway pressure (CPAP) devices?

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 long-term health 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 the 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., apnea-hypopnea index [AHI], apnea index), sleep questionnaires (e.g., Epworth Sleepiness Scale [ESS]), and long-term health outcomes?*

> **KQ 1c**: Do the health 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 long-term health 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 health outcomes across a broader set of studies is primarily addressed in KQ 2.

Analytic Framework

To guide the development of the KQs for the diagnosis and treatment of OSA, we developed an analytic framework (Appendix A Figure A-1) that maps the specific linkages associating the populations and subgroups of interest, the intervention, and outcomes of interest, both intermediate/surrogate and long-term health outcomes. 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 KQs address each aspect of the framework.

Literature Searches

For literature published through 2010, we rescreened for eligibility all studies that were included in existing SRs on OSA diagnosis and treatment conducted for AHRQ.[11, 18-20] For more recent articles, *de novo* literature searches were conducted in Medline (via PubMed), Embase, the Cochrane databases, CINAHL (Cumulated Index to Nursing and Allied Health Literature), ClinicalTrials.gov, and Epistemonikos for primary studies, existing SRs, and published guidelines from January 2010 through March 22, 2021. We also searched the ECRI guidelines Trust[21] for relevant guidelines published in the last 5 years and the U.S. Food and Drug Administration (FDA) medical device databases for summaries of safety and effectiveness that may include study results not published elsewhere.[22]

Literature search strategies included filters to remove nonhuman studies and articles that were not primary studies, SRs, or clinical practice guidelines. The searches included MeSH or Emtree terms, along with free-text words, related to OSA and CPAP. Search strategies were peer reviewed by an independent medical librarian. Appendix A includes the final search strategies.

The reference lists of all relevant existing SRs were screened for additional eligible studies. All prior SRs, including those produced for AHRQ, were used only to identify studies; all eligible studies were re-extracted and re-analyzed anew.

Study Selection

Table 1 presents the major eligibility criteria for each KQ. More detailed criteria are presented in Appendix A. The eligibility criteria were determined in discussion with the Technical Expert Panel and Sponsor, along with public review of the draft protocol.

Briefly, for both KQs, we included studies of adults with OSA, excluding patients with prior strokes, severe heart failure, and in other special populations for whom etiology or management of OSA commonly differs than for most patients. We focused on a specific list of clinical event outcomes and evaluations of mental health conditions, cognitive function, and quality of life (health-related or generic). We did not include sleep and breathing measures (e.g., AHI), symptoms (e.g., sleepiness), or intermediate outcomes (e.g., blood pressure, glucose). We included only long-term treatment and outcomes. For most outcomes, we defined this as a minimum of 1 year. For mental health conditions, cognitive function, quality of life, and sexual function we defined this as a minimum of 6 months. The choices of minimum duration of followup was based on discussions that balanced general understanding of the concept of "long-term" and the expected minimum time required on CPAP treatment to impact outcomes although these intervals might not reflect the durability of any effect.

For KQ 1, we compared CPAP versus no CPAP or other OSA treatments. We restricted evaluation to randomized controlled trials (RCTs) and nonrandomized comparative studies (NRCSs) that included analyses to account for possible confounding, including regression or propensity score analyses. We excluded other observational studies, including pre-post studies and comparisons of CPAP users with healthy controls (without OSA).

For KQ 2, we included only studies that reported an outcome of interest for KQ 1 and *change* in breathing, sleep, or sleepiness measures over a minimum of 6 or 12 months. In particular, we sought studies that provided surrogacy or mediation analyses. We did not assess the validity of single measurements of breathing or sleepiness measures (e.g., measured pretreatment) as predictors of outcomes or treatment effect.

Table 1. Study eligibility criteria, per Key Question

Eligibility Categories	Criteria
Both KQs: Population [A]	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, or with mild cognitive impairment*
Both KQs: Intervention / Comparator [A]	*Exclude studies of surgical interventions for sleep apnea or bariatric surgery*
Both KQs: Outcomes [A]	Major health outcomes • Death • Cardiovascular and cerebrovascular events or incident diagnosis • Motor vehicle accidents • Composite outcomes that include only major clinical event outcomes (e.g., major adverse cardiovascular events defined as including all-cause mortality) Other health outcomes • Other cardiovascular outcomes • Objective measures of cardiovascular severity (categorized, not continuous measures such as intima media thickness) • Incident hypertension (or regression to normotension) • Incident arrhythmias (or resolution of arrhythmias), including clinically significant ventricular arrhythmias and 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, measures of mental health • Cognitive function: clinical diagnosis (e.g., of dementia) or executive function measures • Quality of life and functional status outcome measures, regardless of whether these are health-related or generic measures and regardless of evidence of validation (generally or in adults with OSA) • Sexual function: clinical diagnosis (e.g., diagnosis of erectile dysfunction or anorgasmia) or their resolution, measures of sexual function • Sequelae of sleep deprivation (e.g., trauma, missed work or school) *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 sleep apnea*
Both KQs: Timing (minimum followup duration) [A]	1 year • Death • Incident cardiovascular or cerebrovascular events • Incident hypertension (or reversion to normotension) • Incident diabetes (or reversion to normoglycemia) 6 months • All other outcomes
Both KQs: Setting	Outpatient only (except for sleep laboratory setting for measurement of sleep and breathing measures) *Exclude acute care hospital settings (including perioperative)*
Both KQs: Publication status	*Exclude conference abstracts and other non-peer reviewed reports, except include data reported only in ClinicalTrials.gov*

Eligibility Categories	Criteria
KQ 1 (CPAP treatment): Population [B]	Obstructive sleep apnea (as per study criteria)
KQ 1: Intervention [B]	CPAP for treatment (not diagnosis or staging) of obstructive sleep apnea At least 1 month of prescribed (or planned) treatment *Exclude* • *Intervention designed only to improve CPAP adherence/compliance (i.e., not an intervention of CPAP, per se)* • *Evaluations of accessories only (e.g., nasal cannulas, head straps, humidifiers)* • *Evaluations of CPAP titration methods, per se, including specific parameters or modes (e.g., starting pressures)* • *Evaluations of other features meant to improve comfort or adherence* • *Other non-CPAP interventions (e.g., different times of monitoring, scoring), including noninvasive ventilation*
KQ 1: Comparators [B]	No CPAP Allow assignment to no CPAP, lack of prescription for CPAP, lack of use of CPAP (e.g., due to nonadherence) Other CPAP modality or protocol (e.g., autoCPAP vs. bilevel PAP) Non-CPAP active interventions for obstructive sleep apnea (e.g., mandibular advancement device, positional therapy) *Exclude* • *Bariatric surgery* • *Surgical treatment of obstructive sleep apnea* • *Comparisons with different accessories, titration methods, features to improve comfort or adherence, CPAP protocols (e.g., different times of monitoring, scoring)*
KQ 1: Outcomes [B]	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
KQ 1: Timing (minimum followup duration) [B]	As listed above, for both KQs
KQ 1: Design [B]	Randomized controlled trials Nonrandomized comparative studies Restrict to studies that use modeling or other analytic methods, including randomization, to minimize confounding bias (due to inherent differences between people who receive one or the other intervention *Exclude* • *Case-control studies* • *Pre-post studies (observational comparison of before and after CPAP treatment in a single group of participants)*
KQ 2 (intermediate/surrogate measures): General [B]	For KQ 2, include studies that measured a change in the intermediate/surrogate measure (e.g., AHI) over a period of time and report on outcomes of interest. We included studies that provide formal evaluation of validity of the intermediate / surrogate measure for the health outcome and other studies that reported sufficient data to analyze a potential association between the change in the measure and the health outcome. Studies had to compare two groups (either two interventions or two groups based on participant characteristics, such as gender).
KQ 2: Population [B]	As listed above, for both KQs
KQ 2: Intermediate/surrogate measures [B, C]	Sleep and breathing measures Indices based on apneas or hypopneas (e.g., AHI, RDI) or other respiratory events such as RERAs, oxygen desaturations Note that studies must report a change in the measure over time *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)*

Eligibility Categories	Criteria
KQ 2: Outcomes [B]	As listed above, for both KQs Note that each study must report both change in one or more intermediate / surrogate measures (i.e., sleep and breathing measures) *and* one or more outcomes of interest
KQ 2: Timing (minimum followup duration) [B]	As listed above, for both KQs
KQ 2: Design [B]	Comparative studies informing on validity or person-level associations of change in sleep and breathing measure(s) with outcome(s) Surrogacy or mediation analyses Person-level association between *change* in measure and *change or incident* outcome N ≥30 analyzed for a given association between intermediate/surrogate measure and outcome

Abbreviations: AHI = apnea-hypopnea index, AutoCPAP = auto-adjusting CPAP (also known as APAP), Bilevel PAP = bilevel positive airway pressure (device), CPAP = continuous positive airway pressure (device), KQ = Key Question, RDI = respiratory depression index, RERA = respiratory effort related arousals.

[A] See sections below for eligibility criteria specific to either KQ 1 or 2.
[B] These criteria are in addition to those listed for "Both KQs".
[C] Variables of interest evaluated regarding their association with health outcomes.

Data Extraction

For KQ 1 and 2, we extracted data directly into the Systematic Review Data Repository (SRDR) at https://srdr.ahrq.gov/. We used separate forms for KQ 1 (on treatment effect or association with health outcomes) and KQ 2 (on correlation between intermediate/surrogate measures and health outcomes). Studies reporting data for both KQs were including in both forms.

We extracted information on study characteristics, eligibility criteria, participant characteristics, intervention and comparator details, outcome definitions, and results (including event numbers, effect sizes, and P-values). A series of questions captured various aspects of how OSA diagnoses were made, how sleep studies were conducted, and how sleep and breathing measures were defined. For KQ 1, for each outcome reported by RCTs we determined whether the analysis was intention-to-treat (ITT) or "as-treated" (excluding dropouts or other missing participants). Similarly for NRCSs, we determined whether the analyses were "as-prescribed" (of all participants who were prescribed, or not prescribed, CPAP, conceptually analogous to ITT) or restricted to either adherent or all CPAP users.

Study- and outcome-level risk of bias assessment was conducted during data extraction within SRDR. All studies were reported in English language articles.

Risk of Bias Assessment

We evaluated each study for risk of bias and methodological quality. Because we included a variety of study designs, we incorporated items from three different commonly used tools and tailored the set of items for each study design.

For RCTs, we used all the items from Cochrane Risk of Bias 2.0 tool,[23] 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 were assessed for each outcome.

For NRCSs, we included assessments of specific elements from the Risk Of Bias In Non-randomised Studies – of Interventions (ROBINS-I) tool,[24] in particular related to selection bias (comparability of groups). The questions were assessed for each outcome (e.g., whether each outcome was adjusted for potential confounders).

Since we downgraded RCTs for issues related to randomization and blinding, to ensure consistency across all studies, we deemed that NRCSs could at best be at moderate risk of bias. We further downgraded the risk of bias of NRCSs that did not use propensity score (or equivalent) analyses to control for inherent differences among participants who used (or were prescribed) CPAP and those who did (or were) not. For studies that reported propensity score analyses, we evaluated the type of analysis used (propensity score matching, weighting, stratification, or covariate adjustment), the proportion of the total sample included in a matched analysis, whether non-baseline variables were included, and the comparability of baseline characteristics between groups after propensity score adjustment.[25] Inadequate propensity score analyses were considered to increase the risk of bias.

Data Synthesis and Analysis

Within the main report, data are summarized either in succinct tables that focus on outcome, interventions, and comparative results or in forest plots or succinct summary tables (for most topics). Appendix D includes the more detailed, study-level results for each topic. Appendix C contains detailed tables that describe study and participant characteristics, intervention (and comparator) details, outcomes (and definitions), and arm- and comparison-level results. Appendix C also includes tables providing study-level risk of bias assessments.

For KQ 1 (CPAP treatment effect), where feasible and appropriate, we conducted random effects model pairwise meta-analyses to compare interventions, using the restricted maximum likelihood approach. Details are in Appendix A. Meta-analyzed studies needed to be similar in population, interventions being compared, and reported outcomes.

We first meta-analyzed and summarized findings and conclusions from RCTs only. As exploratory analyses, we secondarily added the adjusted NRCSs to the meta-analyses and summary findings and conclusions. This was done under the assumption that adjusted NRCSs are likely to provide similar findings as RCTs,[26] but also allows for an assumption that NRCS evidence is inherently flawed and should be omitted from synthesis. It was also understood that NRCS data may provide associations that can be tested in future rigorous trials. To evaluate the role of the NRCSs in the combined meta-analysis summary estimates, we performed a sensitivity analysis on meta-analyses that included both RCT and NRCS evidence that allowed us to visualize how the meta-analytic results change as the NRCS evidence is given increasing weight from none (only RCT evidence is considered) to equal weight with the RCT evidence (standard meta-analysis). Each graph was assessed qualitatively to make a judgment about the relative findings of RCTs and NRCSs. Appendix D displays all such graphs. For each meta-analysis, we report the I^2 value, which estimates the percentage of the overall variability that is attributable to statistical heterogeneity.

For each RCT comparing CPAP with no CPAP, we calculated the number-needed-to-treat (NNT) or to harm (NNH) based on the reported numbers of events for each categorical (binary) outcome (no RCT reported NNT or NNH). For outcomes with meta-analyses of RCTs, we also estimated a summary NTT based on the summary effect size and the separately meta-analyzed control rate (event rate in the no CPAP arm).

For KQ 2, we used a metaregression approach to assess evidence on surrogacy and mediation.[27] We included RCTs and observational studies with adjustment for confounding that compared two or more groups (either interventions or subgroups), reported changes in breathing measures (e.g., AHI) or both pre- and post-intervention values, and at least one outcome of interest. We evaluated mean effects between arms by assessing the net difference (difference-in-difference) from baseline until the latest reported timepoint or the odds ratio (OR). Continuous measures that were derived from studies that used different scales were first converted to standardized effect sizes. For studies with three or more relevant groups, we first meta-analyzed effects across similar groups (e.g., the two CPAP groups) to allow a single comparison value per study. We used the same method to assess the correlation between sleepiness measures and health outcomes.

For each analysis, we graphically present the mean effect of AHI (or other breathing measure, or sleepiness measure), on the Y axis, against the mean effect of the outcome, on the X axis such that the origin (where X and Y axes meet) is in the middle and the upper left (quadrant A, northwest) and lower right (quadrant D, southeast) quadrants correspond to improvements in intermediate measure (AHI, etc.) correlating with improvements in health outcome (or vice versa). The X-axis was flipped, as needed, to maintain this orientation. For outcomes with at least three studies, we assessed the study-level concordance (the effect of an intervention on the outcome and its effect on the breathing measure were in the same direction in terms of favorability) between outcomes and breathing measures. We quantitatively assessed the degree of concordance using the non-parametric Spearman correlation coefficient (ρ) and its associated P-value.[28] The strength of the (absolute value of the) correlations were categorized as follows: 0.00 to 0.19 "very weak", 0.20 to 0.39 "weak", 0.40 to 0.59 "moderate", 0.60 to 0.79 "strong", 0.80 to 1.0 "very strong". We also evaluated the statistical significance of whether studies were more likely to demonstrate concordance or discordance between intermediate and outcome measure effects, using the nonparametric exact binomial test.

We examined within-study evaluations of heterogeneity of treatment effect (e.g., subgroup differences, regressions with interaction terms). In particular, we sought to evaluate the following factors as potential mediators of treatment effect: body weight, obesity, neck circumference, etc.; weight change (loss or gain); prior cardiovascular disease, cerebrovascular disease, or other major clinical disease/condition; sex/gender; race/ethnicity; "severity" of OSA (as defined by study; e.g., pretreatment AHI); new or prior OSA diagnosis; treatment-naïve versus failed prior treatment; first versus second or more use of CPAP; treatment (CPAP) adherence/compliance; and treatment (CPAP) discontinuation. When feasible, we also looked across studies for possible evidence of heterogeneity of treatment effect, but restricted these analyses to study-level differences (i.e., we did not evaluate factors subject to ecological fallacy, such as mean ages or weights across studies).

We also qualitatively assessed clinical heterogeneity across studies through an analysis of each study's definitions of the components of AHI and oxygen desaturation index (ODI), the set of criteria used to define AHI, and the AHI (or other) threshold used to define OSA. Studies were ranked (sorted) by the "strictness" of their criteria (e.g., a requirement of 4% drop in oxygen saturation was considered stricter than a 3% drop in saturation). Variations in treatment effects across studies were assessed qualitatively considering the strictness of breathing measure requirements. Statistical (quantitative) assessment across studies was not feasible due to the small number of studies reporting any given outcome and the large number and variability of factors involved in assessing strictness across studies.

Of note, we included only analyses reported in published studies. We did not re-analyze data from other studies (beyond simple calculations, such as OR). We also did not request any participant-level data or conduct participant-level meta-analyses.

Determining Outcome-Level Conclusions and Grading the Strength of the Body of Evidence

We evaluated the strength of evidence (SoE) addressing each major analysis for each KQ. We graded the SoE as per the AHRQ Methods Guide.[15, 29] For each SoE assessment for KQ 1 (CPAP treatment effect), we considered the number of studies, the study limitations (i.e., risk of bias), the directness of the evidence to the KQs, the consistency of study results, the precision of any estimates of effect, and other limitations (particularly sparseness of evidence). We assessed two levels of precision for different aspects of assessing study findings and SoE. When (summary) estimates were "imprecise" (the 95% confidence interval [CI] extended beyond both 0.80 and 1.25), overall SoE was downgraded for precision. When (summary) estimates were "highly imprecise" (95% CI extended beyond both 0.50 and 2.0), we deemed there to be insufficient evidence to allow conclusions regarding SoE.

For each outcome, where there was sufficient (i.e., not insufficient) evidence, we summarized the overall conclusion based on a combination of the direction of effect (or association), the statistical significance, and the magnitude of the effect in relation to a minimal clinically important difference (MCID). Preferable, but not required, the studies were powered *a priori* for the MCID for each clinical metric. Where feasible, we based MCIDs on at least one study reporting a relevant MCID for a population as close as we could find to OSA. We cite each MCID we use. However, we did not use, nor did we find, MCIDs for any clinical event outcomes. For these outcomes, we effectively used an effect size of 1.0 as the MCID threshold. However, we did not depend strictly on statistical significance to make conclusions, but used judgment, in concert with SoE assessment, to determine the ultimate conclusions. Based on the language inherent in the AHRQ Methods Guide for grading SoE, in instances where the effect size estimate and confidence interval did not suggest an effect (i.e., when the effect size was close to 1.0 and statistically nonsignificant) we concluded that there was no evidence of (an) effect or, for bulleted Main Points that the intervention does not affect risk of the outcome.

We assessed the factors that determine SoE separately based on RCT data alone, NRCS data alone, and all eligible studies. We determined a formal SoE rating for RCTs alone because these types of studies can provide information on causality. From eligible NRCS, we could determine only strength of association; although some types of NRCSs can emulate an RCT (e.g., properly designed and implemented propensity score studies) if all of the relevant variables are identified and if there is sufficient information for these variables. For a secondary conclusion, we then combined evidence from RCTs and NRCSs, both by meta-analysis and evaluation of SoE, since these analyses may provide some exploratory and suggestive information that may merit further investigation. We did not determine a SoE rating for NRCSs alone (since a SoE assessment ignoring RCTs would be of questionable value).

For KQ 2 (surrogate/intermediate outcome assessment), we considered whether there was adequate evidence from within-study analyses of correlation between surrogate/intermediate outcomes and health outcomes, the number of studies, the consistency of study findings, and potential limitations stemming from study design and/or analytical approaches.

Based on these assessments, we assigned a SoE rating as being either high, moderate, low, or insufficient to estimate an effect. Similarly, for both KQs, if studies reported highly inconsistent

findings across studies or if only one study reported an outcome, the SoE was deemed to be insufficient.

Applicability

For KQ 1 (effect of CPAP), we assessed the applicability of the included studies to people eligible for Medicare coverage (based on age or co-existing disability). Applicability was 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, race/ethnicity, and geographic location of the study.

Contextual Questions

Contextual Question 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?

The apnea-hypopnea index (AHI) is the most common metric used to summarize the findings of a sleep study. As such, it is used to operationalize the diagnosis of sleep apnea, examine where individuals are on the spectrum of their respiratory condition, and assess treatment response in clinical and research settings. Questions have been raised about whether AHI and related metrics are an adequate aggregation of the findings of a sleep study, or whether alternative or additional polysomnography, other sleep study, or biometric measures should be used for diagnosis, severity assessment, prognosis and treatment monitoring.[30, 31]

We present other metrics and the evolution of standard definitions for AHI (the most well-known metric) in the following paragraphs that organize critiques of over-reliance on AHI as the key metric.

Narrow Scope of Apnea-Hypopnea Index

AHI, and related indices like the apnea index (AI) and oxygen desaturation index (ODI) incompletely measure the severity of obstructive sleep apnea (OSA) and may not truly represent the primary determinants of adverse outcomes. In particular, AHI does not include the severity of arousals, microarousals, autonomic imbalance, sleep stage fragmentation, or hypoxemia. Interruptions in sleep may be more directly associated than AHI with the primary symptoms of OSA such as excessive daytime sleepiness and neurocognitive effects.[32-35] Other available measures such a respiratory effort related arousals (RERA), the arousal index (a broader measure of sleep disturbances), and sleep stage stability may also be used in clinical practice to evaluate sleep disruption and severity of OSA.[36] However, measures of the effect of apnea and hypopnea, such as arousal or hypoxemia, may more directly result in symptoms and clinical sequelae. For example, arousal intensity (measured by electroencephalography) and heart rate response to arousal may be better measures of the sleep disruptive effects of OSA.[34, 35] Similarly, the severity of hypoxemia as measured by the hypoxic burden has been reported to be a predictor of cardiovascular mortality.[37]

Measures of Intensity of Exposure Versus Cumulative Exposure to Reduced Inhaled Air During a Night of Sleep

The hallmark of sleep apnea (obstructive, central, or mixed) is diminished volume of inhaled air during sleep, which can result in oxygen desaturation, sympathetic system activation, and

microarousals (from sleep).[38] It has been postulated that an individual's total exposure to "diminished inhaled air during sleep" contributes to sequelae, much like total exposure to cigarette smoke or ionizing radiation contribute to increased risk of malignancies.

Commonly used metrics, including AHI, AI, ODI, respiratory disturbance index (RDI), RERA, hypopnea index, and others, quantify *number of respiratory events per unit time.* As such, they may provide incomplete information if cumulative exposure to reduced inhaled air—which incorporates sleep duration—is important. These metrics of cumulative exposure include total number of events, total duration of events by type, and time spent with oxygen saturation below a specific threshold.

It is unclear that rate metrics should be preferred over metrics that directly assess cumulative exposure to reduced air intake during a night of sleep. However, measures of cumulative exposure are uncommonly used in sleep medicine. In contrast to rates (measures that include a unit of time), direct measures of cumulative exposure during a night of sleep capture the total number of events experienced. As an analogy, the cumulative exposure to smoking, in terms of pack-years is preferred over the rate of cigarettes smoked per year. For example, consider two people who have the same AHI of 10 events per hour of sleep but who sleep for 7 and 5 hours, respectively, every night. On the basis of AHI, they are observationally equivalent. However, the first is exposed to 2 more hours of diminished inhaled air during sleep. The total number of events per night (70 versus 50), possibly stratified by type of event, is better than AHI at distinguishing levels of exposure to diminished inhaled air during sleep. The analogous observation holds for other indices, e.g., the ODI measures the rate of oxygen desaturation events, but not their total number.

Another direct metric is the total duration of respiratory events. This measure may be informative since the mean duration of respiratory events varies both among patients and for each patient over time. A 10-second respiratory event is unlikely to be physiologically equivalent to an event that lasts 2 minutes. The total duration of respiratory events during sleep would distinguish patients who have the same AHI, same sleep duration (and, thus, same mean number of events), but who differ in the average total duration of their respiratory events. A potentially important implication of this issue of measurement type in reference to AHI, women tend to have a lower AHI, with shorter duration apneas and hypopneas, than men. Thus, the prevalence and severity of OSA, as defined by AHI, is lower in women than in men, while the sequelae of disordered breathing may be the same, if not worse among women for comparable degrees of severity found in men.[39] There has been the finding that women's apneas and hypopneas tend to cluster more during rapid eye movement (REM) sleep than non-REM sleep compared to men, resulting in a lower overall (nightly) AHI among women.[40]

Several other metrics that have been proposed elaborate on the theme of better capturing the total exposure to the respiratory events that characterize the syndrome, using weighting schemes that aim to measure the physiological impact of the respiratory events. For example, one can consider the total duration of respiratory events weighted by the oxygen desaturation that accompanies them,[41] hypoxic burden[37] (oxygen desaturation area under the curve during sleep), the morphology of the oxygen desaturation events,[42] and other analyses that try to infer physiologic impact by quantifying sympathetic system activation (e.g., via time and frequency domain analysis of heart rate). A simple weighted measure is the total duration spent with oxygen desaturation below 90 percent (or another cutoff). These weighting schemes aim to measure the physiological impact of the respiratory events.

Empirical research has demonstrated that patients who have similar AHI can differ greatly in these other metrics.[41, 42] Generally, AHI is not well-correlated with symptoms and signs of OSA.[43] However, it is an open question whether alternative metrics are more strongly correlated with symptoms and signs.

Other metrics have been proposed for clinical assessment that combine physiologic and other measures to assess subjective and objective disease burden. These metrics integrate anatomic (e.g., redundant pharyngeal mucosa, tonsil size), anthropometric (e.g., body mass index [BMI], neck circumference), subjective (e.g., daytime sleepiness reflected by the Epworth Sleepiness Scale [ESS]), and physiologic measurements (e.g., AHI, lowest oxygen desaturation).[44, 45] We do not describe them further because they are not purely physiologic metrics.

Changes in the Definitions of Key Respiratory Events

The definition of key events, such as hypopneas, has evolved over time in in both research and clinical settings. Changes in event definitions can influence the meaning and interpretation of all metrics, including AHI, AI, RDI, RERA, and the metrics of cumulative exposure discussed in the previous section. We describe in some detail the evolution of definitions of hypopneas, because AHI is the most commonly used metric.

The AHI includes both apneas (complete cessations of airflow) and hypopneas (partial reductions of airflow) per hour of sleep that are deemed to be physiologically consequential. Generally, it is agreed that 10 seconds is a reasonable minimum duration for an event, in that within 10 seconds of no breathing an unambiguous drop in oxygen saturation can manifest. However, various versions of AHI use different definitions for what constitutes a physiologically consequential obstruction, for apneas, hypopneas, and oxygen desaturation. This has resulted in variability in the application of "standard" definitions in research settings (see Results, *Definitions of Breathing Measures Across Studies*) and in clinical practice across laboratories (e.g., Manser et al. 2002).[46]

In 1999, an American Academy of Sleep Medicine (AASM) Task Force recommended counting only hypopneas accompanied by a 3 percent oxygen desaturation or a RERA.[2] In 2007, the updated AASM guideline allowed for two definitions for hypopnea, namely at least a 30 percent reduction in airflow accompanied by an oxygen desaturation of 4 percent or at least a 50 percent reduction in airflow accompanied by an oxygen desaturation of 3 percent. The subsequent 2012 AASM update defined hypopneas as drops of at least 30 percent in airflow associated with a range of physiologic responses, including arousal in the electroencephalogram, surrogates of arousals, and measures of sympathetic system activation, such as change in heart rate.[4] It thus allows for two types of definitions: (1) a drop of 30 percent in airflow associated with at least a 3 percent desaturation or evidence of an event-related arousal (or both), or (2) a drop of 30 percent in airflow associated with at least a 4 percent desaturation. The most recent 2020 AASM update dropped this second definition of hypopnea.[5] Based on what was reported in the AASM guidelines and in commentaries we found regarding the guidelines, the changes in the definitions of events (e.g., whether to use a 3 or 4 percent oxygen desaturation threshold) were made based on evolving evidence about the implications or physiologic meaning of apneas and, in particular, hypopneas. However, we could not find explicit reference to supporting evidence, particularly of studies evaluating different breathing measure definitions in reference to health outcomes.

The newer definitions count more events as hypopneas when calculating the AHI, primarily related to decreasing the required oxygen desaturation from 4 to 3 percent. Thus, in a reanalysis

of 6441 polysomnography recordings, with alternative definitions for hypopneas, the mean AHI was 5.4 (95% confidence interval [CI] 1.8 to 13.4) with the 2007 AASM definition (at least 30 percent airflow reduction and 4 percent desaturation) versus 13.4 (95% CI 6.8 to 24.1) with the 2012/2018 AASM definition (at least 30% airflow reduction plus ≥3 percent desaturation and/or arousal).[47] Nomograms have been developed to translate measurements from one definition to the other, accounting for uncertainty. Statistical analyses demonstrate divergence of alternative AHI definitions for smaller values and convergence for higher values. Thus, the same person may be classified differently using the same numerical cutoffs for different AHI definitions. In the aforementioned analysis, 48 percent of patients had an AHI of less than 5 events per hour with the 2007 definitions, versus 17 percent with 2012/2018 definitions.

A prudent interpretation of these critiques encourages research on alternative metrics to better capture patients' respiratory processes during sleep and suggests that we should refrain from relying only on AHI to classify moderately severe disease. Understanding AHI measurements and how to interpret them is important to contextualize the large volume of research that has been conducted thus far.

Contextual Question 2

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

Different questionnaires related to OSA have been developed for distinct goals: (1) to screen for the condition, (2) to assess health status (primarily symptoms), and (3) to measure impact on daily living and health-related quality of life (QoL), including generic and specific instruments. We briefly describe one tool per goal, and comment on their analytic validity (in terms of formal psychometric analyses), clinical validity (sensitivity and specificity to detect a definition of OSA), and clinical utility (whether using versus not using it when managing patients improves clinical and patient-centered outcomes).

Screening for OSA

A commonly used tool is STOP (snoring, tired, observed, pressure) and its variant, STOP-BANG (STOP plus BMI, age, neck size, gender), which were introduced in 2008 in the context of preoperative evaluations.[48] STOP includes questions on four subjective items (snoring, tiredness during daytime, observed apnea, and high blood pressure). STOP-BANG adds four anthropometric and demographic items (STOP with BMI >35 kg/m^2, age >50 years, neck circumference >40 cm, and male gender). STOP-BANG, in particular, has been used extensively in preoperative assessment,[49] sleep labs,[50] the general population,[51] and in patients with chronic diseases to screen for OSA.[52] It has been translated into numerous languages.

In the simplest scoring model for STOP-BANG, the eight items are scored 1 or 0, for affirmative and negative answers, respectively. Higher scores indicate higher likelihood of OSA. Typically, a score of 0 to 2 indicates low likelihood of OSA; a score of 5 to 8 high likelihood of OSA, and a score of 3 or 4 intermediate risk. Because isolated items in STOP-BANG do not have the same predictive potential,[53] various scoring algorithms have been proposed that do not count all items equally.[49] Thus, the predictive ability of STOP-BANG to detect OSA depends on (1) the definition of OSA, (2) the scoring algorithm used for STOP-BANG, and (3) the cutoff in the overall score used. STOP-BANG was developed using rigorous psychometric methods[48]

(including reliability analysis and factor analysis) and has substantial analytic validity. Its clinical validity, in terms of sensitivity and specificity to detect various polysomnography-based definitions of OSA, has been examined in various populations, and varies based on the specifics. However, we did not identify empirical evidence on its clinical utility regarding effects on long-term health outcomes in comparative studies (i.e., whether using versus not using the symptom score improves patient outcomes).

Symptom Severity

The ESS is the most commonly used tool to measure subjective sleepiness (sleep propensity).[54, 55] It comprises eight questions about the likelihood of dozing off or falling asleep (with four options, ranging from 0 = would never doze to 3 = high chance of dozing) in different situations, some very soporific, others less so (e.g., lying down to rest in the afternoon, sitting inactive in a public space, in a car while stopped in traffic). Higher scores correspond to increasing sleep propensity. Typically, a score of at least 10 (out of a maximum of 24) is consistent with clinically important sleepiness.[55]

Because it measures subjective sleepiness, the ESS has been used to predict the presence of OSA (using various definitions). In psychometric analyses, the ESS has good reliability and internal consistency.[56-59] The ESS exhibits moderate or weaker correlations with other constructs (e.g., maintenance of wakefulness test, multiple sleep latency test, AHI and other measurements).[57] In all, we consider it analytically valid.[40] However, ESS has limited clinical validity in detecting various definitions of OSA.[11, 60] We could not identify empirical evidence on its clinical utility, whether using versus not using it improves patient outcomes.

Impact on Daily Living and Quality of Life

Various generic (e.g., 36/12-item Short Form medical survey [SF-36/12],[61] EuroQol-5D) and specialized (e.g., Calgary Sleep Apnea Quality of Life Index [SAQLI],[62] Functional Outcomes Sleep Questionnaire [FOSQ][63]) health status tools, health indices, and health-related QoL tools have been used in sleep apnea assessments.[64] Reviewing generic QoL instruments and health indices is outside the scope of this report. As will be evident in the SR, below, most RCTs of CPAP have evaluated generic QoL tools (SF-36/12, EuroQol-5D).

For assessment of the effect of a treatment on a health condition, it is important to consider the more specific concept of health-related QoL. There is a wide range of definitions of health-related QoL.[65-67] We briefly comment on FOSQ, a disease-specific health status instrument that measures impact of sleepiness on activities of daily living.[63]

FOSQ is a self-administered, disease-specific health status instrument that measures impact of sleepiness on activities of daily living.[63] It evaluates five dimensions of functional status: (1) activity level, (2) vigilance, (3) intimacy and sexual relationships, (4) general productivity, and (5) social outcomes. The long form of FOSQ includes 30 questions, for which responders choose from a 4- or 6-point scale, depending on the question. The possible scores range from 0 to 100, with higher values indicating more adverse impact on daily living. Shorter forms of the instrument (e.g., FOSQ-10, scores ranging from 5 to 20) have also been developed.

In psychometric analyses, the FOSQ has good reliability, construct validity, concurrent validity, internal consistency, and discriminant validity. A commonly used shortened version (FOSQ-10) has been validated against other QoL measures in 149 CPAP-treated patients with OSA (who were selected for having an AHI ≥15, but who had a mean BMI of 64 [SD 29]);[68] although we found no studies that have validated it against established OSA-related health-

related QoL outcomes. The originally constructed, but less commonly used, 30-question FOSQ has been validated in 135 adults with OSA (AHI ≥15, mean 45.6 [SD 26.3]).[69] It was found to be valid across several validity domains, including responsiveness to CPAP use and adherence. There is also some information that FOSQ has limited clinical validity in detecting various definitions of OSA (i.e., screening).[11, 60] We could not identify empirical evidence on its clinical utility, whether using versus not using it improves patient outcomes.

Contextual Question 3

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

Positive airway pressure therapy, namely, CPAP, is commonly considered to be a first-line treatment for OSA. Positional therapy, weight loss, and behavioral modifications are (often ancillary) interventions that do not require particular devices. Several oral appliances (e.g., mandibular advancement devices) have been cleared or approved by the FDA to be legally marketed in the U.S. They may be used in patients with mild to moderate OSA and in those who do not tolerate CPAP.[11, 70, 71] They reportedly help reposition the lower jaw to alleviate upper airway obstruction during sleep. For some patients with specific anatomic variations, surgery may be indicated,[72] including uvulopalotopharyngoplasty,[73] maxillomandibular advancement,[74] and adenotonsillectomy (particularly in children). These surgeries reportedly correct structural airflow obstructions with the intent of increasing the volume of inhaled air during sleep.

Newer treatments include nasal expiratory positive airway pressure (EPAP) devices,[75] oral pressure therapy,[76] and implantable hypoglossal nerve stimulation devices.[77, 78] The FDA has cleared or approved several of these devices for use in the U.S. The EPAP device is a single-use mechanical device that has mechanical valves with very low inspiratory resistance but high expiratory resistance. One EPAP valve is adhered to each nostril. The rationale for the device is that high expiratory resistance results in positive pressure throughout exhalation. This splints open the upper airway during exhalation, purportedly making it more resistant to collapse on subsequent inspiration.[75, 79]

Oral pressure therapy applies a vacuum to the mouth, which is purported to stabilize upper airway structures in patients with OSA. The oral pressure device uses a mouthpiece with a lip seal and a vacuum pump. The device creates a partial vacuum (negative pressure) gradient that is purported to draw the soft palate anteriorly into a stable contact with the tongue, widening the nasopharyngeal airways for nasal breathing during sleep. The negative pressure is isolated in the oral cavity, because of the stable seal between the soft palate and the tongue and does not extend to the nasopharynx.[76]

Hypoglossal nerve stimulation by means of implantable devices is being examined for patients with moderate to severe OSA.[77, 78, 80] The premise is that upper-airway stimulation, via unilateral stimulation of the hypoglossal nerve, increases the patency of the upper airways, which activates the genioglossus muscle, moving the tongue anteriorly and opening up the nasopharyngeal passages. Such devices are intended to synchronize stimulation with the ventilatory effort during sleep.

CPAP Device Premarket Notification Clearances

Numerous CPAP devices are available on the market. According to Section 510(k) of the Food, Drug and Cosmetic Act, device manufacturers must provide a Premarket Notification to the FDA of their intent to market a medical device at least 90 days in advance. This allows the FDA to determine whether the device is equivalent to an already-cleared device. Specifically, "medical device manufacturers are required to submit a premarket notification if they intend to introduce a device into commercial distribution for the first time or reintroduce a device that will be significantly changed or modified to the extent that its safety or effectiveness could be affected. Such change or modification could relate to the design, material, chemical composition, energy source, manufacturing process, or intended use."[81] Upon premarket notification by manufacturers, the FDA provides devices "clearance" upon review of the 510(k) submission. After rigorous review of the devices, the FDA determines whether to grant "approval" for the device to be legally marketed in the U.S. In addition, devices may be granted a "de novo" categorization for novel medical devices for which there is no legally marketed predicate (previously approved) device.

We examined the Premarket Notification [Section 510(k)] records on CPAP devices to determine whether new versions of CPAP devices were cleared (and/or approved or granted de novo categorization) on the basis of new clinical research or on the basis of substantial equivalence to an already-cleared device.

We searched the Devices@FDA database[82] without date restrictions (last search on March 1, 2020) using the terms "sleep apnea," "OSA," "continuous positive airway pressure," and "CPAP;" the names of major manufacturers ("Phillips-Respironics", "ResMed", "Simons", "Fisher & Paykel"); and product codes for non-continuous ventilator devices ("BZD"); and additional codes listed on the returned records ("LRK", "MNQ"). We identified 812 unique Premarket Notification records, of which 163 referred to CPAP devices used for treating sleep apnea in adults. Of these, 143 records were approved in 1996 or later and had downloadable documents; we analyze these. The remaining 20 records without downloadable documents were approved/cleared between 1976 and 1994.

For the 163 510(k) Premarket Notification records on CPAP devices, we extracted an identification number, the device name and manufacturer, and the approval date (see Appendix Table D-4). From the 143 that had downloadable documents we also extracted whether a clinical study was mentioned and whether they cited another approval document to claim equivalence with an approved device. Upon examination of the publicly available data in FDA database, it was apparent that we did not have access to the complete Premarket Notification records. Publicly available data via Devices@FDA database included only: basic device and manufacturer information provided by the submitter; predicate device(s); reason for submission; submitted information about substantial equivalence with previously cleared devices, intended use, and device description; and the clearance letter written by the FDA. The available records do not include explicit references to any clinical studies that may have supported the claims. There was also no explicit reporting of, or reference to, any new (unpublished) clinical data. However, such information might exist in other inaccessible parts of the Premarket Notification records.

The large majority of Premarket Notification records for CPAP devices cite other previously approved/cleared device records to support claims equivalence. We evaluated the more-recent 143 CPAP device records with available data. The 143 CPAP device records also cited 78 records of CPAP accessories, which are included in the analysis.

We found a total of 266 citation relationships. Figure 1 graphs the citation relationships reported in the 143 device records. The figure demonstrates 13 groups of records that are connected with citation relationships. The largest relationship group includes 169 device records. The other relationship groups include many fewer records: 3 groups have between 9 and 12 records, and nine groups have between 1 and 4 records.

Among the 143 CPAP device records, four (large red dots in Figure 1) did not cite other records but they were each cited by at least one other record. These include: Apex Medical XT1 CPAP model 9S-005 (document number K070609, approved in 2007), Apex Medical CPAP RT 21XX (document number K022650, approved in 2004), Orion Nasal CPAP System (Bird Products Corp., document number K020730, approved in 2002), and Sullivan Autoset Portable II Nasal CPAP System (ResMed, Ltd., document number K970771, approved in 1997).

Half of the device records (n = 76, empty circles) both cited and were cited by other records. Many device records (n = 61, blue circles) only cited other device records. Two device records neither cited nor were cited by other available device records (gray circles); they were both approved based on predicate devices marketed prior to 1976.

Figure 1. Citations between records to support equivalence claims
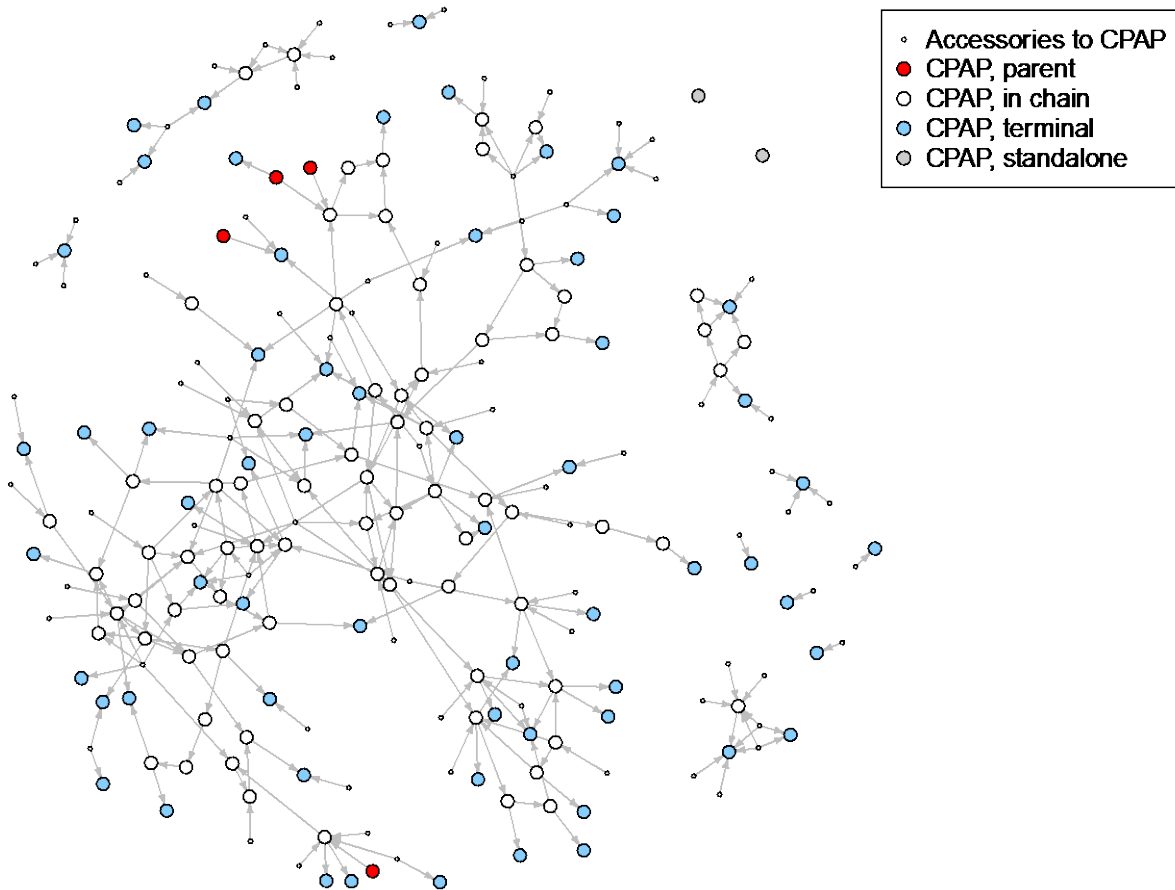
Shown are citation relationships (266 grey arrows) between 143 continuous positive airway pressure (CPAP) device records (large circles) and 78 CPAP accessory records (small circles). The arrows point from the *record being cited* to the *citing record*, to denote the "flow of information" in an equivalence claim. The 4 CPAP records shown as large red dots ("CPAP, parent") are cited by other CPAP device records but do not cite other device records themselves. The 61 records shown as large blue dots ("CPAP, terminal") are CPAP device records that cite other records of CPAP devices (large circles) or CPAP accessories (small circles). Two CPAP device records (large gray circles) are "standalone", in that they are not cited by other devices nor citing

other devices. The remaining 76 CPAP records are shown as large empty circles and are part of a chain of citation relationships ("CPAP, in chain").

Contextual Question 4

What are the variable features of marketed CPAP devices?

Many CPAP device manufacturers exist, each marketing several devices. We reviewed in detail the websites and user manuals for the CPAP device lines from Philips/Respironics, Fisher & Paykel and ResMed, the three manufacturers with the largest market shares in the United States. All three device lines include basic and more feature-rich devices. For each device line, we recorded information in terms of tunable breathing settings, automation, humidity control, monitoring and documentation of use, communication, connectivity and software solutions for patients and providers, and accessories. Because all device lines include comparable features in some of their models, we describe device types generally.

Tunable Breathing Settings and Automation

CPAP devices differ in the flexibility they allow in setting inspiratory and expiratory pressure parameters. Too low an inspiratory pressure will not increase the inhaled volume of air during sleep. Too high inspiratory and expiratory pressures may induce pressure intolerance, a common complaint, or aerophagia or swallowing of air, which is less common. The following features are purported to help alleviate pressure-induced discomfort. We summarize the various proprietary features of these devices in terms of four attributes:

1. Ability to ramp up the inspiratory pressure to a target value over some period of time (e.g., 45 minutes). Most devices have this feature, which purports to help mitigate the discomfort of positive pressure experienced by some patients.
2. Ability to set the inspiratory and expiratory pressures independently. All devices in the reviewed device lines have the ability to set the expiratory pressure up to 3 cm of water lower than the inspiratory pressure. Bilevel devices (BPAP) allow setting inspiratory and expiratory pressures that differ substantially (by more than 3 cm of water pressure). By allowing lower air pressure during exhalation, this feature purports to increase comfort by easing exhalation for patients who need high pressure settings.
3. Ability to titrate pressures dynamically throughout sleep. Some devices have automated algorithms that can vary the inspiratory pressure (and/or the expiratory pressure, for BPAP devices) dynamically. This feature may be used for a titration study (to determine the optimal pressures to set the device) or even during routine use (to allow titration nightly to adjust for night-to-night variation in an individual patient). The algorithms that govern pressures are proprietary. To optimize respiratory pressures, devices use signals from pressure transducers, microphones, and other sensors.[83] These devices are often referred to as auto-adjusting or autoCPAP.
4. Various advanced add-ons, including the ability to vary settings to target a minimum breathing rate, or a minimum tidal volume per breath. These features are found in devices marketed for patients with either obstructive or mixed/central sleep apneas.

Humidity Control

Several devices use heated humidification that may help reduce nasal congestion or dryness and are purported to thus improve adherence to use.

Monitoring and Documentation of Use

Devices differ in their ability to automatically detect changes in the monitored signals, infer and record specific events (e.g., flow limitation, hypopnea, obstructive airway apnea, clear airway apnea, periodic breathing, respiratory-related arousal, types of snores, and presence and types of leaks), calculate indices (e.g., AHI), and even record waveforms. Many devices have the ability to document adherent usage in a tamper-resistant fashion (and communicate these data to the patient's provider), which may be needed for insurance coverage.

Communication, Connectivity, and Software Solutions

Most devices offer numerous input/output options through SD (secure digital) cards, Bluetooth, WiFi, or cellular modem, giving the ability for daily time submission of data to the provider. Separate software suites exist for patients and for providers.

Peripherals

All device lines have a large variety of customizable accessories for interfacing with the patient (e.g., face masks, nasal masks, nasal pillow, custom fits by facial structure), connecting with the device (tubes and pneumatic plumping components), adding function (humidifiers), and other features (e.g., rechargeable batteries). In addition, a variety of materials are used with the goals of minimizing skin irritation and allergies.

Contextual Question 5

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

Patient-centered outcomes are those that patients consider important or relevant. They are a fundamental aspect of patient-centered outcomes research. They include clinically important events (such as death, incident diabetes), symptoms (such as sleepiness), surrogate outcomes commonly measured in clinical practice (e.g., blood pressure, blood sugar), and others. Many commonly measured patient-reported outcomes, such as (health-related) QoL or sexual function, may also be considered to be important to patients. In contrast, many outcomes frequently measured in clinical practice or in research are generally not considered to be patient-centered, including most specific laboratory measures, although these types of results may be the biological linkage to health-related QoL.

CPAP treatment has short- and long-term therapeutic goals. The primary short-term goals of CPAP therapy include alleviating symptoms of fatigue and somnolence; improving activities of daily living, health-related QoL, and cognitive function; resolving snoring; and reducing the risk of work or motor vehicle accidents. In addition to continuation of the short-term goals, the primary long-term goals of CPAP include reducing the risks of a range of chronic disease outcomes and premature death. Primarily based on large database observational studies, but also based on physiologic models and animal studies, long-term OSA has been associated with cardio- and cerebrovascular disease,[84, 85] hypertension,[86] impaired cognitive function (and Alzheimer disease),[87] type 2 diabetes,[88] anxiety and depression symptoms,[89] impaired QoL,[90] and impaired sexual function in both men and women.[91, 92] Although causality has not been established, preventing or alleviating these outcomes are clinical goals of OSA treatment.

For most patients, AHI is not a patient-centered outcome, since its exact value is of secondary importance compared with the clinical sequelae of OSA. However, AHI is commonly measured to evaluate short-term adequacy or effectiveness of treatment. In clinical practice, a reduction in AHI is used to set CPAP pressures and, in both clinical and research settings, as a proxy of the short-term outcome goals. In practice, devices are initially titrated to achieve an AHI reduction to normal levels, but are subsequently adjusted for comfort and/or to alleviate persistent symptoms.

Results

The Results Chapter is organized primarily by Key Question (KQ) and then by outcome (category). High-level summary tables and forest plots describing overall findings across studies are included in the main report. Appendixes C and D include detailed summary tables describing each study and other detailed information.

Overview of the Evidence Base Addressing All Key Questions

The literature database searches yielded 20,804 citations; in addition, we found and screened the references from 87 existing systematic reviews (SR). From the existing SRs, we found 797 citations to retrieve for further screening. Ultimately 52 primary studies (reported in 80 articles) were eligible and included. The reasons for exclusion of articles were: duplicate analysis (no unique data) (590 articles), KQ 2: no apnea-hypopnea index (AHI) change data (258 articles), no outcome of interest (197 articles), too short followup (145 articles), SR (95 articles), no results given (78 articles), full text not available (53 articles), KQ 2 participants <30 total (29 articles), excluded population (28 articles), not primary study (or SR) (17 articles), no comparator of interest (KQ 1) (13 articles), KQ 1 not continuous positive airway pressure (CPAP) device (14 articles), KQ 1 participants <10/arm (11 articles), not obstructive sleep apnea (OSA) (9 articles), case-control/cross-sectional study (9 articles), nonrandomized comparative study (NRCS) without multivariable analysis (KQ 1) (8 articles), and conference abstract (7 citations). See the literature flow figure in Appendix D (Figure D-1) and the list of rejected unique studies in Appendix B for more details.

Definitions of Breathing Measures Used Across Studies

Key Points

- Studies were highly inconsistent in the criteria used to define sleep study measures (apneas, hypopneas, and oxygen desaturation), even among studies stating that definitions are based on the same standard criteria.
- Most studies (60%) did not fully and explicitly report the definitions of sleep study measures used.

Findings

As displayed in Table 2, we categorized each of the 52 eligible studies (for both KQs), including randomized controlled trials (RCTs), NRCSs, and single group (cohort) studies, by the reported definitions used for apnea, hypopnea, and oxygen desaturation, together with the cited criteria for evaluating these, and the types of polysomnography (or home sleep study polygraphy) monitors used. Several issues are evident.

The majority of studies did not explicitly report full criteria or definitions; one-third (17/47, 36%) of studies that evaluated the apnea-hypopnea index (AHI) (excluding 5 studies that evaluated only oxygen desaturation index [ODI] or were registry studies without AHI data) omitted any explicit apnea or hypopnea definitions; only 19 of 47 (40%) fully explicitly reported apnea and hypopnea definitions.

The reported actual criteria (thresholds used) are highly heterogeneous across studies, with few studies using the same sets of criteria. Among the 47 studies that reported at least some

apnea or hypopnea definitions, 13 defined apnea as 100 percent airflow cessation, but 8 used lower thresholds down to 75 percent (well within the definition of hypopnea used by most studies). Among the 20 studies that explicitly reported hypopnea criteria (and used a single hypopnea definition), almost half (9/20) required at least 50 percent airflow cessation and about half allowed at least 30 percent (or in one case 25%) airflow cessation. There was similar heterogeneity in minimum thresholds to define oxygen desaturation. Among 35 studies that explicitly reported oxygen desaturation thresholds (either as part of the hypopnea definition or to determine ODI), a bit more than half (20/35, 57%) used a 4 percent desaturation threshold and about one third (13/35, 37%) used a 3 percent threshold. One study used a 3 percent threshold to define hypopnea, but a 4 percent threshold to define ODI. One study used a 3 percent threshold for their OSA definition of AHI ≥15, but a 4 percent threshold for their OSA definition of AHI ≥10.

Among the 32 studies that cited published criteria to define apneas or hypopneas, 26 (81%) cited some version of the American Academy of Sleep Medicine (AASM) criteria. However, among the 17 that cited AASM criteria and explicitly reported apnea, hypopnea, or oxygen desaturation definitions, there was no discernable consistency in choice of a threshold and citation of a specific AASM version. For example, among the eight studies that cited the ASSM 1999 criteria, three explicitly defined apnea as 100 percent airflow cessation and two used a 90 percent threshold (3 did not report an apnea definition). Among the six studies that cited the AASM 1997 criteria, three each used either a 3 percent or a 4 percent threshold to define oxygen desaturation. However, it is important to note that it was generally unclear whether authors were citing the AASM (or other criteria) as the source for the set of criteria used or to cite a specific criterion (e.g., only hypopnea).

As will be described more fully in the following section, *Key Question 1: CPAP Versus No CPAP Treatment Effect*, we could not discern whether the clinical findings of studies (i.e., effect sizes for specific health outcomes) differed in relation to how apnea, hypopnea, ODI, or OSA were defined. For most health outcomes, studies were statistically homogeneous (the effect sizes did not significantly differ from each other) and/or estimates were imprecise. Therefore, possible differences in results could not be elucidated based on any participant or study characteristic, including definition of sleep study measures and OSA.

Table 2. Apnea, hypopnea, and desaturation criteria used by eligible studies.

Study PMID	Design [A]	Monitor Type	Apnea Threshold	Hypopnea Threshold	Hypopnea Desaturation %	ODI Desaturation %	Criteria Cited
Barbé 2012 22618923	**RCT CPAP vs No**	**III home**	**100%**	**50%**	**4%**	n/a	**NR**
Lindberg 2012 22499826	Cohort	III home	100%	50%	4%	4%	NR
Meurice 2007 17638595	RCT CPAP vs. CPAP [A]	NR	100%	50%	4%	n/a	AASM 1997
Tegelberg 2012 23620682	Cohort	III home	100%	50%	4%	n/a	AASM 1999
Campos-Rodriguez 2005 16100147	**NRCS CPAP vs No**	**I lab or III home**	**100%**	**50%**	**4%**	n/a	**NR**
Crawford-Achour 2015 25700873	*NRCS CPAP vs No*	*III home*	*100%*	*50%*	*3%*	*3%*	*NR*
ONSLEEP 33006309	*NRCS CPAP vs No*	*I lab*	*100%*	*50%*	*3%*	*n/a*	*NR*
Sforza 2017 28225159	Cohort	III home	100%	50%	3%	n/a	AASM 1999
Jaoude 2014 24452812	Cohort	I lab	100%	30%	4%	n/a	AASM 1999
Navarro-Soriano 2021 32029279	*NRCS CPAP vs No*	*I lab*	*100%*	*30-90%*	*4%*	*n/a*	*SNCSAHS 2005*
Wu 2015 (Beijing) 25412159	*NRCS CPAP vs No*	*I or III lab*	*100%*	*30%*	*4%*	*n/a*	*Other*
Peppard 2000 10805822	Cohort	I lab	100%	Not used	4%	n/a	NR
Corral 2017 28636405	RCT Other [C]	I lab or III home	100%	NR	3%	n/a	AASM 2007
Pelletier-Fleury 2004 14738231	**RCT CPAP vs No**	**I lab**	**100%**	**NR**	**3%**	n/a	**Other**
ISAACC 31839558	**RCT CPAP vs No**	**I lab**	**100%**	**NR**	**4%**	n/a	**SNCSAHS 2005**
Berg 2020 32665778	RCT CPAP vs Other	III lab or hotel	90%	50%	3%	n/a	*AASM 2007*
RICCADSA 26914592	**RCT CPAP vs No**	**III home**	**90%**	**1: 50%** **2: 30%**	**1: n/a (not used)** **2: 4%**	**4%**	**AASM 1999**
Banhiran 2014 24458949	NRCS Other [C]	NR	90%	30-89%	4%	n/a	AASM 2007
Huang 2015 25125635	**RCT CPAP vs No**	**I lab**	**90%**	**30-90%**	**4%**	n/a	**AASM 1999**
López-Padilla 2016 27198943	*NRCS CPAP vs No*	*I lab or III home*	*90%*	*30-90%*	*3%*	*n/a*	*AASM 2007 or SNCSAHS 2005*
Uchôa 2017 28823814	Cohort	III home	90%	30%	3%	n/a	AASM 2012
Bjornsdottir 2015 25431105	*NRCS CPAP vs No*	*II home*	*80%*	*1: 50%* *2: 30%*	*1: n/a (not used)* *2: 4%*	*4%*	*NR*
Lisan 2019 30973594	*NRCS CPAP vs No*	*II home*	*75%*	*25-69%*	*4%*	*n/a*	*NR*
Bloch 2017 28982804	RCT CPAP vs. CPAP [A]	I lab or III home	NR	50%	3%	n/a	AASM 1999
Ou 2015 26068440	*NRCS CPAP vs No*	*I lab or II home*	*NR*	*1: >50%* *2: ≤50%*	*1: n/a (not used)* *2: 3%*	*4%*	*AASM 1999*
Hasselbacher 2018 29808422	Cohort	III home	NR	30%	4%	n/a	AASM 2007
Woodson 2018 29582703	Cohort	I lab	NR	30%	4%	n/a	AASM 2007
BestAIR 28419387	**RCT CPAP vs Sham**	**I lab or II home**	**NR**	**NR**	**3% or 4%**	n/a	**NR**
MOSAIC 23111478, 24508706	**RCT CPAP vs No**	**III or IV home**	**NR**	**NR**	**NR**	**4%**	**NR**

Study PMID	Design [A]	Monitor Type	Apnea Threshold	Hypopnea Threshold	Hypopnea Desaturation %	ODI Desaturation %	Criteria Cited
de Ruiter 2018 28913630	RCT Other [C]	I lab	NR	NR	NR	4%	AASM 2012
SAVE 27571048, 25669180	**RCT CPAP vs No**	**III home**	**n/a**	**n/a**	**n/a**	**4%**	**NR**
PREDICT 25172769	**RCT CPAP vs No**	**NR home**	**NR**	**NR**	**NR**	**4%**	**Other**
Gagnadoux 2017 28947040	NRCS Other [C]	III home	NR	NR	NR	3%	AASM 2007
Lin 2015 25766707	Cohort	IV home	n/a	n/a	n/a	3%	n/a
Shaw 2016 26926656	**RCT CPAP vs No**	**III home**	**NR**	**NR**	**NR**	**3%**	**NR**
Aarab 2017 28083705	**RCT CPAP vs No vs Other [D]**	**NR**	**NR**	**NR**	**NR**	**n/a**	**AASM 1999**
Botros 2009 19958890	*NRCS CPAP vs No*	*I lab*	*NR*	*NR*	*NR*	*n/a*	*AASM 1999*
Budweiser 2013 23088487	*NRCS CPAP vs No*	*I lab*	*NR*	*NR*	*NR*	*n/a*	*AASM 1999*
Chang 2020 32421898	*NRCS CPAP vs No*	*NR*	*NR*	*NR*	*NR*	*n/a*	*NR*
de Vries 2019 31596213	RCT CPAP vs. Other [D]	I lab or II home	NR	NR	NR	n/a	AASM No date
Fernández-Julián 2017 29152745	NRCS Other [C]	II home	NR	NR	NR	n/a	AASM 2012
Jara 2018 29800001	*NRCS CPAP vs No*	*I lab*	*NR*	*NR*	*NR*	*NR*	*AASM 2015*
Jennum 2015 25914563	*NRCS CPAP vs No*	*Any*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Kingshott 2000 10712335	Cohort	I lab or II home	NR	NR	NR	NR	NR
Kushida 2011 21804670	RCT CPAP vs. CPAP [A]	NR	NR	NR	NR	NR	NR
APPLES 23204602	**RCT CPAP vs Sham**	**I lab**	**NR**	**NR**	**NR**	**n/a**	**AASM 1999**
Lau 2013 23766914	Cohort	NR	NR	NR	NR	NR	NR
Lin 2006 16735919	Cohort	NR	NR	NR	NR	NR	NR
Monasterio 2001 11587974	**RCT CPAP vs No**	**I lab or II home**	**NR**	**NR**	**NR**	**n/a**	**Other**
Myllylä 2019 30848437	*NRCS CPAP vs No*	*I lab or NR home*	*NR*	*NR*	*NR*	*n/a*	*AASM 1999*
Schipper 2017 28550476	*NRCS CPAP vs No*	*I lab or II home*	*NR*	*NR*	*NR*	*n/a*	*AASM 2007*
Schulz 2019 29773460	Cohort	I lab	NR	NR	NR	n/a	NR
Sheth 2021 33729910	*NRCS CPAP vs No*	*NR*	*NR*	*NR*	*NR*	*n/a*	*NR*
Wu 2016 (Yangzhou) 26993342	**RCT CPAP vs No**	**I lab or II home**	**NR**	**NR**	**NR**	**NR**	**AASM 2007**

Studies ordered by 1) apnea threshold; 2) hypopnea threshold, minimum; 3) hypopnea desaturation threshold (each highest to lowest, then NR or n/a).

Monitor types are indicated and color coded by lab (light blue), home (light purple), either lab or home (light orange), any (no shading), or NR (light grey).

Highest thresholds are color coded in red (apnea 100%, hypopnea 50%, desaturation 4%).

Lower thresholds are color coded in orange (apnea 90%, hypopnea 30%, desaturation 3%).

Lowest thresholds are color coded in yellow (apnea <90%, hypopnea <30%).

Variable definitions, when multiple criteria were used are color coded in peach.

The AASM criteria are colored in shades of blue-grey, with brighter/bluer shades indicating more recent years. Spanish National Consensus in Sleep Apnea-Hypopnea Syndrome criteria are colored in green. Other criteria are colored in pink.

The shading and text format in cells do not provide unique information not included in the table text. RCTs of CPAP versus no (or sham) CPAP are in **bold**, larger font. NRCSs of CPAP versus no CPAP are in *italic*, larger font. Other included studies are in smaller font.

The subtlety of whether a threshold was greater than (>) or greater than or equal to (≥) is omitted.

Abbreviations: AASM = American Academy of Sleep Medicine, n/a = not applicable, Cohort = single-group cohort analysis, CPAP = continuous positive airway pressure, NR = not reported, NRCS = nonrandomized comparative study, ODI = oxygen desaturation index, PMID = PubMed Identifier, RCT = randomized controlled trial, SNCSAHS = Spanish National Consensus in Sleep Apnea-Hypopnea Syndrome.

A Randomized controlled trials of CPAP vs. no CPAP are in bold, larger (10 point) font. Nonrandomized comparative studies (with adjustment for potential confounders) of CPAP vs. no CPAP are in italic, larger (10 point) font.
B Or other positive airway pressure, such as auto-adjustable.
C Not a comparison of CPAP.
D E.g., CPAP vs. mandibular advancement device.

Key Question 1: CPAP Versus No CPAP Treatment Effect

Key Points

- Comparative studies (RCTs and adjusted NRCSs) provide, at best, low strength of evidence (SoE) on the effect of CPAP on long-term health outcomes. We did not evaluate sleepiness, other symptoms, or other important outcomes such as intermediate outcomes (including blood pressure and glucose). We did not evaluate other study designs, including comparisons between patients with OSA who use CPAP and healthy controls or pre-post studies.
- 31 studies (14 RCTs and 17 NRCSs) compared CPAP with no CPAP, of which 12 RCTs primarily reported intention-to-treat (ITT) analyses, 6 NRCSs provided "as-prescribed" analyses (of all patients prescribed vs. not prescribed CPAP), and 13 (2 RCTs and 11 NRCSs) provided analyses of CPAP adherent patients versus nonadherent or nonusers. Two RCTs used sham CPAP as the comparator.
- Few comparative studies were explicitly powered for long-term health outcomes: 3 RCTs were designed to be powered for composite cardiovascular (CV) outcomes (SAVE, RICCADSA, ISAACC); however, RICCADSA was ultimately underpowered. One RCT (APPLES) was powered for neurocognitive outcomes. No eligible study was explicitly powered for any other outcome. Patients enrolled in RCTs had variable, but generally poor adherence; where reported, between 35 and 64 percent of patients used their devices for at least 4 hours per night (at least 70% of nights). The NRCSs that reported "as-prescribed" analyses mostly did not have data related to CPAP adherence. The NRCSs that reported analyses of CPAP users versus nonusers mostly compared adherent users (>4 hours/night) to those who used their devices less.
- Five RCTs did not provide evidence that CPAP reduces the risk of all-cause death (low SoE). The RCTs included patients with moderate or severe OSA (AHI ≥15) who were mostly at increased risk of CV death, but without excessive daytime sleepiness. The RCTs were at moderate risk of bias. The studies had relatively short duration of followup (2 to about 5 years) and provided a nonsignificant summary estimate (effect size [ES] 0.89, 95% CI 0.66 to 1.21).
  - Inclusion of six NRCSs (with about 3 to 11 years of followup) adjusted for various potential confounders in propensity analyses or logistic regression suggests that CPAP may be associated with decreased risk of all-cause mortality in longer-term followup among all adults with moderate to severe OSA, with or without sleepiness symptoms (low SoE).
  - The overall summary ES (RCTs and NRCSs combined) was 0.61 (95% CI 0.49 to 0.76).

- o Both within RCTs and across RCTs and NRCSs, ITT analyses (of RCTs) or "as-prescribed" analyses (of patients prescribed CPAP in NRCSs) yielded similar summary ESs as analyses of adherent versus nonusers.
- o Comparative studies (both RCTs and adjusted NRCSs) provide insufficient evidence to support whether effect sizes may differ in different subgroups of patients, including by severity of OSA (as assessed by AHI or similar measures), severity of symptoms (e.g., as measured by Epworth Sleepiness Scale [ESS]), age, other risk factors or comorbidities.
- Four RCTs do not provide evidence that CPAP reduces the risk of CV mortality (low SoE). Addition of one NRCS does not change the conclusion; the overall summary ES was imprecise (0.79, 95% CI 0.50 to 1.26; low SoE).
- Five RCTs do not provide evidence that CPAP affects the risk of stroke (summary ES 0.99, 95% CI 0.73 to 1.34; low SoE) or of acute myocardial infarction (summary ES 1.05, 95% CI 0.78 to 1.41; low SoE). Both summary ESs are imprecise. One NRCS did not alter the conclusion regarding stroke.
- Four RCTs do not provide evidence that CPAP affects the risk of atrial fibrillation (AFib). The summary ES was imprecise (0.94, 95% CI 0.58 to 1.51; low SoE).
- Seven RCTs do not provide evidence that CPAP affects the risk of various composite CV outcomes. Effect sizes were mostly imprecise and ranged from 0.42 to 1.10 across studies, all statistically nonsignificant (low SoE). Six of the RCTs were evaluated as ITT analyses of all users, in which adherence ranged from 35 to 64 percent. Additional evidence from five NRCSs adjusted for various potential confounders in propensity score analyses or logistic regression did not change this conclusion. The range of ESs among the NRCSs was similar to that of the RCTs (0.38 to 0.88; 2 of which were statistically significant). The NRCSs all evaluated adherent CPAP users.
- There is insufficient evidence regarding the effect of CPAP on other CV outcomes, including transient ischemic attacks (TIA), angina-related outcomes, coronary artery revascularization, congestive heart failure (CHF) outcomes, or AFib. For each outcome, there were sparse studies and/or effect estimates were highly imprecise.
- Two RCTs provide low SoE that CPAP does not affect the likelihood of driving accidents after 1 or about 3 to 4 years of use.
- Three RCTs do not provide evidence that CPAP affects the risk of incident type 2 diabetes mellitus (DM); the summary ES was imprecise (1.02, 95% CI 0.69 to 1.51; low SoE). An additional adjusted NRCS, though statistically significant, does not alter the conclusion (overall summary ES 0.88, 95% CI 0.57 to 1.35).
- Comparative studies do not provide evidence that CPAP results in clinically meaningful changes in depression symptoms (4 RCTs), anxiety symptoms (4 RCTs), executive cognitive function (4 RCTs and 1 NRCS), or quality of life (QoL, 10 RCTs and 1 NRCS); although the small differences were statistically significant (all low SoE).
- There is insufficient evidence regarding the effect of CPAP on other outcomes, including hypertension (HTN), sexual function, and days of work missed. For each outcome, there were sparse studies, effect estimates were highly imprecise, and/or studies reported highly inconsistent results. We did not evaluate sleepiness or intermediate outcomes.
- Based on the evaluated long-term, comparative studies, there is insufficient evidence regarding the risk of adverse events specific related to CPAP use.

- Comparative studies do not provide evidence regarding whether CPAP use is more (or less) effective in any specific subgroup. For all health outcomes, studies provide insufficient evidence to suggest whether the effect of CPAP on all-cause mortality varies by baseline AHI, age, body weight, or pulmonary function. Within and across studies, associations did not clearly or consistently differ based on diagnostic criteria for OSA, definitions of AHI and other sleep study measures, or whether analyses were based on CPAP prescription (e.g., ITT analyses) or CPAP adherence.

Evidence Base

Thirty-one studies (in 46 publications) compared CPAP with no (or rarely sham) CPAP treatment. Their design, eligibility criteria, adherence information, and reported outcomes are summarized in Table 3. These included 14 RCTs[12, 93-119] and 17 NRCSs[120-135] with multivariable adjustments for outcomes of interest. The 14 RCTs included a total of 7449 participants. One of the NRCSs was relatively large (n = 25,389); the remaining 16 NRCSs included 9073 participants. Among the 17 NRCSs, 3 used propensity score matching, which attempts to make treatment and non-treatment groups comparable with respect to factors that predict each individual's likelihood of using CPAP. The remaining 14 NRCSs used traditional logistic regression multivariable analyses to control for specific possible confounders. However, it is unclear whether the NRCSs, particularly those that reported adherent analyses, could have fully accounted for possible biases related to self-selection regarding CPAP use and adherence.

Study eligibility criteria were variable across studies. Among the 14 RCTs, 5 included adults with OSA regardless of other major comorbidities,[93, 97, 105, 108, 109] 5 were restricted primarily to participants with cardiovascular disease (CVD) or cerebrovascular disease (CeVD) (1 post-revascularization[112]),[12, 101-103, 112] 3 excluded patients with CVD (1 also with CV risk factors[119]),[99, 104, 119] and 1 was restricted to patients with type 2 DM.[118] One RCT was restricted to older adults (≥65 years)[109] and 7 trials excluded older patients (>65-75 years).[12, 99, 101, 102, 105, 108, 119] Most of the 14 RCTs included patients with at least "moderate" OSA (e.g., AHI ≥15); 1 RCT (Monasterio 2001[104]) excluded patients with "severe" OSA (AHI >30) and a single study was restricted to patients with "severe" OSA (the SAVE study in which their measure of ODI ≥12 was determined to be equivalent to AHI ≥30).[12]

Only 5 of the 17 NRCSs restricted their participant samples by comorbidities: 1 was restricted to patients who recently had percutaneous coronary intervention with a drug-eluting stent,[135] 1 excluded patients with CVD or CeVD or CV risk factors,[124] 1 included only patients with HTN,[129] 1 included only patients with type 2 DM,[134] and 1 excluded patients with DM.[121] Among the NRCSs, 3 were restricted to older adults (≥60, 65, or 80 years).[124, 127, 132]

Eighteen of the studies (12 RCTs, 6 NRCSs) compared participants who were prescribed CPAP with those not prescribed CPAP or given a sham CPAP device (ITT in RCTs, "as-prescribed" analyses in NRCSs). Thirteen studies (2 RCTs, 11 NRCSs) compared adherent CPAP users mostly with all nonusers, including both those who never started CPAP and those who were nonadherent. Two RCTs gave the control participants sham CPAP devices.[95-98, 100, 101]

Each study is described in some detail, including study design, analysis type (ITT/"as-prescribed" vs. adherent), eligibility criteria, participant characteristics, reported power analyses, and risk of bias concerns. Full design and arm details, baseline details, and risk of bias details for all included studies are in Appendix C, Tables C-1 through C-8.

As is presented in summary tables below and discussed in various outcome-specific sections, adherence with CPAP use varied across studies. We were unable to discern any obvious patterns

across studies regarding low versus high lack of adherence, including duration of the studies. Among the 13 RCTs that reported on adherence, 6 reported adherence rates of 35 to 43 percent (at 6 months in 3 RCTs, 12 months in 1 RCT, and at means of 40 and 44 months in 2 trials),[12, 97, 101, 103, 105, 109] 3 RCTs reported adherence rates of 60 to 64 percent (at 6 months and 1 and 4 years),[99, 112, 118] and 1 RCT reported an adherence rate of 86 percent (at 3 years).[102] Average CPAP usage varied from 2.5 to 5.8 hours per night. None of the eligible RCTs evaluated reasons why participants were nonadherent. Three RCTs (APPLES, Barbé 2012, and ISAACC) reported that adherent CPAP users were heavier (or had larger neck circumference) and had more abnormal sleep study measures (such as higher AHI, lower minimum oxygen saturation, or longer sleep latency).[97, 99, 103]

Of note, in regards to Key Question 1b (within-study concordance between apnea and hypopnea indices, sleep questionnaires, and health outcomes), no study reported within-study correlations among outcomes (e.g., effect on AHI and effect on all-cause mortality). Our analysis of these potential correlations across studies is described in the section, below, *Key Questions 1b and 2: Intermediate and Surrogate Measures*.

Also note that where each included study is described for the first time (in detail), we have put the study or author/name in bold font to assist the reader to find the relevant description when the study is included in subsequent sections.

Table 3. Summary of design and reported outcomes for CPAP versus no CPAP studies

Study	Design	Analysis	Total N	Duration (mo)	Population	Age Restriction	AHI Eligibility	ESS	Adherence	All-Cause Mortality	CV Mortality	Stroke/TIA	AMI	Angina	Revascularization	CHF	Atrial Fibrillation	Composite CV	Accident	Hypertension	Diabetes	Mental Health	Cognitive	QoL/Function	Sexual Function	Other
Aarab 2017[93,94]	RCT	ITT	37	6	All	none	≥5	≥10	83% noc (6 mo)	.	.	.	.	.	.	.	.	.	.	.	.	X	.	.	X	.
APPLES[95-98]	RCT (sham)	ITT	1098	6	All	none	>10	none	42% (6 mo)	.	.	.	.	.	.	.	.	.	.	.	.	.	others P	QWB SAQLI	.	.
Barbé 2012[99]	RCT	ITT	723	48	No CVD	≤70	≥20	≤10	64% (48 mo)	X	X	X	X	.	.	X	.	X	.	X	.	.	.	.	.	.
BestAIR[100,101]	RCT (sham)	ITT	169	12	CVD	45-75	≥10	none	43% (6 mo)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	SF-36	.	.
Huang 2015[102]	RCT	As-treated	83	36	CVD	45-75	≥15	none	86% (36 mo)	.	.	.	.	.	.	.	.	X	.	X	X	.	.	.	.	.
ISAACC[103]	RCT	ITT	1255	40	CVD (ACS)	none	≥15	≤10	38% (40 mo)	X	X	X	X	X	X	X	X	P	.	.	X	.	.	EQ-5D	.	.
Monasterio 2001[104]	RCT	ITT	125	6	No CVD	none	10-30	not severe	4.8 hr/noc	.	.	.	.	.	.	.	.	.	.	.	.	.	TMT	FOSQ	.	.
MOSAIC[105-107]	RCT	ITT	188	60	All	45-75	>7.5 (ODI)	<10	38% (6 mo)	X	.	.	.	.	.	.	.	X	.	.	.	.	.	SF-36 EQ-5D SAQLI	.	.
Pelletier-Fleury 2004[108]	RCT	As-treated	171	6	All	≤70	NR	none	4.8-5.5 hr/noc	.	.	.	.	.	.	.	.	.	.	.	.	.	TMT	.	.	.
PREDICT[109]	RCT	ITT	231	12	All	≥65	>7.5 (ODI)	≥9	35% (12 mo)	.	.	X	X	X	.	.	X	X	X	.	.	X	MMSE TMT	SAQLI	.	.
RICCADSA[110-113]	RCT	ITT	244	57	CVD (revasc)	none	≥15	<10	60% (12 mo)	X	X	X	X	.	X	.	.	P	.	X	.	.	.	SF-36	.	.
SAVE[12,114-117]	RCT	ITT	2687	44	CVD	45-75	≥30 (ODI ≥12)	<16	42% (44 mo)	X	X	X	X	X	X	X	X	P	X	.	X	X	.	SF-36 EQ-5D	.	Missed work
Shaw 2016[118]	RCT	ITT	302	6	DM	none	≥15 (ODI)	none	61% (6 mo)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	SF-36	.	.
Wu 2016 (Yangzhou)[119]	RCT	ITT	136	6	No CVD or RF	30-65	≥15	none	NR	.	.	.	.	.	.	.	.	.	.	.	.	.	MMSE	.	.	.
Bjornsdottir 2015[120]	NRCS	Adh vs. nonuse	562	24	All	none	≥15	none	100% (pp)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	SF-12	.	.
Botros 2009[121]	NRCS	Adh vs. nonuse[A]	266	32	No DM	none	≥20	none	NR	.	.	.	.	.	.	.	.	.	.	.	X	.	.	.	.	.
Budweiser 2013[122]	NRCS	Adh vs. nonuse	83	36	All	none	NR	none	NR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.
Campos-Rodriguez 2005[123]	NRCS	Adh vs. nonadh	749	48	All	none	≥10	none	variable	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Chang 2020[136]	NRCS (ps)	As-prescribed	1918	59	All	none	NR	none	NR	.	.	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Crawford-Achour 2015[124]	NRCS	As-prescribed	126	120	No CVD	≥65	>30	none	NR	.	.	.	.	.	.	.	.	.	.	.	.	.	MMSE TMT	.	.	.
Jara 2018[151]	NRCS	Adh vs. nonuse	182	12	All	none	≥5	none	100% (pp)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.
Jennum 2015[125]	NRCS	As-prescribed	25389	36	All	none	NR	none	NR	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Lisan 2019[126]	NRCS (ps)	As-prescribed	255	133	All	≥40	NR	none	NR	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
López-Padilla 2016[127]	NRCS	Adh vs. nonuse[A]	155	53	All	≥80	≥20	none	100% (pp)	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Myllylä 2019[128]	NRCS	Use vs. nonuse[B]	2060	104	All	none	≥15	none	100% (pp)	.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.	.
Navarro-Soriano 2021[129]	NRCS	Adh vs. nonuse[A]	163	58	HTN	none	≥15	none	100% (pp)	.	.	.	.	.	.	.	.	X	X	X	.	X	.	.	.	.
ONSLEEP[130,131]	NRCS (ps)	As-prescribed	1274	79	All	none	>5	none	NR	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Ou 2015[132]	NRCS	As-prescribed	124	60	All	≥60	≥20	none	67% (60 mo)	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
O1	NRCS	Adh vs. nonadh	283	71	All	none	≥5	none	100% (pp)	.	.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.
Sheth 2021[134]	NRCS	Adh vs. nonadh	578	30	DM	none	NR	none	100% (pp)	.	X	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.
Wu 2015 (Beijing)[135]	NRCS	Adh vs. nonuse[A]	295	60	CVD (revasc)	none	≥15	none	100% (pp)	.	.	.	.	.	X	.	.	X	.	.	.	.	.	.	.	.

33

A and nonadherent users.   B Long-term CPAP users versus those who discontinued CPAP.

Abbreviations: . = outcome not reported, ACS = acute coronary syndrome, Adh = adherent, AHI = apnea-hypopnea index eligibility criteria, AMI = acute myocardial infarction, CHF = congestive heart failure, CV = cardiovascular, CVD = cardiovascular disease, DM = diabetes mellitus, EQ-5D = EuroQoL-5D scale, ESS = Epworth Sleepiness Scale eligibility criteria, FOSQ = Functional Outcomes of Sleep Questionnaire, hr = hour(s), HTN = hypertension, ITT = intention-to-treat, MMSE = Mini-Mental State Exam, mo = months, noc = night(s), nonadh = nonadherent, NR = not reported, NRCS = nonrandomized comparative study, ODI = oxygen desaturation index, P = outcome for which study was powered (RCTs colored bright orange, NRCSs colored bright green), pp = per protocol, ps = adjusted with propensity score, QoL = quality of life, QWB = Quality of Well-Being, RCT (sham) = RCT using sham device comparator, RCT = randomized controlled trial, revasc = coronary revascularization, RF = cardiovascular risk factors, SAQLI = Sleep Apnea Quality of Life Index, SF = Short Form, TIA = transient ischemic attack, TMT = Trail Making Test (A or B), X = outcome reported (RCTs colored orange, NRCSs colored green), The trial acronyms are defined in the text.

Mortality and Cardiovascular Outcomes (CPAP Versus No CPAP)

Table 4 summarizes the evidence from RCTs and adjusted NRCSs regarding the effect of CPAP (versus no CPAP) on long-term mortality and CV outcomes. In brief, comparative studies provide at most low SoE for all CV outcomes. The low SoE for the conclusions suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding that the findings are stable or that the estimate of effect is close to the true effect.

We conclude that the RCTs do not provide evidence of an effect of CPAP on long-term all-cause mortality, CV death, stroke, myocardial infarction, or a heterogeneous collection of composite CV outcomes. However, analyses of adherent CPAP users within RCTs and adjusted NRCS analyses of both all CPAP users and adherent users found associations between CPAP use and reduced mortality. Thus, we conclude that comparative studies (RCTs and adjusted NRCSs), together, are associated with reduced mortality (again, with low SoE, suggesting that future trials are needed to determine the accuracy of the conclusion). NRCSs did not alter conclusions for other CV outcomes. There is insufficient evidence from comparative studies for TIAs, angina, coronary artery revascularization, and CHF.

Table 4. Evidence profile for CPAP versus no CPAP: Mortality and cardiovascular outcomes

Outcome	Study Design	No. Studies (Subjects)	Risk of Bias	Consistency	Precision	Directness	Other	Overall SoE	Conclusion Statements Summary ES/NNT (95% CI)
Death, all-cause	**RCT (ITT)**	**5 (5097)**	**Moderate**	**Consistent**	**Imprecise**	**Indirect**	**None**	**Low**	**No evidence of effect 0.89 (0.66, 1.21) NNT 268 (NNH 142, NNT 86)**
	RCT (adherent users)	2 (2650)	Moderate	Consistent	Imprecise	Indirect	None	not evaluated	Possible assn with reduced death 0.73 (0.51, 1.04)
	adj NRCS (as-prescribed)	4 (28,453)	Moderate	Consistent	Precise	Direct	None	not evaluated	Assn with reduced death 0.55 (0.42, 0.74)
	adj NRCS (adherent users)	*2 (904)*	*High*	*Consistent*	*Precise*	*Indirect*	*None*	*not evaluated*	*Assn with reduced death 0.41 (0.26, 0.64)*
	Overall	*11 (34,454)*	*Moderate*	*Inconsistent*	*Precise*	*Direct*	*None*	*Low*	*CPAP may reduce death ITT/as-prescribed 0.66 (0.53, 0.81) Users 0.55 (0.38, 0.81)*
Death, cardiovascular	**RCT**	**4 (4909)**	**High**	**Consistent**	**Imprecise**	**Direct**	**None**	**Low**	**No evidence of effect ES 0.99 (0.64, 1.53) NNT 5656 (NNH 108, NNT156)**
	adj NRCS	1 (2558)	High	N/A	Precise	Direct	Single study	not evaluated	No conclusion
	Overall	*5 (7457)*	*High*	*Consistent*	*Imprecise*	*Direct*	*None*	*Low*	*No evidence of effect ES 0.79 (0.50, 1.26)*
Stroke	**RCT**	**5 (5140)**	**Moderate**	**Consistent**	**Imprecise**	**Direct**	**None**	**Low**	**No evidence of effect ES 0.99 (0.73, 1.35) NNT 5364 (NNH 154, NNT 198)**
	adj NRCS	1 (1918)	High	N/A	Imprecise	Direct	None	not evaluated	
	Overall	*6 (7058)*	*Moderate*	*Consistent*	*Imprecise*	*Direct*	*None*	*Low*	*No evidence of effect ES 0.90 (0.70, 1.17)*
TIA	**RCT**	**4 (4891)**	**Moderate**	**Consistent**	**Highly imprecise**	**Direct**	**None**	**Insufficient**	**No conclusion**
AMI	**RCT**	**5 (5140)**	**Moderate**	**Consistent**	**Imprecise**	**Direct**	**None**	**Low**	**No evidence of effect ES 1.05 (0.78, 1.41) NNH 653 (NNH 80 , NNT 147)**

Outcome	Study Design	No. Studies (Subjects)	Risk of Bias	Consistency	Precision	Directness	Other	Overall SoE	Conclusion Statements Summary ES/NNT (95% CI)
Angina [A]	**RCT**	**3 (4172)**	**Moderate**	**Consistent**	**Imprecise**	**Direct**	**Sparse [A]**	**Insufficient**	**No conclusion**
Revascularization, coronary artery [B]	**RCT**	**2 (1499)**	**Moderate**	**Consistent**	**Precise**	**Direct**	**Sparse [C]**	**Insufficient**	**No conclusion**
	adj NRCS	1 (295)	Moderate	N/A	Precise	Direct	Single study	Insufficient	No conclusion
	Overall	*2 (2982)*	*High*	*Inconsistent [D]*	*Imprecise*	*Direct*	*None*	*Insufficient*	*No conclusion*
CHF [E]	**RCT**	**3 (4665)**	**Moderate**	**Consistent**	**Highly imprecise**	**Direct**	**Sparse [E]**	**Insufficient**	**No conclusion**
	adj NRCS	1 (163)	High	N/A	Highly imprecise	Direct	Single study	Not evaluated	No conclusion
	Overall	*4 (4828)*	*Moderate*	*Consistent*	*Highly imprecise*	*Direct*	*Sparse*	*Insufficient*	*No conclusion*
AFib	**RCT**	**3 (4173)**	**Moderate**	**Consistent**	**Imprecise**	**Direct**	**None**	**Low**	**No evidence of effect ES 0.89 (0.48, 1.63) NNT 223 (NNH 42, NNT 48)**
	adj NRCS	1 (163)	High	N/A	Highly imprecise	Direct	Single study	Not evaluated	No conclusion
	Overall	*4 (4336)*	*Moderate*	*Consistent*	*Imprecise*	*Direct*	*None*	*Low*	*No evidence of effect ES 0.94 (0.58, 1.51)*
Composite CV outcomes	**RCT**	**7 (5401)**	**High**	**Consistent**	**Imprecise**	**Indirect [F]**	**Clinical heterog**	**Low**	**No evidence of effect ES range 0.42-1.10 (all NS) [G] Range NNH 63 to NNT 10**
	adj NRCS	5 (3337)	High	Inconsistent	Precise	**Indirect [F]**	Clinical heterog	not evaluated	CPAP associated with lower risk (in patients not restricted by CV history), but no evidence of effect in patients with CV risk ~0.37 [H] and 0.83
	Overall	*12 (8738)*	*High*	*Inconsistent*	*Imprecise*	*Indirect [F]*	*Clinical heterog*	*Low*	*No evidence of effect*

Evaluations of RCT evidence base are in bold font. Evaluations of all studies (RCTs and adjusted NRCSs together) are in italic font.

Abbreviations: adj NRCS = adjusted nonrandomized comparative studies, AFib = atrial fibrillation, AMI = acute myocardial infarction, Assn = association, CHF = congestive heart failure, CI = confidence interval, Clinical heterog = clinically heterogeneous (e.g., different populations, eligibility criteria), CV = cardiovascular, ES = (summary) effect size, HR = (summary) hazard ratio, N/A = not applicable, NNH = (summary) number-needed-to harm, NNT = (summary) number-needed-to-treat, NS = not statistically significant, OR = (summary) odds ratio, RCT = randomized controlled trials, sig = statistically significant, SoE = strength of evidence.

[A] One study reported on risk of incident angina and two studies reported on risk of hospitalization for unstable angina, providing an imprecise estimate.
[B] An additional RCT (SAVE) also evaluated risk of any major artery revascularization; as a unique study, it provided insufficient evidence.
[C] Each study evaluated a similar, but distinct outcome (revascularization in patients with acute coronary syndrome, repeat coronary revascularization in patients with prior revascularization).
[D] Inconsistent findings regarding repeat revascularization.
[E] Two studies reported on risk of incident CHF and two studies reported on risk of hospitalization for CHF.

Thirteen studies, including six RCTs and eight NRCSs with adjustment for confounding evaluated one or more mortality or CV event outcomes. (Unadjusted analyses of specific outcomes from otherwise eligible NRCSs are omitted.) For studies reporting mortality and CVD outcomes, we provide a brief description of each study, first RCTs then NRCSs, in alphabetical order. Throughout the report, the name (or first author and publication date) of each study is in bold text in the paragraph with the brief study description.

Randomized Controlled Trials

Barbé 2012 randomly allocated relatively young, otherwise healthy participants to CPAP (n = 357) or no CPAP (n = 366) for 4 years.[99] The trial was conducted and analyzed with an ITT approach, in which all participants were analyzed based on their allocated (i.e., prescribed) intervention. Participants were included if they had AHI ≥20, with *no* daytime hypersomnolence (ESS ≤10), and were 70 years of age or less. Participants were *excluded* if they had CVD or other chronic disease. Despite randomization, the median baseline AHI differed between the two arms (CPAP 42, no CPAP 35, P<0.1). The mean baseline ESS in both arms was 6.5. The two arms were similar in terms of age (mean: 51.9 years) and gender (males 85.6%). Most participants were obese (mean body mass index [BMI] 31.2 and mean neck circumference 42.2 cm). At baseline, 51.6 percent of the participants had HTN. Somewhat more CPAP users were current smokers than non-CPAP users (32% vs. 26%, P<0.1). The study was powered based on combined incident CVD event or HTN, which was not an outcome of interest for the current report. At 4 years, 64 percent of CPAP users were adherent (≥4 hours/night); median CPAP usage was 5.5 hours per night (IQR 2.2 to 6.3). The study was rated as high risk of bias due to failure to account for baseline differences in AHI (and time with oxygen saturation <90%) between groups and lack of participant or clinician blinding (although outcome assessors were blinded).

Huang 2015 randomly allocated 83 participants to CPAP (n = 42) and no CPAP (n = 41) for a median of 3 years.[102] Participants who were allocated to CPAP but had poor adherence (mean CPAP use <4 hours/night) were excluded from analysis. This applied to 9.5 percent (4/42) of participants randomized to CPAP. However, some additional results data were reported for some of the excluded participants (where feasible, we included these data). Overall, the study was analyzed on "as-treated" groups. Participants were included if they had both coronary artery disease (CAD) and uncontrolled HTN on treatment, and at least "moderate" OSA (AHI ≥15); all participants were 45 to 75 years of age, with a mean age of approximately 62 years. Most participants were overweight, with a mean BMI of 27.7 and a mean neck circumference of 41.1 cm. The study was not explicitly powered for any health outcome (but instead for a 5 mmHg drop in systolic blood pressure). The study was rated as moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded).

The **ISAACC** trial (Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome) randomly allocated participants with OSA (AHI ≥15) to CPAP treatment (n = 629) or usual care (n = 626) for a median of 3.4 years.[103] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP adherence. The trial also reported a propensity-score matched analysis of adherent CPAP users. Participants were adults who had been admitted to hospital for acute coronary syndrome who were found to have

OSA but had an ESS score of ≤10 (not excessive daytime sleepiness). The mean baseline AHI was similar between the two groups (CPAP 36.4, no CPAP 35.5), as were mean ESS (5.3), age (60 years), gender (males 84%), and BMI (29.5). Participants in the two groups had similar prevalence of preexisting diseases, including HTN (56%), DM (26%), and stroke (3%). Adherence was poor, with a mean of only 2.8 hours of use per night (median 2.2 hours); only 38 percent of those assigned to CPAP had good adherence (mean ≥4 hours/night). The study was powered for a composite outcomes of CV death or nonfatal acute myocardial infarction, stroke, hospitalization for CHF, unstable angina, or TIA. The study was rated as moderate risk of bias due to lack of participant or clinician blinding (with no mention of outcome assessor blinding).

A substudy of the **MOSAIC** (Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular) trial evaluated 188 participants who were randomly allocated to CPAP (n = 94) or no CPAP (n = 94) for 2 years.[107] The trial was conducted and analyzed with an ITT approach. The complete MOSAIC trial included 391 participants.[105] All 391 participants in the original trial were followed for 6 months while only 188 of them were followed for 2 years. Participants were included if they were 45 to 75 years of age, had confirmed OSA (ODI >7.5), and either no sleepiness symptoms (ESS<10) or sleepiness symptoms that were "not accompanied by patient concerns." The median ODI in the two arms were similar (CPAP 10.2, no CPAP 9.4) as was mean ESS (CPAP 7.9, no CPAP 8.0). Participants were on average 58 years old, 76 percent were male, and participants were mostly obese (mean BMI 32.4, mean neck circumference 43 cm). Participants in the two arms had similar prevalence in preexisting diseases, including MI (6%), HTN (77%), type 2 DM (16%), and tobacco use (12%). The study was powered for changes in blood pressure or cholesterol, not CV or other events. Adherence (CPAP use ≥4 hr/night) was relatively poor (across RCTs), at only 38 percent at 6 months (not reported at 2 years); although 71 percent expressed a wish to continue usage. Average usage was about 2.5 hours per night (including nonusers). No explanation was provided for the high dropout rate at 2 years (n = 188 vs. 391 at 6 months). The study was rated as high risk of bias due to lack of blinding and high dropout rate at 2-year followup (without assessment of potentially attrition bias).

The **PREDICT** trial (acronym undefined) randomly allocated participants to CPAP (n = 114) or no CPAP (n = 117) for 1 year.[109] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP adherence. Participants were included if they were 65 years or older with newly diagnosed OSA (ODI ≥7.5) with sleepiness symptoms (ESS ≥9). The median baseline AHI was similar between the two groups (CPAP 28.1, no CPAP 29.4), as were mean ESS (11.6), age (71.1 years) and gender (males 82%). Participants were obese with mean BMI of 33.7 and neck circumference of 43.3 cm. Participants in the two groups had similar prevalence of preexisting diseases, including CeVD (13%), ischemic heart disease (33%), CHF (7%), and HTN (73%). The trial was powered for a change in ESS score, and the authors noted that it might not be adequately powered for CV events and other outcomes. At 1 year, only 35 percent of participants were adherent (>4 hours/night) with CPAP. The study was rated as moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded).

The **RICCADSA** (Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA) trial randomly allocated 244 participants to CPAP (n = 122) or no CPAP (n = 122) for 4.75 years.[112] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP adherence. Participants were included if they had non-sleepy (ESS <10) OSA (AHI ≥15) and had a diagnosis of CAD with coronary revascularization. The mean baseline AHI was similar between the two arms (CPAP 28.3, no

CPAP 29.3), as was ESS (both 5.5). The two arms were similar in terms of age (mean 66 years) and gender (male 84%). Participants were generally overweight, with a mean BMI of 28.5. Participants in the two arms had similar prevalence in preexisting diseases, including CAD (100%), coronary revascularization (100%), MI (50%), HTN (64%), pulmonary disease (7%), type 2 DM (24%), obesity (28%), and current smoker status (16%). The study aimed to be powered for a composite outcome of CV mortality, acute MI, and new revascularization; however, the study authors noted that it was underpowered because CPAP adherence among the participants was lower than initially expected (60% at 1 year, implicitly defined as ≥4 hours/night). CPAP usage at 1 year averaged 5.8 (SD 1.7) hours per night. The study was rated as high risk of bias due to lack of blinding and high crossover (20% of participants in the no CPAP arm started CPAP during the study period).

The **SAVE** (Sleep Apnea cardioVascular Endpoints) trial randomly allocated 2687 participants to CPAP (n = 1346) or no CPAP (n = 1341) for 3.7 years.[12, 114-117] The trial was conducted and analyzed with an ITT approach; participants were not excluded from analysis based on CPAP adherence. The trial also reported a propensity-score matched analysis of adherent CPAP users. However, to be eligible, patients had to be adherent (≥3 hour/night) during a 1-week run-in trial; this criterion excluded about 16 percent of potential participants. Participants were included if they had a diagnosis of moderate-to-severe OSA and had a diagnosis of CAD or CeVD. The study reported that for their measure of ODI, a threshold of ≥12 was determined to be equivalent to AHI ≥30 (to diagnose "moderate-to-severe" OSA).[12, 114] Patients with very ESS scores (>15) were excluded. The study was designed to be enriched in patients at increased risk for CVD to increase power to detect an effect within the size and followup constraints of an RCT. As described in more detail below, in the section on composite CV outcomes (the primary outcome for this trial), the original planned sample size (5000 participants) was reduced during the study because of difficult recruitment, but better-than-expected CPAP adherence, allowing a revised planned sample size of 2500. The mean baseline AHI was similar between the two arms (CPAP 29.0, no CPAP 29.6), as was mean baseline ESS (CPAP 7.3, no CPAP 7.5). The two arms were similar in terms of age (mean 61 years), gender (male 81%), BMI (mean 28.6), and neck circumference (mean 40.7 cm). Participants in two arms had similar prevalence in preexisting diseases, including CAD (51%), MI (34%), CeVD (49%), stroke (44%), HTN (78%), type 2 DM (30%), and tobacco use (15%). The study was powered for a composite CV outcome (CV death, nonfatal MI, nonfatal stroke, and any hospitalization for unstable angina, CHF, or TIA). The study was not explicitly powered for specific CV outcomes or death, and the authors noted that it was not adequately powered for stroke or TIA. For the duration of the study, 42 percent were adherent with CPAP (≥4 hours/night), with mean usage of 3.3 (SD 2.3) hours per night. The study was rated as moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded). Of note, the study reported numerous analyses and sensitivity analyses. Their primary analysis was an unadjusted Cox proportional-hazards regression (survival) model. We primarily used these analyses (although for consistency with other RCTs, we used odds ratio [OR] results within meta-analyses). They also reported several sensitivity analyses using other analytic techniques. We also summarize their "as-treated" analyses of those adherent with CPAP therapy (≥4 hours/night for first 2 years). This analysis was conducted as a propensity-score matching one-to-one of adherent CPAP users (n = 561) and participants not assigned to CPAP (n = 561). Propensity-score matching was performed with baseline variables related to CPAP adherence and CV outcomes, including study site, type of CVD, ESS score (categorized at a threshold of 11), ODI,

prior adherence to sham CPAP (during the trial run-in phase), sex, age, obesity, smoking, DM, HTN, CHF, and others. The propensity score analysis applied appropriate methods; thus, the risk of bias was unchanged from the primary analysis.

Adjusted Nonrandomized Comparative Studies

No NRCS incorporated an *a priori* power calculation or made claims about their power to detect a given effect size for any outcome. All NRCSs included multivariable adjustments for potential confounders between groups (CPAP vs. no CPAP or adherence groups) and outcomes.

CPAP Versus No CPAP (NRCSs)

Chang 2020 was an adjusted NRCS comparing participants in the Taiwan's National Health Insurance Research Database newly diagnosed with OSA (by ICD-9 code) who received CPAP in the first year following the diagnosis (n = 959) with a propensity-score matched cohort of those who did not (n = 959).[136] The study conducted an "as-prescribed" analysis (comparing those prescribed CPAP vs. those not prescribed CPAP), conceptually similar to an ITT analysis. Patients were followed for 5 years. The majority of the participants were male (83%). The two groups were comparable in terms of age distribution (52% were 40 to 59 years of age) and preexisting diseases (HTN 56%, type 2 DM 16%, CVD 34%). No information on baseline AHI, ESS, or BMI were provided. The study had no information on adherence. The study was rated as high risk of bias as an (unblinded) nonrandomized study that implemented propensity score matching, but provided insufficient descriptive information.

Jennum 2015 was an adjusted NRCS comparing participants in the Danish National Patient Registry diagnosed with OSA (by ICD-10 code) who received CPAP (n = 13,631) within 2 years of diagnosis with those who did not (n = 11,758).[125] The study conducted an "as-prescribed" analysis (comparing those prescribed CPAP vs. those not prescribed CPAP). Patients were followed for 3 years. The majority of the participants were male (78%). The two arms were comparable in terms of age distribution (55% were 40 to 59 years of age). No information on baseline AHI, ESS, BMI, or preexisting diseases were provided. The study had no information on adherence. The study was rated as high risk of bias related to inadequate reporting regarding OSA diagnosis, intervention, with the use of a regression model for adjustment of baseline confounders.

Lisan 2019 was an adjusted NRCS comparing participants from the Sleep Heart Health Study (**SHHS**) who were prescribed CPAP (n = 51) with those not prescribed CPAP (n = 204) for 11 years.[126] The study conducted an "as-prescribed" analysis (comparing those prescribed CPAP vs. those not prescribed CPAP). Participants were included if they were over 40 years and able and willing to undergo a home polysomnography. The definition of OSA or criteria for use of CPAP were not reported. The study conducted a propensity score-matched analysis, matching one CPAP participant to four non-CPAP participants. The two groups were similar in terms of pretreatment AHI (mean 29.5), respiratory disturbance index (RDI), (25.8), ESS (mean 11.1), age (mean 64 years), gender (male 76%), BMI (mean 31.3), and neck circumference (mean 41.1 cm). Participants in the two arms also had similar prevalence in preexisting conditions, including MI (9%), HTN (45%), type 2 DM (15%), and current tobacco use (6%), among others. The study had no information on adherence. The study was rated as moderate risk of bias based on being an (unblinded) nonrandomized study that used a propensity score analysis and had generally complete reporting; although adherence was unknown.

Myllylä 2019 was an adjusted NRCS evaluating patients who had been prescribed CPAP for OSA.[128] The study compared participants who continued to use CPAP for at least 5 years (n = 1030) with those who discontinued using CPAP against their doctor's advice (n = 1030). The study followed participants for a median of 8.7 years. Thus, this study conducted based on actual use of CPAP. Participants were included if they commenced CPAP for at least "moderately" severe OSA (AHI ≥15). The two groups were similar in AHI (mean 27.5) but different in ESS (CPAP 9.4, no CPAP 8.3). The two groups were similar in terms of mean age (56.0 years) and gender (males 76%). The mean BMI was higher in the CPAP arm than no CPAP arm (32.7 vs. 31.5), as were the prevalence of several preexisting conditions, including type 2 DM or impaired fasting glucose (CPAP 40%, no CPAP 35%), HTN: (CPAP 77%, no CPAP 70%). Participants in the CPAP group had lower rates of other preexisting conditions, including CVD (CPAP 5.4%, no CPAP 13%), current tobacco use (CPAP 22%, no CPAP 29%), and COPD (CPAP 4.6%, no CPAP 7.6%). Among CPAP users, 89 percent used CPAP for at least 4 hours per night (overall median 6.4 hours/night). The study was rated as high risk of bias related to inadequate reporting regarding OSA diagnosis, intervention, with the use of a regression model for adjustment of baseline confounders.

Nakamura 2021 reported an adjusted NRCS, based on Okinawa Nakamura Sleep (**ONSLEEP**) registry, comparing participants who started CPAP (n = 1274) with a propensity-score matched cohort those who did not (n = 1274) at a median of 6.6 (interquartile range 2 to 10.7) years of followup.[131] The study defined CPAP users as those who agreed to use CPAP (and used it for at least 30 days), regardless of adherence; CPAP nonusers were either not prescribed CPAP or did not commence CPAP use despite a prescription for use. We categorized this as an "as-prescribed" analysis. Participants were included if they were diagnosed with OSA (AHI>5). The propensity score analysis included only 28 percent of CPAP users with available data and 60 percent of nonusers. Analyzed patients were 76 percent male and on average 52 years of age with a mean BMI of 27.3. Among nonusers 14 percent had DM (vs. 8% of CPAP users) and 6 percent had a history of stroke (vs. 4%); these were considered to be significant differences between groups. There were 12 percent with a history of heart disease. Median AHI was slightly higher among CPAP users (median 23.9 vs. 18.4), as was percent of sleep time with oxygen saturation below 90 percent (2.3% vs. 1.0%). The percent of patients with ESS >10 was similar in both groups (55%). Adherence data were not available in the large majority (84%) of CPAP users. The study was rated to be at moderate risk of bias for all-cause death in an (unblinded) nonrandomized study that implemented propensity score matching adjusting for baseline imbalances. The CV death analysis was deemed to be high risk of bias since reported propensity-score matched analyses were unadjusted.

Ou 2015 was an adjusted NRCS that compared participants who used CPAP (n = 36) with those not using CPAP (n = 88) for 5 years.[132] We categorized the study as an "as-prescribed" analysis, conceptually similar to a quasi-ITT in that all patients were offered CPAP and the analysis compared those who agreed to start CPAP (regardless of adherence) with those who refused. Participants were included if they were older adults (≥60 years) with moderate to severe sleep apnea (AHI≥20). Both mean baseline AHI and ESS were significantly different between the two arms (AHI: CPAP 45.3, no CPAP 36.0, P = 0.001; ESS: CPAP 8.1, no CPAP 6.2, P = 0.007). The study also reported a mean ODI of 34.3, with no differences between two arms. The two arms were similar in terms of age (mean 73 years), gender (male 84%), and BMI (mean 30.9). Among those prescribed CPAP, 67 percent were adherent (4 hours/night, ≥5 nights/week). Participants in the two arms had similar prevalence in preexisting diseases, including CAD

(48%), HTN (68%), type 2 DM (16%), and stroke (13%). Notably, several of the baseline characteristics were not adjusted for in the multivariable analysis although differences existed between the two arms, including ESS, habitual snoring, and chronic insomnia. The study included too few participants to allow complete adjustment for potential confounders. The study was rated as high risk of bias primarily related to inadequate adjustment for differences between groups.

Adherent CPAP Use Versus Nonadherent CPAP Use (NRCSs)

Campos-Rodriguez 2005 was an adjusted single center based NRCS evaluating OSA patients using PAP (either CPAP or bilevel pressure ventilation [BPAP, about 9%]), based on their level of adherence.[123] Nonusers were excluded from analysis; comparisons were between those with good adherence: >6 hours per night (n = 322), 1 to 6 hours per night (n = 342), and <1 hour per night (n = 85). Patients were followed for a mean of 48.5 months (range 0 to 103 months). Participants were included if they were adults with OSA with an AHI ≥10. The subgroups of patients differed from each other for many breathing and sleep measures (oxygen and carbon dioxide partial pressures, spirometry measures, percent of time below 90% saturation, and AHI), but multivariable adjustment included only variables that were statistically significant predictors of death within the model (HTN, age, and spirometry). Mean AHI in the good adherence group was 60.0, the adherence group 52.1, and poor adherence group 48.8. Across groups, mean age was about 55 years, about 80 percent were male, and mean BMI was 35.6. Most patients were obese (83%), had HTN (60%), and were smokers (66%); many had DM (35%) chronic obstructive pulmonary disease (17%). The study was rated as high risk of bias, primarily based on their use of a regression model with incomplete adjustment of baseline confounders.

López-Padilla 2016 reported an adjusted NRCS of older adults (≥80 years) with OSA.[127] The study conducted an analysis of adherent CPAP users (≥4 hours/night, n = 79) versus a combination of nonadherent patients (n = 53) and those not prescribed CPAP (n = 23). Median followup (CPAP treatment) was 53 months (interquartile range 41 to 77 months). Participants were included if they were older and had at least "moderately" severe OSA (AHI ≥20). Patients with chronic respiratory failure were excluded. The mean AHI was 49.2, mean ODI 39.8, and mean ESS 12.8. The participants were about half men (54%) and were on average obese (BMI 32.8). Comorbidities were common, including HTN (78%), CHF (35%), tobacco use (35%), DM (30%), hyperlipidemia (28%), arrhythmias (26%), ischemic heart disease and stroke (15% each). The study was rated as high risk of bias for being an (unblinded) nonrandomized study that used a regression model for adjustment of baseline confounders.

Navarro-Soriano 2021 was a secondary analysis from an existing RCT (HIPARCO study), which had randomly allocated participants to CPAP and no CPAP.[129] Included patients had at least moderately severe OSA (AHI ≥15) and resistant HTN. Patients with follow-up data on the incidence of fatal or non-fatal CV events were included in this observational analysis. This secondary NRCS compared adherent CPAP use (>4 hour/night, n = 95) with combined nonadherent (n = 39) and those not offered CPAP (n = 29). Participants were followed for a median of 58 months. Patients in the CPAP group had significantly higher BMI than nonusers (34.6 vs. 32.6). Other background characteristics were similar between groups, including age (mean 58), AHI (mean 40), ESS (mean 9.2), gender (male 68%), prior CV events (21%) and type 2 DM (40%). The study was rated as high risk of bias for being an (unblinded) nonrandomized study that used a regression model for adjustment of baseline confounders.

Schipper 2017 was an adjusted NRCS comparing participants who were adherent with CPAP (n = 162) for at least 4 hours per night for more than 70 percent of nights with nonadherent CPAP users (n = 121) for 6 years.[133] Thus, this study conducted an analysis based on actual use of CPAP. Participants were included if they were newly diagnosed OSA (AHI ≥5). The two arms were similar in the baseline AHI (median 25.0), mean age (54.2 years), gender (male 77%), and mean BMI (30.0). Participants in the two arms had similar prevalence of preexisting diseases, including HTN (60%) and CVD or CeVD (15%). However, the prevalence of other preexisting diseases was significantly different between the two arms, including type 2 DM (CPAP 12%, no CPAP 24%), tobacco use (CPAP 15%, no CPAP 27%). The study was rated as high risk of bias for being an (unblinded) nonrandomized study that used a regression model for adjustment of baseline confounders.

Sheth 2021 was an adjusted NRCS that compared patients with type 2 DM prescribed CPAP for OSA (AHI ≥5) who were CPAP adherent (usage for ≥4 h/night for ≥70% of nights; n = 260) with those who were CPAP nonadherent (n = 318).[134] The study followed patients for 2.5 years. There were no significant differences between those who were CPAP adherent and non-adherent in terms of age (54 years), gender (male 46%), BMI (42), preexisting diseases (CVD 30%, peripheral vascular disease 13%, stroke 13%), or AHI (45.8). The study was rated as high risk of bias for being an (unblinded) nonrandomized study that used a regression model for adjustment of baseline confounders. The study also did not adequately report results, particularly in terms of analyzed participants.

Wu 2015 (Beijing) was an adjusted NRCS that evaluated participants with moderate to severe OSA (AHI ≥15) and who were undergoing percutaneous coronary intervention for CAD.[135] The study compared those who used and were adherent with autoCPAP (≥4 hours/night, ≥70% of nights, >3 months; n = 128) with those who refused (n = 79) or were nonadherent with CPAP (n = 88). Patients were followed for a median of 4.8 years. The median baseline AHI was different between the two arms (CPAP 46.3, no CPAP 40.1), but the two arms were similar in terms of age (mean 55.1 years), gender (males 84%), and BMI (mean 29.7). Participants in the two arms had similar prevalence in preexisting diseases, including previous MI (14%), non-ST-segment elevation acute coronary syndrome (52%), HTN (74%), type 2 DM (34%), current tobacco use (25%), stroke (10%), and chronic kidney disease (4.4%). The study was rated as high risk of bias for being an (unblinded) nonrandomized study that used a regression model for adjustment of baseline confounders.

All-Cause Mortality

Eleven studies (5 RCTs and 6 NRCSs) comparing CPAP prescription or use and no CPAP reported on all-cause mortality (Table 5).[12, 103, 107, 112, 123, 125-127, 131, 132, 137, 138] Studies evaluated a wide range of followup times (from 2 to 11 years). The studies included patients with OSA but with different diagnostic methods, including specific AHI thresholds or database classification (e.g., ICD 9 or 10 codes). All five RCTs reported ITT analyses and three of the NRCSs reported "as-prescribed" analyses comparing patients provided CPAP (who started using CPAP) and those not provided CPAP. One of the RCTs and three of the NRCSs compared adherent users and nonusers.

Randomized Controlled Trials (All-Cause Mortality)

The five RCTs evaluated variable populations of patients at 2 to about 5 years of followup. Barbé 2012 and the MOSAIC trial included otherwise healthy, somewhat younger adults (≤70

years in Barbé 2012, 45 to 75 years in MOSAIC) with at least "moderate" OSA (AHI ≥20 or ODI ≥7.5, respectively). Barbé 2012 explicitly excluded patients with CVD or other chronic diseases and those with excessive daytime sleepiness (ESS >10). In MOSAIC, few patients had prior CVD, but the trial included only patients with sleepiness ESS ≥10. The ISAACC and RICCADSA trials included high-risk populations, namely those who had been admitted to hospital for acute coronary syndrome (ISAACC) or recent coronary revascularization (RICCADSA). Both required at least "moderate" OSA (AHI ≥15). The SAVE trial included adults with CVD or CeVD with "severe" OSA (ODI ≥12, which the researchers had determined to be equivalent to AHI ≥30). None of the trials was explicitly powered for all-cause mortality. All were analyzed as ITT for CPAP versus no CPAP prescription. Adherence (use of CPAP for ≥4 hours per night) varied across studies at about 40 percent in the ISAAC, MOSAIC, and SAVE trials and about 60 percent in Barbé 2012 and the RICCADSA trial (see Table 5). SAVE and ISAACC also reported analyses of adherent users.

In all RCTs, patients and clinicians were not blinded, but the SAVE and ISAACC trials reported that outcome assessors were blinded. These two trials were, thus, rated to be at moderate risk of bias, but the other trials were further downgraded to high risk of bias due to high dropout rate (MOSAIC), high crossover rate (RICCADSA), and failure to account for baseline differences in sleep study measures.

Across RCTs, effect sizes (odds ratio [OR], hazard ratio [HR], or incidence density ratio) ranged from 0.33 to 2.6, all statistically nonsignificant. The number-needed-to-treat (NNT) or harm (NNH) to prevent (or add) one death ranged from a NNT of 47 to a NNH of 70 across studies, which corresponds to the range of effect sizes across studies. The summary OR across the RCTs was nonsignificant at 0.89 (95% confidence interval [CI] 0.66 to 1.21) (Figure 2). Under an assumption of a control rate (risk of death without CPAP use) of 3.5 percent (the meta-analyzed control rate across RCTs), the summary NNT was 186 (95% CI NNH 142 to NNT 86). The effect sizes did not clearly correlate with adherence rates in the ITT analyses.

In addition to their primary ITT analyses, SAVE and ISAACC reported propensity-score matched analyses of adherent CPAP users, with a nonsignificant, but somewhat stronger, HR in SAVE: 0.60 (95% CI 0.32 to 1.10), but near identical HR in ISAACC: 0.80 (95% CI 0.52 to 1.23).

Adjusted Nonrandomized Comparative Studies (All-Cause Mortality)

Across studies, all-cause mortality was reported at a wide range of followup times from 3 to 11 years, overlapping with the RCTs but with mostly longer-term followup. The NRCSs mostly compared adults on CPAP versus no CPAP regardless of AHI/ODI threshold or comorbidities; however, Ou 2015 and López-Padilla 2016 were restricted to older adults (≥60 years or ≥80 years, respectively) with AHI ≥20. The event rates (percent who died) in the NRCSs tended to be higher than in the RCTs, particularly among the two noted studies of older adults and the study with longest term followup (Lisan 2019, 11 years). One study compared users of CPAP versus nonusers and did not report on adherence (defined by nightly use);[127] a second study compared different groups of CPAP users based on their nightly adherence (overall 43% used CPAP >6 hours/night).[131] Among the four "as-prescribed" analysis NRCSs,[125, 126, 131, 132] three did not report data on adherence, but Ou 2015 reported a 67 percent adherence rate.[132]

The six NRCSs were adjusted for potential confounders or patient characteristics in propensity score analyses or logistic regression. The specific adjustments made are presented in the footnotes to Table 5. The studies mostly adjusted for age, but only two NRCSs (Jennum 2015

and ONSLEEP) adjusted for sex. All but one study adjusted for either CVD or risk factors (in particular DM, but also BMI, HTN, and smoking); Jennum 2015 adjusted for the Charlson Comorbidity Index. Three NRCSs (Campos-Rodriguez 2005, ONSLEEP, Ou 2015) adjusted for AHI, ESS, or pulmonary function test readings.) Campos-Rodriguez 2005 further adjusted for CPAP adherence.

NRCSs also evaluated two different categories of analyses. Four NRCSs (Jennum 2015, Lisan 2019, ONSLEEP, and Ou 2015) evaluated "as-prescribed" analyses of people provided CPAP devices.[125, 126, 131, 132] Two studies (Campos-Rodriguez 2005 and López-Padilla 2016) evaluated analyses of adherent CPAP users.[123, 127]

Two of the NRCSs that conducted "as-prescribed" analyses (Lisan 2019 and ONSLEEP) used adequate propensity score analyses and were, thus, at moderate risk of bias; although, neither study reported on adherence. (Since we downgraded RCTs for issues related to randomization and blinding, to ensure consistency across all studies, we deemed that NRCSs could at best be at moderate risk of bias.)

CPAP Versus No CPAP

The four adjusted NRCSs that reported "as-prescribed" analyses all reported statistically significant associations between CPAP prescription and reduced risk of all-cause mortality. Adjusted HRs ranged from 0.08 (in a small, poorly analyzed study) to 0.67 (in a large database analysis); the two propensity-score analyses found intermediate adjusted HRs. The summary adjusted HR across the NRCSs was statistically significant at 0.55 (95% CI 0.42 to 0.74), with moderate heterogeneity (I^2=55%) (Figure 2). It can be noted that, like all studies included in meta-analyses of adjusted NRCSs, each NRCS used a unique method to adjust for potential confounders.

CPAP Users Versus CPAP Nonusers

Two studies that evaluated adherent CPAP use. Campos-Rodriguez 2005 compared very adherent users (>6 hr/night), moderately adherent users (1-6 hr/night), and very nonadherent users (<1 hr/night).[123] López-Padilla 2016 evaluated a comparison of adherent users versus a combination of nonadherent users and those not prescribed CPAP. Both studies reported statistically significant associations, without establishing causality, between adherent CPAP use and lower risk of all-cause mortality (Table 5, Figure 2). Neither study attempted to account for the inherent, underlying differences between adherent and nonadherent individuals. The findings were generally consistent with the "as-prescribed" analysis NRCSs, but with wider confidence intervals. Notably, Campos-Rodriguez compared three cohorts of patients based on the number of hours per night they used their devices, finding a trend such that high levels of use (>6 hours/night, mean 7.6) had a stronger effect (adjusted OR 0.10) than moderate (1 to 6 hours/night, mean 3.9) levels of use (adjusted OR 0.28); although death rates were similar between the two groups (3.4% vs. 4.6%, respectively).

All Studies (All-Cause Mortality)

The summary estimate across all studies (RCTs and NRCSs) was 0.61 (95% CI 0.49 to 0.76), with moderate statistical heterogeneity (Figure 2). Separately analyzing the ITT or "as-prescribed" analyses and the adherent analyses (across RCTs and NRCSs), we found generally similar summary associations. The ITT/"as-prescribed" analyses (all five RCTs and four of the NRCSs, respectively) yielded a summary effect size of 0.66 (95% CI 0.53 to 0.81) with a small to moderate degree of statistical heterogeneity; although, effect size estimates ranged from 0.08

(Ou 2015 NRCS, statistically significant) to 2.6 (Barbé 2012, RCT, nonsignificant). The CPAP adherent analyses (including two *post hoc* analyses from the RCTs and two adjusted NRCSs) yield a summary effect size of 0.55 (95% CI 0.38 to 0.81) with moderate heterogeneity and a range of HRs of 0.28 (Campos-Rodriguez 2005, NRCS, statistically significant) to 0.80 (ISAACC, RCT, nonsignificant). There was no evidence of a difference in effect size between the two sets of analyses (P = 0.41).

RCTs Versus NRCSs (All-Cause Mortality)

The summary effect size for the NRCSs was statistically significant while the summary RCT estimate was not. Meta-regression between RCTs and (all) NRCSs suggests that the average effect size among the NRCSs was 58 percent smaller (stronger, favoring CPAP) than the average RCT (P = 0.061 between study designs).

An exploratory sensitivity analysis evaluating the change in results as the meta-analysis borrows strength from the NRCS evidence is shown in Appendix D, Figure D-2. The figure recapitulates that with no NRCS evidence (only RCT evidence), the summary OR is nonsignificant and imprecise. However, adding in even a small degree of weight (<1%) from the NRCSs yields a statistically significant effect size similar to the estimate based on full inclusion of the NRCS information, suggesting a significant effect (or association) of CPAP use on reducing the risk of all-cause mortality.

The reasons for the differences between study designs is not inherently clear. The difference in statistical significance may be largely related to power issues among the RCTs, but would not account for the difference in average effect sizes. Study event rates (percent who died) tended to be higher in the NRCSs than the RCTs, related to both longer followup duration for most NRCSs and higher risk of death due to older age in at least two NRCSs. Another important difference between study designs is that RCTs, in contrast to the NRCSs, mostly excluded patients with excessive daytime sleepiness. While the RCTs may have been more rigorous in design than the NRCSs and less subject to reporting biases, the NRCSs may better represent real world experiences.

Heterogeneity of Treatment Effect (All-Cause Mortality)

Note that a discussion about variations in estimates of effect or association across studies based on different criteria used to define apnea, hypopnea, and oxygen desaturation, and different AHI thresholds appears in its own section ("*Evaluation of Heterogeneity, Across Outcomes, Based on AHI Definitions Used*") after the summaries of all outcomes.

Three NRCSs compared incidence of all-cause mortality among subgroups of participants. Comparisons included AHI (5-29 vs. ≥30), age, BMI (<25, 25-29, ≥30), and pulmonary function status (impaired vs. normal). It is unclear whether the NRCSs were adequately powered for the subgroup analyses; although Jennum 2015 was a large study. Note that none of the RCTs reported on subgroup effects, except as implied by adherent-user analyses.

CPAP Adherence

As noted above, the two RCTs that reported secondary analyses of adherent users (≥4 hours/night) found statistically similar results as their primary ITT analyses (ISAACC: HR 0.80 vs. 0.82, with similar statistical significance; SAVE: HR 0.60 vs. 0.91, nonsignificant but smaller P-value in adherent users analysis). Also as noted, Campos-Rodriguez 2005 found an apparent trend in adjusted ORs compared with minimal device use (<1 hour/night) between moderate

users (1 to 6 hour/night, mean 4.6; OR 0.28) and high users (≥6 hours/night, mean 7.6; OR 0.10); however, there was no significant difference between the adherent groups.

Across studies, there was no consistent association between mean CPAP use (adherence) and risk of death (compared with no use). Three RCTs had reported adherence rates of about 40 percent, with effect sizes ranging from 0.33 to 0.82); two had adherence rates of about 60 percent with effects sizes of 0.76 and 2.6. Differences across all studies that reported adherence rates related more to study design (RCT vs. NRCS) than reported adherence.

Sleepiness (Epworth Sleepiness Score)

None of the studies compared patients based on their degree of daytime sleepiness. Across studies, there was no apparent relationship among studies that excluded patients with (or without) excessive daytime sleepiness (ESS ≥10), or included patients regardless of ESS. However, the RCTs generally excluded patients with ESS ≥10 (MOSAIC allowed some patients with sleepiness, but mean ESS was 8; SAVE excluded only patients with ESS <15, but mean ESS was 7.4). Among NRCSs that restricted based on ESS or reported average ESS, all except Ou 2015 either excluded patients with low ESS (Campos-Rodriguez 2005) or had average scores >10 (Lisan 2019, ONSLEEP, López-Padilla 2016). Only Ou 2015 had a mean ESS <10 (6.7). Among NRCSs, Ou 2015 had the largest effect size (although with a wide confidence interval; see Figure 2). While studies of patients with excessive daytime sleepiness (all of which were NRCSs) found significant associations between CPAP use and lower risk of death, the studies that excluded patients with excessive daytime sleepiness (all of which were RCTs) did not find such a causal relationship. None of the studies, however, evaluated heterogeneity of treatment effect based on the presence of excessive daytime sleepiness. Thus, overall, across RCTs and adjusted NRCSs, the relationship between pre-treatment ESS and the effect of CPAP on all-cause mortality is unclear. One cannot conclude that CPAP is more effective in patients with excess daytime sleepiness.

AHI Subgroups

An earlier publication of the ONSLEEP NRCS reported adjusted HRs among different subgroups, including by AHI.[130] Among participants with AHI between 5 and 29, the adjusted HR for all-cause mortality was 0.94 (95% CI 0.47 to 1.86). Based on reported data, we were able to calculate that for those with AHI more than 30, the adjusted HR for all-cause mortality was 0.48 (95% CI not reported). However, the study did not report on the statistical significance of an interaction between AHI category and CPAP use.

Age Subgroups

The Jennum 2015 NRCS found similar, not statistically significantly different, HRs among different age groups: 0.72 (95% CI 0.39 to 1.33) for patients aged 20-39 years, 0.67 (95% CI 0.56 to 0.79) for patients aged 40-59 years, and 0.62 (95% CI 0.54 to 0.71) for patients aged ≥60 years.[125] We calculated these HRs based on their report of a model that included interaction terms between CPAP use and age group.

Gender Subgroups

The Jennum 2015 NRCS reported that "the interaction term for gender and CPAP treatment was significant, indicating that women had higher [relative] mortality than men when they received CPAP treatment," with a P value for the interaction term of 0.036, suggesting a weaker association between CPAP use and death in women than in men.[125] However, the study appears

to have included interaction terms for both age and gender in the same model, but did not translate these to separate HR estimates for men and women (as they did for age subgroups). Thus, we cannot estimate the different associations in the two genders.

Body Weight Subgroups

The ONSLEEP NRCS also reported analyses by BMI category. The adjusted HR for all-cause mortality among participants with BMI <25 was 0.39 (95% CI 0.18 to 0.83), for those with BMI 25-29 0.66 (95% CI not reported), and for those with BMI ≥30 2.27 (95% CI not reported). Statistical significance between groups was not reported.[130]

Pulmonary Function Subgroups

The ONSLEEP NRCS also reported data to suggest that the adjusted HR of all-cause mortality with CPAP among participants with impaired pulmonary function (based on pulmonary function testing) was 0.39 (95% CI not reported), while among those with normal pulmonary function HR was 1.00 (95% 0.48 to 2.09). Statistical significance between groups was not reported.[130]

Applicability (All-Cause Mortality)

The RCTs are mostly applicable to people with moderate or severe OSA (by AHI or ODI criteria) who have known CVD and did not have excessive daytime sleepiness (ESS <10); only the smallest of the three RCTs, MOSAIC, included participants without CVD. In contrast, the NRCSs mostly included all adults with OSA (variably defined), with a wide range of baseline AHI, but mostly undescribed comorbidities. The NRCSs mostly included patients with or without excessive daytime sleepiness, but average ESS scores were mostly >10.

In terms of applicability to the Medicare population, two NRCSs (Ou 2015 and López-Padilla 2016) were restricted to older adults (≥65 or ≥80 years) with no statistically significant difference in results compared with other studies, except these studies had higher death rates, yielding greater power and statistically significant effect sizes.

Summary of Effect of CPAP on All-Cause Mortality

The RCTs and NRCSs yielded different conclusions regarding the effect (or association) of CPAP on all-cause mortality (Table 4). Five RCTs, of patients with moderate or severe OSA, who were mostly at increased risk of CV death but without excessive daytime sleepiness, did not provide evidence that CPAP reduces the risk of all-cause mortality. This conclusion does not imply that CPAP has been proven to be ineffective to reduce all-cause mortality. In addition, the finding was deemed to have low strength of evidence (SoE) because of moderate risk of bias among the studies, indirectness due to concerns about the applicability of the studies to typical adults with OSA (related to risk factors and sleepiness symptoms), relatively short duration of followup, and imprecision (summary effect size 0.89, 95% CI 0.66 to 1.21).

The NRCSs, with mostly included all patients with OSA, with or without excessive daytime sleepiness, in contrast, all found statistically significant associations between CPAP and lower risk of all-cause mortality. This was the case for analyses of both CPAP prescription and adherent CPAP use. Most NRCSs were at high risk of bias, but the two moderate risk of bias NRCSs (that reported propensity score analyses) had consistent findings as the other NRCSs.

Meta-analyses of all ITT or "as-prescribed" analyses (of patients prescribed CPAP) and of all analyses of patients adherent to CPAP yielded similar summary effect sizes, favoring CPAP: ITT/"as-prescribed" 0.55 (95% CI 0.42 to 0.74); adherent 0.41 (0.26 to 0.64).

Across both RCTs and NRCSs, we concluded that that CPAP may reduce the risk of all-cause mortality (low SoE). Studies, overall, had moderate risk of bias (and most NRCSs had high risk of bias) and the RCTs and NRCSs were inconsistent with each other in that the RCTs found a nonsignificant causal effect, but NRCSs found larger, statistically significant associations. The explanation for the difference between the RCTs and NRCSs is unclear. The RCTs mostly included higher risk patients at higher risk of death, but still had relatively low death rates over about 2 to 5 years of followup. The NRCSs mostly included all adults with OSA, had mostly longer followup (about 3 to 11 years), and mostly higher rates of death. The studies also included somewhat different populations in terms of symptoms. The RCTs mostly excluded patients with excessive daytime sleepiness, but half (or more) of patients in the NRCSs had high ESS (>10). While none of the studies could effectively blind patients or their clinicians to treatment, the RCTs mostly followed strict protocols, including randomization, and fully reported results, thus, reducing bias. Two of the NRCSs reported appropriately conducted propensity score analyses (which may simulate randomization), but the degree to which nonsignificant NRCSs may be underreported is unclear. However, the NRCSs may be reporting more "real world" experiences with CPAP. Furthermore, we cannot rule out that the apparent difference in findings between RCTs and NRCSs relates to insufficient power among the RCTs; although this would not fully account for the difference in summary effect estimates.

Thus, the RCTs do not provide evidence that CPAP affects the risk of all-cause mortality (summary ES 0.89, 95% CI 0.66 to 1.21, low SoE). The RCTs are at moderate risk of bias, provide an imprecise estimate of effect, mostly excluded patients with excessive daytime sleepiness, and may have been too short in duration to provide adequate power. We also note that even though the RCTs were statistically nonsignificant, across comparative studies, all but one RCT had effect sizes <1 (i.e., appeared to favor CPAP).

The NRCSs, in contrast, found statistically significant associations between CPAP prescription (or use) and lower risk of all-cause mortality. Taking the RCT effects and NRCS associations together, we conclude that, overall, the comparative studies find that CPAP (use) may be associated with reduced risk of death (low SoE). However, RCTs and NRCSs were inconsistent in terms of statistical significance and magnitude of effect size, and studies were at moderate to high risk of bias.

The studies provided insufficient evidence to support whether effect sizes may differ in different subgroups of patients, including by severity of OSA (as assessed by AHI or similar measures), severity of symptoms (e.g., as measured by ESS), age, gender, other risk factors or comorbidities. The comparative studies provide inconsistent evidence regarding the interplay between CPAP adherence and association of CPAP use and risk of death: RCTs did not find a difference in effect based on CPAP adherence, summary estimates of ITT and adherent analyses were similar, but NRCSs that analyzed cohorts based on adherence did find significant differences favoring more adherent users. The RCT evidence is mostly applicable to patients with CV risk factors, but without excessive daytime sleepiness. The NRCS evidence may pertain to the effect of longer duration treatment and may be more applicable to typical adults with OSA who qualify for possible CPAP use.

The low SoE for the conclusions regarding all-cause mortality suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding that the findings are stable or that the estimate of effect is close to the true effect.

Figure 2. Meta-analysis of CPAP versus No CPAP: All-cause mortality

Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Effect Size (95% CI)
Barbé 2012	48	8/357	3/366	IDR 2.6 (0.70, 11.8)
ISAACC 2020*	40	27/629	32/626	HR 0.82 (0.49, 1.36)
MOSAIC 2014	24	1/94	3/94	OR 0.33 (0.03, 3.19)
RICCADSA 2016	57	7/122	9/122	OR 0.76 (0.28, 2.12)
SAVE 2016*	44	40/1346	43/1341	HR 0.91 (0.59, 1.40)
RCT (all ITT)		**83/2543**	**90/2549**	**0.89 (0.66, 1.21) (I²=0%)**
Jennum 2015	36	NR/13,631	NR/11,758	adjHR 0.67 (0.61, 0.75)
Lisan 2019	133	12/81	84/311	adjHR 0.58 (0.35, 0.96)
ONSLEEP 2021	79	53/1274	94/1274	adjHR 0.47 (0.33, 0.66)
Ou 2015	60	2/36	19/88	adjHR 0.08 (0.01, 0.44)
NRCS "as-prescribed"		**NR/15,022**	**NR/13,431**	**0.55 (0.42, 0.74) (I²=55%)**
Campos-Rodriguez 2005	48	16/342†	8/85†	adjOR 0.28 (0.11, 0.69)†
López-Padilla 2016	53	35/79	48/76	adjHR 0.46 (0.27, 0.78)
NRCS Users v. Nonusers		**51/421**	**56/161**	**0.41 (0.26, 0.64) (I² N/A)**
All ITT or "as-prescribed"*				**0.66 (0.53, 0.81) (I²=36%)**
All Users v. Nonusers*				**0.55 (0.38, 0.81) (I²=41%)**
OVERALL				**0.61 (0.49, 0.76) (I²=45%)**

Favors CPAP   Favors No CPAP

0.05   0.1   0.3   0.5   1.0   2.0   3.0
Effect Size (log scale)

Abbreviations: adj = adjusted, CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, HR = adjusted hazard ratio, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, IDR = incidence density ratio, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, ITT = intention-to-treat, MOSAIC = Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular trial, NR = not reported, NRCS = nonrandomized comparative study, ONSLEEP = Okinawa Nakamura Sleep, OR = odds ratio, RCT = randomized controlled trial, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

* SAVE and ISAACC reported both ITT and CPAP adherent (users) analyses (see Table 5). Both the all ITT or "as-prescribed" summary estimate and the users versus nonusers summary estimate include these trials.
† PAP use 1-6 hours per night versus <1 hour per night.

Table 5. CPAP versus no CPAP: All-cause mortality

Study PMID	Design	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
Barbé 2012 22618923[99]	RCT	ITT (No)	48	CPAP No CPAP	64% overall	8/357 (2.2) 3/366 (0.8)	Incidence density ratio 2.6 (0.70, 11.8) NNH 70 (NNH 31, NNT 270)
ISAACC 31839558[103]	RCT	ITT (No)	40	CPAP No CPAP	38% overall	27/629 (4.3) 32/626 (5.1)	HR 0.82 (0.49, 1.36) NNT 122 (NNH 66, NNT 32) [Adherent analysis: HR 0·80 (0.52, 1.23)
MOSAIC 24508706[107]	RCT	ITT (No)	24	CPAP No CPAP	38% at 6 mo	1/94 (1.1) 3/94 (3.2)	OR 0.33 (0.03, 3.19) [C] NNT 47 (NNH 50, NNT 16)
RICCADSA 26914592[112]	RCT	ITT	57	CPAP No CPAP	60% at 1 y	7/122 (5.7) 9/122 (7.4)	OR 0.76 (0.28, 2.12) [C] NNT 61 (NNH 22, NNT 13)
SAVE 27571048[12]	RCT	ITT (No)	44	CPAP No CPAP	42% overall	40/1346 (3.0) 43/1341 (3.2)	HR 0.91 (0.59, 1.40) NNT 426 (NNH 93, NNT 65) [Adherent analysis: HR 0.60 (0.32, 1.10)]
Campos-Rodriguez 2005[123]	NRCS	Adherence analysis (No)	48.5	PAP use >6 hr/night	Mean 7.6 hr/night	11/322 (3.4)	**Adjusted OR 0.10 (0.04, 0.29) [I]** vs. <1 hr/night
.	.	.	.	PAP use 1-6 hr/night	Mean 3.9 hr/night	16/342 (4.6)	**Adjusted OR 0.28 (0.11, 0.69) [I, J]** vs. <1 hr/night
.	.	.	.	PAP use <1 hr/night	Mean 0.3 hr/night	8/85 (9.4)	.
Jennum 2015 25914563[125]	NRCS	As-prescribed (No)	36	CPAP No CPAP	NR	NR/13,631 NR/11,758	**Adjusted HR 0.67 (0.61, 0.75) [D]**
Lisan 2019 30973594[126]	NRCS	As-prescribed (No)	133	CPAP No CPAP	NR	12/81 (14.8) 84/311 (27.0)	**Adjusted HR 0.58 (0.35, 0.96) [E]**
López-Padilla 2016 27198943[127]	NRCS	Adherent vs. nonuse / nonadherent (No)	53	CPAP use No CPAP use	100% (adherent)	35/79 (44.3) 48/76 (63.2)	**Adjusted HR 0.46 (0.27, 0.78) [K]**
ONSLEEP 33006309[131]	NRCS	As-prescribed (No)	79	CPAP No CPAP	NR	53/1274 (4.2) 94/1274 (7.4)	**Adjusted HR 0.47 (0.33, 0.66) [F]**
Ou 2015 26068440[132]	NRCS	As-prescribed (No)	60	CPAP No CPAP	67% overall	2/36 (5.6) 19/88 (21.6)	***No CPAP vs. CPAP*: Adjusted HR 12.2 (2.28, 64.7) [G, H]**

Statistically significant results are in bold font.

Abbreviations: . = no information, CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, hr/noc = hours per night, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, ITT = intention-to-treat, MOSAIC = Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular trial, NNH = number needed to harm, NNT = number needed to treat, NR = not reported, NRCS = nonrandomized comparative study, ONSLEEP = Okinawa Nakamura Sleep, OR = odds ratio, RCT = randomized controlled trial, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.
[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[C] Calculated.
[D] Adjusted for age, gender, education, Charlson Comorbidity Index.
[E] Propensity matching (including age, sex, ethnicity, education level, BMI, AHI, smoking, driving a car, the respiratory disturbance index, the total sleep time, neck circumference, hypertension, diabetes, history of myocardial infarction and angina, history of stroke, heart failure, history of a pacemaker, cholesterol and high-density lipoprotein levels, and the Epworth sleepiness score).
[F] Propensity score analysis: age, sex, lifestyle (smoking, drinking), comorbid conditions including cardiovascular disease and/or stroke, other comorbid conditions, ESS, and AHI; further adjusted for age, sex, BMI, ESS status, AHI status, oxygen saturation.
[G] Adjusted for hypertension, coronary heart disease, diabetes, AHI.
[H] This HR is for *no CPAP versus CPAP*, opposite of the typical analysis direction. We inverted the HR for CPAP vs. no CPAP in the meta-analysis: approximate HR 0.08 (95% CI 0.02, 0.44).
[I] Adjusted for age, smoking habit, hypertension, diabetes, pulmonary function testing, and categories of CPAP adherence.
[J] Effect size used in meta-analysis.
[K] Adjusted for age, sex, BMI, alcohol intake, diabetes, stroke, and ischemic cardiac disease. Note that study restricted to age ≥80 years old.

Cardiovascular Mortality

Five studies (4 RCTs and 1 NRCS) comparing CPAP and no CPAP reported on incidence of CVD mortality after about 4 to 7 years (Table 6).[12, 99, 103, 112, 131] The studies included patients with OSA but with different diagnostic methods, including specific AHI thresholds, or in the case of the NRCS, inclusion in a registry. None of the studies was explicitly powered for CV mortality and all were conducted by ITT or "as-prescribed" analysis (CPAP treatment/prescription vs. no treatment/prescription). One RCT also reported an "as-treated" analysis.

Randomized Controlled Trials (Cardiovascular Mortality)

The four RCTs evaluated different, but overlapping, sets of patients at about 4 years of followup. Barbé 2012 included otherwise healthy adults 70 years or younger with at least "moderate" OSA (AHI ≥20). RICCADSA included adults with recent coronary revascularization also with at least "moderate" OSA (AHI ≥15). The SAVE trial included adults with CVD or CeVD with "severe" OSA (ODI ≥12). The ISAACC trial included patients who had been admitted to hospital for acute coronary syndrome and had an EES score of ≤10.

In all RCTs, patients and clinicians were not blinded, but outcome assessors were blinded in Barbé 2012, SAVE and ISAACC. The SAVE trial and ISAACC were thus, rated to be at moderate risk of bias, but the other two trials were further downgraded to high risk of bias due to failure to adjust for significant differences in sleep study measures pre-treatment (Barbé 2012) or high crossover rate (RICCADSA).

Across trials, effect sizes (OR or HR) ranged from 0.41 to 3.08, all statistically nonsignificant and mostly highly imprecise. The NNT to prevent one CV death ranged from 31 to a NNH of 357 across studies. The summary effect size across the RCTs was imprecise at 0.99 (95% CI 0.64 to 1.53) (Figure 3). Under an assumption of a control rate of 1.8 percent (based on meta-analysis), the summary NNT was 5656 (95% CI NNH 108 to NNT 156). The SAVE trial reported both an ITT and a propensity score matched analysis of adherent CPAP users (see Table 6). The HR of the adherent analysis appeared to favor CPAP, but was highly imprecise and statistically similar to the ITT analysis.

Adjusted Nonrandomized Comparative Study (Cardiovascular Mortality)

A single NRCS, ONSLEEP, evaluated CPAP treated and untreated patients in a national registry.[130, 131] The study was deemed to of high risk of bias since the analysis of CVD mortality was not adjusted for remaining differences between groups after propensity score matching (whereas further adjustment was conducted for all-cause mortality). The NRCS reported a HR of 0.54 (95% CI 0.28 to 1.03) over a mean of 6.5 years of follow up (Figure 3).

All Studies (Cardiovascular Mortality)

Combining the single NRCS with the RCTs yielded a summary effect estimate of 0.79 (95% CI 0.50 to 1.26; I^2 =26%), with the heterogeneity resulting from the difference between the ONSLEEP NRCS and the RCTs. Overall, the comparative studies did not provide definitive evidence that CPAP reduces CV mortality (Figure 3). The sensitivity analysis showing the change in results as the analysis borrows strength from the NRCS evidence is shown in Appendix D, Figure D-3. Regardless of the weight given to the NRCS in the meta-analysis, the OR remains unchanged.

Heterogeneity of Treatment Effect (Cardiovascular Mortality)

None of the trials reported subgroup analyses. All studies were conducted by ITT (or "as-prescribed") analysis.

The four RCTs were statistically homogenous ($I^2 = 0\%$). Thus, no potential differences in effects could be elucidated based on eligibility criteria (e.g., history of CVD) or definition of OSA based on either severity (i.e., AHI threshold) or details about how AHI was measured. In a *post hoc* analysis, SAVE found no statistical difference in HR in its ITT and adherent users analyses and no significant association in *post hoc* CPAP dose-response analyses and CV end points. Across studies, there was no correlation between CPAP adherence and effect sizes.

Applicability (Cardiovascular Mortality)

Among the RCTs, about half the included participants (55%) were in the SAVE trial, which was restricted to adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD. All RCTs excluded patients with excessive daytime sleepiness. The effect estimates from the other RCTs were highly imprecise. The summary estimate is most applicable to patients who would meet eligibility criteria for the SAVE trial. The NRCS, in contrast was conducted in generally healthy, relatively younger (≤70 years) participants. None of the studies provided analyses specifically of patients who would be eligible for Medicare (due to age or disability).

Summary of Effect of CPAP on Cardiovascular Mortality

Four RCTs did not provide evidence that CPAP reduces the risk of CV mortality over about 3 to 5 years of followup (low SoE) (Table 4). This conclusion does not imply that CPAP has been proven to be ineffective to reduce CV mortality. Study estimates, individually and by meta-analysis were imprecise (summary ES 0.99 (95% CI 0.64 to 1.53). Inclusion of one NRCS of a generally healthy population followed for about 6.5 years yielded a somewhat smaller effect size, but did not change the conclusion (summary ES 0.79, 95% CI 0.50 to 1.26). The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

Figure 3. Meta-analysis of CPAP versus No CPAP: Cardiovascular mortality
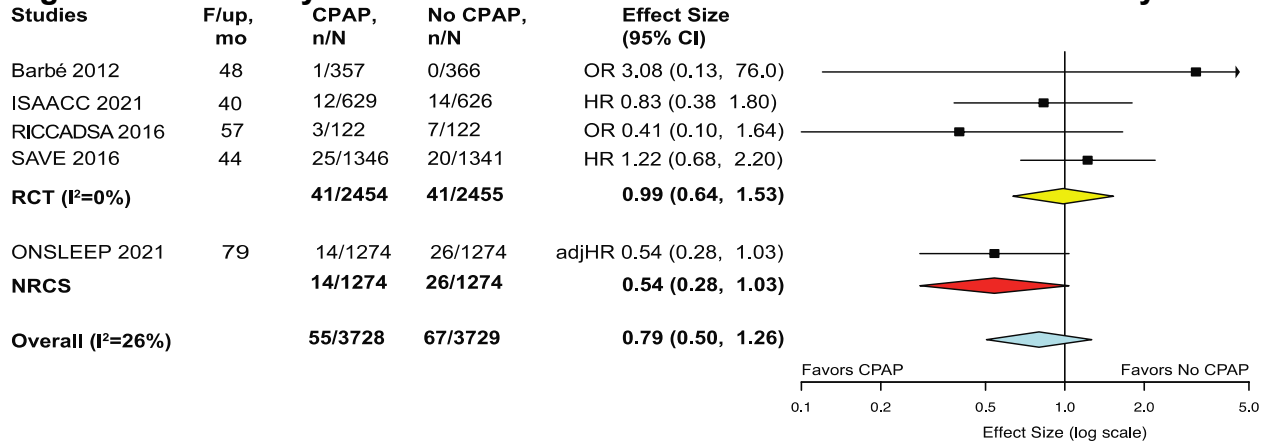
Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Effect Size (95% CI)
Barbé 2012	48	1/357	0/366	OR 3.08 (0.13, 76.0)
ISAACC 2021	40	12/629	14/626	HR 0.83 (0.38 1.80)
RICCADSA 2016	57	3/122	7/122	OR 0.41 (0.10, 1.64)
SAVE 2016	44	25/1346	20/1341	HR 1.22 (0.68, 2.20)
RCT (I²=0%)		**41/2454**	**41/2455**	**0.99 (0.64, 1.53)**
ONSLEEP 2021	79	14/1274	26/1274	adjHR 0.54 (0.28, 1.03)
NRCS		**14/1274**	**26/1274**	**0.54 (0.28, 1.03)**
Overall (I²=26%)		**55/3728**	**67/3729**	**0.79 (0.50, 1.26)**

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, HR = hazard ratio, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, NA = not applicable, NR = not reported, NRCS = nonrandomized comparative study, ONSLEEP = Okinawa Nakamura Sleep, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

Table 6. CPAP versus no CPAP: Cardiovascular mortality

Study PMID	Design	Analysis (Powered? [A])	Followup Duration (mo)	Arm	n/N (%)	CPAP Adherence	Effect Size (95% CI) Calculated NNT (95% CI) [B]
Barbé 2012 22618923[99]	RCT	ITT (No)	48	CPAP No CPAP	1/357 (0.3) 0/366 (0)	64% overall	OR 3.08 (0.13, 76.0) [C] NNH 357 (NNH 121, NNT 373)
ISAACC 31839558[103]	RCT	ITT (No)	40	CPAP No CPAP	12/629 (1.9) 14/626 (2.2)	38% overall	HR 0.83 (0.38, 1.80) NNT 304 (NNH 80, NNT 52)
RICCADSA 26914592[112]	RCT	ITT (No)	57	CPAP No CPAP	3/122 (2.5) 7/122 (5.7)	60% at 1 yr	OR 0.41 (0.10, 1.64) [C] NNT 31 (NNH 60, NNT 12)
SAVE 27571048[12]	RCT	ITT (No)	44	CPAP No CPAP	25/1346 (1.9) 20/1341 (1.5)	42% overall	HR 1.22 (0.68, 2.20) NNH 273 (NNH 75, NNT 166) [Adherent analysis: HR 0.90 (0.41, 2.01)]
ONSLEEP 33006309[131]	NRCS	As-prescribed (No)	79	CPAP No CPAP	14/1274 (1.1) 26/1274 (2.0)	NR	Adjusted HR 0.54 (0.28, 1.03)

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, NR = not reported, NRCS = nonrandomized comparative study, OR = odds ratio, PMID = PubMed Identifier, RCT = randomized controlled trial, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.
[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[C] Calculated.

Stroke and Transient Ischemic Attack

Seven studies evaluated stroke or TIA. Five RCTs comparing CPAP and no CPAP reported on incidence of stroke and/or TIA after 1 or about 4 years (Table 7).[12, 99, 103, 109, 112] The studies included patients with OSA but with different diagnostic methods, including specific AHI thresholds. None of the trials was explicitly powered for either stroke or TIA. All reported ITT analyses. The SAVE trial also reported propensity score matched adherent CPAP users analyses. Two adjusted NRCSs also reported on stroke.[134, 136] One conducted an "as-prescribed" analysis with propensity score matching of adults with OSA followed for 5 years. The second NRCS evaluated patients with type 2 DM and OSA and compared adherent users (≥4 hours per night, ≥70% of nights) with nonadherent users.

The seven studies evaluated different, but overlapping, sets of patients. Three trials (ISAACC, RICCADSA, and SAVE) were restricted to patients with existing CVD or CeVD. One trial (Barbé 2012) excluded patients with CVD or chronic conditions and older adults (>70 years). One trial (PREDICT) was restricted to older adults (≥65 years), regardless of CVD history. One NRCS included all patients with OSA and the other included patients with type 2 DM. Four trials included patients with at least "moderate" OSA (AHI ≥15 or 20, or ODI ≥7.5); SAVE was restricted to patients with severe OSA (ODI ≥12). The Sheth 2021 NRCS provided CPAP for patients with AHI ≥5.

Two RCTs were deemed to be at high risk of bias either because baseline imbalances between groups were not accounted for (Barbé 2012) or high crossover between groups (RICCADSA). The other three RCTs were at moderate risk of bias for lack of patient and clinician blinding (although PREDICT and SAVE blinded outcome assessors). Both NRCSs were deemed to be at high risk of bias. One (Chang 2020) conducted a propensity score analysis, but provided an inadequate description of their methods and results. The Sheth 2021 NRCS conducted a simple regression and did not adequately report their results.

Stroke

All five trials evaluated risk of stroke (Figure 4). Effect sizes (OR or HR) ranged from 0.49 to 1.59 across studies, all statistically nonsignificant. NNT ranged from 41 to a NNH of 340 across studies. Most trials were highly imprecise. The summary effect size was 0.99 (95% CI 0.73 to 1.35; $I^2 = 0\%$), which was imprecise and thus finds no evidence of an effect of CPAP on risk of stroke. Under an assumption of a control rate of 1.9 percent (based on meta-analysis), the summary NNT was 5364 (95% CI NNH 154 to NNT 198). In a propensity score matched analysis of adherent CPAP users, the SAVE trial reported a stronger effect size that was just statistically significant (see Table 7), but the study did not report whether this analysis was significantly different than its primary ITT analysis. The two adjusted NRCSs (Chang 2020, which reported an "as-prescribed" analysis, and Sheth 2021, which compared adherent and nonadherent CPAP users) both reported imprecise effect estimates of stroke (adjusted HR 0.68 and adjusted OR 0.85). Addition of the NRCS to the RCTs did not substantially alter the summary estimate (ES 0.90, 95% CI 0.70 to 1.17, $I^2 = 0\%$).

Transient Ischemic Attack

Four trials evaluated risk of TIA (Figure 5). SAVE and ISAACC specified the outcome as hospitalization due to TIA. Three of the trials were highly imprecise; the largest trial, SAVE, found somewhat more patients in the CPAP group were hospitalized for TIA than in the no

CPAP group, but the estimate was still imprecise. Effect sizes ranged from 0.41 to 1.99 and NNT ranged from 120 to a NNH of 314 across trials. Across all four trials (combining TIA and hospitalization for TIA) the summary effect size was 1.21 (95% CI 0.56 to 2.63; I^2 =16%), which was highly imprecise, precluding a conclusion regarding the effect of CPAP on risk of TIA. Under an assumption of a control rate of 0.7 percent (based on meta-analysis), the summary NNH was 658 (95% CI NNH 86 to NNT 313). The SAVE trial's analysis of adherent CPAP users yielded an even more imprecise estimate (Table 7).

Heterogeneity of Treatment Effect (Stroke and TIA)

None of the RCTs reported subgroup analyses. The NRCS (Chang 2020) stratified their analyses of stroke outcomes by gender and by age groups. Across all groups adjusted HRs were mostly highly imprecise and all were nonsignificant. There was no evidence of differences between groups. For both stroke and TIA outcomes, studies were mostly highly imprecise and were homogeneous across studies, precluding cross-study comparisons. The SAVE trial found a stronger effect on stroke in its analyses of adherent CPAP users than its ITT analysis, but they reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Applicability (Stroke and TIA)

The findings are likely most applicable to patients who would meet eligibility criteria for the SAVE trial (since it was by far the largest study), which was restricted to adults with severe OSA (ODI ≥12) and preexisting CAD or CeVD, but excluding patients with excessive daytime sleepiness. The effect estimates from the other RCTs were highly imprecise. Although, their findings were imprecise, the PREDICT trial may be most applicable to the Medicare population since the trial was restricted to older adults. No trial specifically included participants who would qualify for Medicare based on disability.

Summary of Effect of CPAP on Stroke and Transient Ischemic Attack

Five RCTs do not provide evidence that CPAP reduces the risk of stroke (Table 4), based on a somewhat imprecise estimate of effect (summary ES 0.99, 95% CI 0.73 to 1.35); SoE is low. This conclusion does not imply that CPAP has been proven to be ineffective to reduce stroke. Findings from two high risk of bias adjusted NRCSs do not alter these conclusions. The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

Four RCTs yielded a highly imprecise estimate of the effect of CPAP on risk of TIA and, thus, provide insufficient evidence.

Figure 4. Meta-analysis of CPAP versus No CPAP: Stroke incidence

Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Effect Size (95% CI)
Barbé 2012	48	3/357	2/366	OR 1.54 (0.26, 9.29)
ISAACC 2020	40	8/629	5/626	HR 1.59 (0.52, 4.85)
PREDICT 2014	12	0/114	0/117	OR 1.03 (0.02, 52.16)
RICCADSA 2016	57	3/122	6/122	OR 0.49 (0.12, 1.99)
SAVE 2016	44	67/1346	68/1341	HR 0.98 (0.69, 1.39)
RCT (I²=0%)		81/2568	81/2572	0.99 (0.73, 1.35)
Chang 2020	59	NR/959	NR/959	HR 0.68 (0.38, 1.23)
Sheth 2020	30	14/~179	34/~219	OR 0.85 (0.41, 1.79)
NRCS				0.74 (0.47, 1.17)
Overall (I²=0%)				0.90 (0.70, 1.17)
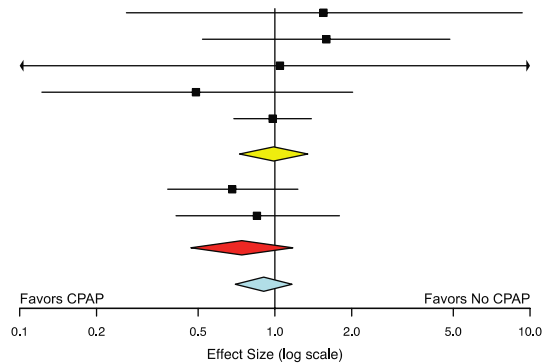
Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

Figure 5. Meta-analysis of CPAP versus No CPAP: Transient ischemic attack incidence

Studies	F/up, mo	CPAP n/N	No CPAP n/N	Effect Size (95% CI)
Barbé 2012	57	2/357	5/366	OR 0.41 (0.08, 2.11)
ISAACC 2020	40	4/629	2/626	HR 1.99 (0.37, 10.9)
PREDICT 2014	24	1/114	2/117	OR 0.51 (0.05, 5.69)
SAVE 2016	44	16/1346	9/1341	HR 1.78 (0.78, 4.04)
Overall (I²=17%)		23/2441	18/2450	1.21 (0.56, 2.63)
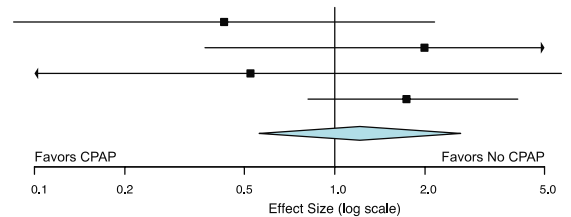
Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, Sleep Apnea cardioVascular Endpoints trial, TIA = transient ischemic attack.

Table 7. CPAP versus no CPAP: Stroke and transient of ischemic attack

Study PMID	Design	Outcome	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
Barbé 2012 22618923[99]	RCT	Stroke	ITT (No)	48	CPAP No CPAP	64% overall	3/357 (0.8) 2/366 (0.5)	OR 1.54 (0.26, 9.29) [C] NNH 340 (NNH 66, NNT 109)
ISAACC, 31839558[103]	RCT	Stroke	ITT (No)	40	CPAP No CPAP	38% overall	8/629 (1.3) 5/626 (0.8)	HR 1.59 (0.52, 4.85) NNH 319 (NNH 68, NNT 118)
PREDICT 25172769[109]	RCT	Stroke	ITT (No)	12	CPAP No CPAP	35% at 1 yr	0/114 (0) 0/117 (0)	OR 1.03 (0.02, 52.2) [C] NNT Not calculable
RICCADSA 26914592[112]	RCT	Stroke	ITT (No)	57	CPAP No CPAP	60% at 1 yr	3/122 (2.5) 6/122 (4.9)	OR 0.49 (0.12, 2.00) [C] NNT 41 (NNH 44, NNT 14)
SAVE 27571048[12]	RCT	Stroke	ITT (No)	44	CPAP No CPAP	42% overall	67/1346 (5.0) 68/1341 (5.1)	HR 0.97 (0.69, 1.35) NNT 1074 (NNH 64, NNT 57) [Adherent analysis: HR 0.56 (0.32, 1.00)]
Chang 2020, 32421898[136]	NRCS	Stroke	As-prescribed (No)	59	CPAP No CPAP	NR	NR/959 (NR) NR/959 (NR)	adjHR 0.68 (0.38, 1.23)
Sheth 2021, 33729910[134]	NRCS	Stroke	Adherent vs. nonadherent (No)	30	CPAP adherent Nonadherent	100%	14/~179 [D] 34/~219 [D]	adjOR 0.85 (0.41, 1.79)
Barbé 2012 22618923[99]	RCT	TIA	ITT (No)	48	CPAP No CPAP	64% overall	2/357 (0.6) 5/366 (1.4)	OR 0.41 (0.08, 2.11) [C] NNT 124 (NNH 163, NNT 45)
PREDICT 25172769[109]	RCT	TIA	ITT (No)	12	CPAP No CPAP	35% at 1 yr	1/114 (0.9) 2/117 (1.7)	OR 0.51 (0.05, 5.69) [C] NNT 120 (NNH 48, NNT 27)
ISAACC, 31839558[103]	RCT	Hospitalization for TIA	ITT (No)	40	CPAP No CPAP	38% overall	4/629 (0.6) 2/626 (0.3)	HR 1.99 (0.37, 10.9) NNH 314 (NNH 93, NNT 226)
SAVE 27571048[12]	RCT	Hospitalization for TIA	ITT (No)	44	CPAP No CPAP	42% overall	16/1346 (1.2) 9/1341 (0.7)	HR 1.88 (0.83, 4.28) NNH 193 (NNH 80, NNT 481) [Adherent analysis: HR 0.22 (0.03, 2.01)]

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, NR = not reported, OR = odds ratio, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, Sleep Apnea cardioVascular Endpoints trial, TIA = transient ischemic attack.

[A] Explicitly powered for this outcome.

[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).

[C] Calculated.

[D] Number of participants analyzed in regression not adequately reported.

Acute Myocardial Infarction

Five RCTs comparing CPAP and no CPAP reported on incidence of acute MI after 1 or about 4 years (Table 8).[12, 99, 103, 109, 112] The studies included patients with OSA but with different diagnostic methods, including specific AHI thresholds. None of the trials was explicitly powered for MI. All reported ITT analyses; the SAVE trial also reported a propensity score matched adherent CPAP users analysis.

The five RCTs evaluated different, but overlapping, sets of patients. Three trials (ISAAC, SAVE, and RICCADSA) were restricted to patients with existing CVD or CeVD. One trial (Barbé 2012) excluded patients with CVD or chronic conditions and older adults (>70 years). One trial (PREDICT) was restricted to older adults (≥65 years), regardless of CVD history. Four trials included patients with at least "moderate" OSA (AHI ≥15 or 20, or ODI ≥7.5); SAVE was restricted to patients with severe OSA (ODI ≥12).

Two RCTs were deemed to be at high risk of bias either because baseline imbalances between groups were not accounted for (Barbé 2012) or high crossover between groups (RICCADSA). The other three RCTs were at moderate risk of bias for lack of patient and clinician blinding (although PREDICT and SAVE blinded outcome assessors). Across the five RCTs there was a wide range of effect size estimates (OR or HR) from 0.25 to 7.38, but estimates were all imprecise (or highly imprecise) and statistically nonsignificant. One study suggested a NNT of 62 and the other three had NNHs ranging from 38 to 472. The summary effect size across the RCTs was nonsignificant at 1.05 (95% CI 0.78 to 1.41) (Figure 6). With the wide confidence intervals within each study, the meta-analysis was statistically homogeneous (I^2 = 0%). Under an assumption of a control rate of 3.2 percent (by meta-analysis), the summary NNH was 653 (95% CI NNH 80 to NNT 147).

Heterogeneity of Treatment Effect (Acute Myocardial Infarction)

None of the trials reported subgroup analyses. The four RCTs were statistically homogenous; thus, no potential differences in effects based on factors of interest could be elucidated. The SAVE trial found no difference in effect in its analysis of adherent CPAP users than its ITT analysis (Table 8). They also reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Applicability (Acute Myocardial Infarction)

The findings are likely most applicable to patients who would meet eligibility criteria for the SAVE or ISAAC trials, which was restricted to adults with preexisting CAD or CeVD and moderate to severe OSA. The effect estimates from the other RCTs were highly imprecise. Although, their findings were imprecise, the PREDICT trial may be most applicable to the Medicare population since the trial was restricted to older adults. No trial specifically included participants who would qualify for Medicare based on disability.

Summary of Effect of CPAP on Acute Myocardial Infarction

Five trials did not provide evidence that CPAP reduces the risk of acute MI (Table 4). This conclusion does not imply that CPAP has been proven to be ineffective to reduce acute MI. The RCTs reported a wide range of estimates of effect of CPAP on acute MI, but all were statistically nonsignificant and the meta-analysis was statistically homogeneous. By meta-analysis, the RCTs failed to find a significant difference in risk of acute MI (summary ES 1.05, 95% CI 0.78 to 1.41). In aggregate, the studies were of moderate risk of bias and provided an imprecise estimate

of effect. The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

Figure 6. Meta-analysis of CPAP versus No CPAP: Acute myocardial infarction incidence
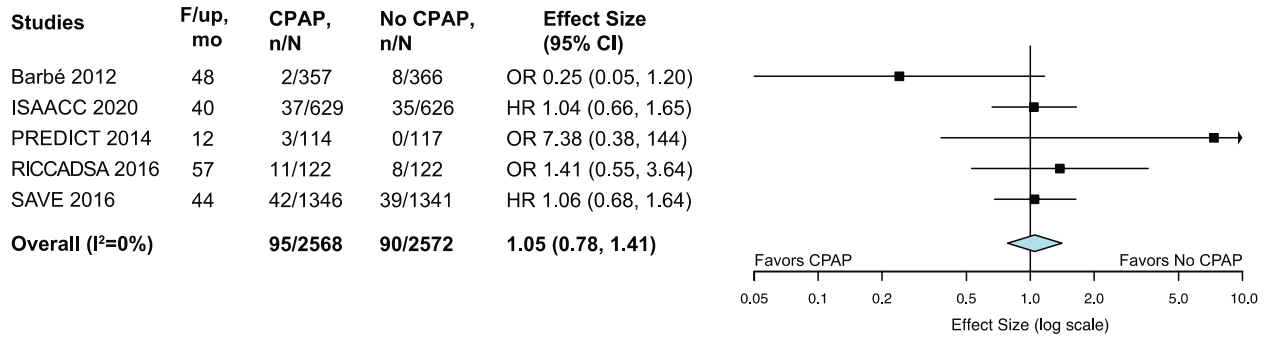
Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Effect Size (95% CI)
Barbé 2012	48	2/357	8/366	OR 0.25 (0.05, 1.20)
ISAACC 2020	40	37/629	35/626	HR 1.04 (0.66, 1.65)
PREDICT 2014	12	3/114	0/117	OR 7.38 (0.38, 144)
RICCADSA 2016	57	11/122	8/122	OR 1.41 (0.55, 3.64)
SAVE 2016	44	42/1346	39/1341	HR 1.06 (0.68, 1.64)
Overall (I²=0%)		**95/2568**	**90/2572**	**1.05 (0.78, 1.41)**

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, PREDICT = trial with undefined acronym, Sleep Apnea cardioVascular Endpoints trial.

Table 8. CPAP versus no CPAP: Acute myocardial infarction

Study PMID	Design	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
Barbé 2012 22618923[99]	RCT	ITT (No)	48	CPAP No CPAP	64% overall	2/357 (0.6) 8/366 (2.2)	OR 0.25 (0.05, 1.20) [C] NNT 62 (NNH 1647, NNT 30)
ISAACC, 31839558[103]	RCT	ITT (No)	40	CPAP No CPAP	38% overall	37/629 (5.9) 35/626 (5.6)	HR 1.04 (0.66, 1.65) NNH 315 (NNH 35, NNT 44)
PREDICT 25172769[109]	RCT	ITT	12	CPAP No CPAP	35% at 1 yr	3/114 (2.6) 0/117 (0)	OR 7.38 (0.38, 144) [C] NNH 38 (NNH 18, NNT 326)
RICCADSA 26914592[112]	RCT	ITT (No)	57	CPAP No CPAP	60% at 1 yr	11/122 (9.0) 8/122 (6.6)	OR 1.41 (0.55, 3.64) [C] NNH 41 (NNH 11, NNT 23)
SAVE 27571048[12]	RCT	ITT (No)	44	CPAP No CPAP	42% overall	42/1346 (3.1) 39/1341 (2.9)	HR 1.06 (0.68, 1.64) NNH 472 (NNH 66, NNT 93) [Adherent analysis: HR 1.19 (0.59, 2.39)]

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, NR = not reported, OR = odds ratio, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, Sleep Apnea cardioVascular Endpoints trial, TIA = transient ischemic attack.

[A] Explicitly powered for this outcome.

[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).

[C] Calculated.

Angina

Three RCTS reported on risk of angina. The PREDICT trial reported on incidence of angina at 1 year[109] and the SAVE trial and ISAACC trial reported on incidence of hospitalization for unstable angina at 3.3 to 3.7 years (Table 9).[12, 103] The PREDICT trial was restricted to older adults (≥65 years), regardless of CVD history while the SAVE and ISAAC trial included only adults with existing CVD or CeVD. The three trials were deemed to be at moderate risk of bias. None was explicitly powered for angina. All reported ITT analyses; SAVE also reported a propensity score matched adherers analysis.

PREDICT provided a highly imprecise estimate of effect of CPAP on incident angina (OR 0.68; 95% CI 0.11 to 4.18; NNT 126). SAVE and ISAACC both found no significant difference in risk of hospitalization for unstable angina (SAVE: HR 1.09, 95% CI 0.82 to 1.45, NNH 155; ISAACC: HR 0.73, 95% CI 0.47 to 1.13, NNT 52). The analysis CPAP adherers yielded a similar estimate in the SAVE trial (Table 9). SAVE also reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Heterogeneity of Treatment Effect (Angina)

None of the studies reported subgroup analyses. Participants in the SAVE and ISAACC trials (who were selected in part based on CV risk factors) were at higher risk of angina than in the PREDICT trial (who were selected based only on older age). However, there was no discernable difference among studies. As noted, the effect of CPAP did not vary based on adherence.

Applicability (Angina)

The findings are most applicable to patients who would meet eligibility criteria for the SAVE or ISAACC trials, which were restricted to adults with preexisting CAD or CeVD.

Summary of Effect of CPAP on Angina

With a single RCT evaluating angina and two RCTs evaluating hospitalization for angina, together with imprecision of estimates in the trials, there is insufficient evidence to determine the effect of CPAP on risk of incident angina or hospitalization for unstable angina (Table 4).

Table 9. CPAP versus no CPAP: Angina

Study PMID	Design	Outcome	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
ISAACC, 31839558[103]	RCT	Hospital admission for unstable angina	ITT (No)	40	CPAP No CPAP	38% overall	34/629 (5.4) 46/626 (7.3)	HR 0.73 (0.47, 1.13) NNT 52 (NNH 127, NNT 22)
PREDICT 25172769[109]	RCT	Angina	ITT (No)	12	CPAP No CPAP	35% at 1 yr	2/113 (1.8) 3/117 (2.6)	OR 0.68 (0.11, 4.18) [C] NNT 126 (NNH 34, NNT 22)
SAVE 27571048[12]	RCT	Hospitalization for unstable angina	ITT (No)	44	CPAP No CPAP	42% overall	99/1346 (7.4) 90/1341 (6.7)	HR 1.09 (0.82, 1.45) NNH 155 (NNH 39, NNT 78) [Adherent analysis: HR 0.99 (0.64, 1.51)]

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, NR = not reported, OR = odds ratio, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.
[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[C] Calculated.

Revascularization

Four studies (three RCTs and one NRCS) comparing CPAP and no CPAP reported on incidence of revascularization (Table 10).[12, 103, 112, 135] Two studies (RICCADSA, and the Wu 2015 [Beijing] NRCS) included only patients with prior coronary artery revascularization. These studies evaluated the risk of repeat coronary artery revascularization. In contrast, ISAACC included patients who were hospitalized for acute coronary syndrome. The SAVE RCT included patients with any history of CVD or CeVD and evaluated the risk of any (major) arterial revascularization, including coronary artery, peripheral artery, or carotid artery (stent insertion). The studies evaluated patients at 3.3 to 5 years of followup. The studies included patients with OSA, but with different diagnostic methods, including specific AHI thresholds.

Randomized Controlled Trials (Revascularization)

The three RCTs evaluated different outcomes (the occurrence of subsequent repeat coronary artery revascularization, first coronary artery revascularization, and any major arterial revascularization). None of the RCTs blinded patients or their clinicians, but SAVE and ISAACC blinded outcome assessors and were deemed to be at moderate risk of bias. RICCADSA was deemed to be at high risk of bias due to a high crossover rate. No trial was explicitly powered for a revascularization outcome. Each trial reported an ITT analysis, but SAVE also reported a *post hoc* propensity score matched analysis of CPAP adherers.

The SAVE trial found no significant difference in risk for any subsequent arterial revascularization procedure at a mean of 3.7 years followup (OR 1.31; 95% CI 0.91 to 1.89; NNH 41) in patients with prior history of any CVD or CeVD (Table 10).[12] The CPAP adherers analysis found a similar result.

ISAACC found no significant difference in risk of revascularization at a mean of 3.3 years (HR 0.98; 95% CI 0.70 to 1.38; NNT 1989) in nonsleepy patients who had been admitted to hospital for acute coronary syndrome.[103]

RICCADSA found no significant difference in risk of repeat coronary revascularization at a mean of 4.7 years (OR 1.25; 95% CI 0.59 to 2.66; NNH 85) in patients with prior coronary artery revascularization.[112]

Adjusted Nonrandomized Comparative Study (Revascularization)

The Wu 2015 (Beijing) NRCS was deemed to be at moderate risk of bias for failure to use propensity score matching (or equivalent methodology). As all NRCSs, the study was not explicitly reported to be powered for this outcome. The study compared adherent CPAP users with those who refused CPAP or were nonadherent. The study found a statistically significant association between *no CPAP* use and increased risk of repeat coronary artery revascularization (adjusted HR of *no CPAP vs. CPAP* 2.13, 95% CI 1.19 to 3.81) at a mean of 4.8 years of followup.[135] Particularly as an analysis of adherent users, this NRCS may have been subject to uncontrolled biases related to self-selection of CPAP use and adherence.

Heterogeneity of Treatment Effect (Revascularization)

None of the studies reported subgroup analyses and no explanations for differences in findings across studies were evident. SAVE reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Applicability (Revascularization)

The findings regarding risk of any vascularization are applicable to people with preexisting CAD or CeVD. By definition, the findings regarding the risk of repeat coronary artery revascularization are intended to be applicable to people who have had prior coronary artery revascularization. None of the studies specifically addressed patients who would be eligible for Medicare enrollment based on either age or disability status.

Summary of Effect of CPAP on Revascularization

Two of the RCTs evaluated related, but distinct outcomes (first coronary revascularization and any revascularization). With only a single trial evaluating any (major) first revascularization (coronary, peripheral, or carotid arteries), there is insufficient evidence to make a conclusion regarding the effect of CPAP on these outcomes (Table 4).

There is also insufficient evidence regarding the effect of CPAP on risk of repeat coronary artery revascularization. The RCT and NRCS effect measures differ in both direction and significance, with the RCTs showing a nonsignificantly lower OR for revascularization in the no CPAP group, while the single NRCS shows a significantly higher adjusted HR for revascularization in the no CPAP group.

Table 10. CPAP versus no CPAP: Revascularization

Study PMID	Design	Outcome	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
ISAACC, 31839558[103]	RCT	Revascularization	ITT (No)	40	CPAP No CPAP	38% overall	66/629 (10.5) 66/626 (10.5)	HR 0.98 (0.70, 1.38) NNT 1989 (NNH 30, NNT 29)
RICCADSA 26914592[112]	RCT	Repeat coronary revascularization	ITT	57	CPAP No CPAP	60% at 1 yr	17/122 (13.9) 14/122 (11.5)	OR 1.25 (0.59, 2.66) [C] NNH 41 (NNH 9, NNT 17)
SAVE 27571048[12]	RCT	Any revascularization	ITT (No)	44	CPAP No CPAP	42% overall	69/1346 (5.1) 53/1341 (4.0)	OR 1.31 (0.91, 1.89) [C] NNH 85 (NNH 36, NNT 250) [Adherent analysis: HR 1.25 (0.79, 1.96)]
Wu 2015 (Beijing) 25412159[135]	NRCS	Repeat coronary revascularization	Adherent vs. nonuse / nonadherent (No)	60	CPAP adherent No CPAP	100% (adherent)	15/128 (11.7) 35/167 (21.0)	**No CPAP vs. CPAP: Adjusted HR 2.13 (1.19, 3.81) [D]**

Statistically significant results are in bold font.

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, RCT = randomized controlled trial, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.
[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[C] Calculated.
[D] This HR is for *no CPAP versus CPAP*, opposite of other comparisons. Adjusted for age, sex, BMI, clinical presentation (i.e., stable angina, NSTACS, STEMI), smoking, hypertension, type 2 diabetes, dyslipidemia, history of myocardial infarction, cerebrovascular disease, peripheral arterial disease, renal failure, heart failure (LVEF 40%), extent of diseased or treated vessel, revascularization type (i.e., emergency vs elective), adjunctive medical therapy.

Congestive Heart Failure

Three RCTs and one NRCS comparing CPAP and no CPAP reported on incidence of either CHF[99, 129] or hospitalization for CHF (Table 11).[12, 103] The studies evaluated distinct populations. The two studies that reported incident CHF included either otherwise healthy participants (RCT Barbé 2012) or resistant HTN (NRCS Navarro-Soriano 2021). The two RCTs that reported hospitalization for CHF included either patients admitted to hospital for acute coronary syndrome (ISAACC) or more general histories of either CVD or CeVD (SAVE). Rates of incident CHF or hospitalization for CHF varied widely across studies, from 0.8 to 7.4 percent.

In all studies, patients and clinicians were not blinded. SAVE and ISAACC trials were deemed to be at moderate risk of bias. Barbé 2012, however, had significant differences in baseline AHI between groups that was not accounted for in their analyses; thus, the study was deemed to be at high risk of bias. Navarro-Soriano 2021 reported on a secondary analysis of a completed RCT in which adherent CPAP users were compared with combined nonadherent users and those not offered CPAP; the NRCS was deemed to be at high risk of bias, primarily for using a simple regression analysis. No study was explicitly powered for a CHF outcome. The three RCTs reported ITT analyses. The Navarro-Soriano 2021 NRCS compared adherent users versus nonadherent users and nonusers and SAVE also reported a *post hoc* propensity score matched analysis of CPAP adherers.

Barbé 2012 and Navarro-Soriano 2021 each found highly imprecise estimates of the difference in risk of incident CHF at 4 to 5 years of followup (Table 11). Similarly, ISAACC and SAVE each found imprecise estimates of the incidence of hospitalization for CHF at about 3 to 4 years of followup. The SAVE analysis of CPAP adherers yielded very similar results.

Heterogeneity of Treatment Effect (CHF)

The RCTs did not report subgroup analyses. SAVE reported no significant association in *post hoc* CPAP dose-response analyses and CV end points. The four RCTs were imprecise, given the relatively low incidence of CHF; thus, no differences could be discerned between trials.

Applicability (CHF)

The four trials covered different populations. In theory, this would broaden applicability beyond any given study characteristics; however, given the imprecise (and, thus, inconclusive) findings, the true applicability is unclear.

Summary of Effect of CPAP on Congestive Heart Failure

With few studies addressing each outcome and imprecise or highly imprecise estimates, there is insufficient evidence to determine the effect of CPAP on risk of CHF or hospitalization for CHF (Table 4).

Table 11. CPAP versus no CPAP: Congestive heart failure

Study PMID	Design	Outcome	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
Barbé 2012 22618923[99]	RCT	CHF	ITT (No)	48	CPAP No CPAP	64% overall	3/357 (0.8) 5/366 (1.4)	OR 0.61 (0.15, 2.58) [C] NNT 190 (NNH 101, NNT 49)
Navarro-Soriano, 2021, 32029279[129]	NRCS	CHF	Adherent vs. nonuse / nonadherent (No)	58	CPAP adherent No CPAP use	100%	5/95 (5.3) 5/68 (7.4)	*No CPAP vs. CPAP* HR 1.3 (0.4, 4.2)
ISAACC, 31839558[103]	RCT	Hospitalization for CHF	ITT (No)	40	CPAP No CPAP	38% overall	22/629 (3.5) 25/626 (4.0)	HR 0.86 (0.49, 1.53) NNT 202 (NNH 62, NNT 39)
SAVE 27571048[12]	RCT	Hospitalization for CHF	ITT (No)	44	CPAP No CPAP	42% overall	17/1346 (1.3) 17/1341 (1.3)	HR 0.98 (0.50, 1.92) NNT 21,235 (NNH 119, NNT 118) [Adherent analysis: HR 0.82 (0.34, 2.03)]

Abbreviations: CHF = congestive heart failure, CPAP = continuous positive airway pressure, CI = confidence interval, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, RCT = randomized controlled trial, Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.
[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[C] Calculated.

Atrial Fibrillation

Three RCTs and one NRCS reported on risk of AFib, with followup between 1 and 4.8 years (Table 12).[12, 103, 129] Two of the trials included patients with different CV risk factors, including hospitalization for acute coronary syndrome (ISAACC) and existing CVD or CeVD (SAVE). The Navarro-Soriano 2021 NRCS included participants with resistant HTN. The PREDICT trial was restricted to older adults (≥65 years), regardless of CVD history. ISAACC evaluated a composite of AFib and other (undefined) arrhythmias. The four trials were deemed to be at moderate risk of bias. All trials were conducted by ITT analysis, but SAVE also reported a propensity score matched analysis of CPAP adherers. None was explicitly powered for AFib. The NRCS compared adherent users versus nonadherent users and nonusers.

The three RCTs all found imprecise or highly imprecise estimates of effect, with a range of effect sizes (OR or HR) from 0.48 to 1.47. Meta-analysis also yielded an imprecise effect (ES 0.89, 95% CI 0.48 to 1.63, I^2 = 49%), with a moderate degree of statistical heterogeneity (related to the stronger effect size seen in PREDICT compared with other trials) (Figure 7). Under an assumption of a control rate of 4.0 percent (by meta-analysis), the summary NNT was 233 (95% CI NNH 42 to NNT 48). Addition of the Navarro-Soriano 2021 NRCS did not substantively change the summary effect size (0.94, 95% CI 0.58 to 1.51, I^2 = 28%) (Figure 7). The SAVE CPAP adherers analysis found a similar result.

Heterogeneity of Treatment Effect (AFib)

None of the studies reported subgroup analyses. SAVE reported no significant association in *post hoc* CPAP dose-response analyses and CV end points. The PREDICT trial (of older adults not selected for CV risk factors) had a relatively high incidence of AFib and found a stronger effect than the other trials, but the explanation is unclear.

Applicability (AFib)

The four studies covered different populations, but mostly with risk factors for CVD. In theory, this would broaden applicability beyond any given study characteristics; however, given the imprecise (and, thus, inconclusive) findings, the true applicability is unclear.

Summary of Effect of CPAP on Atrial Fibrillation

Three RCTs do not provide evidence that CPAP affects the risk of AFib, with low SoE (Table 4). The low SoE for the conclusions suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding that the findings are stable or that the estimate of effect is close to the true effect Inclusion of a single NRCS comparing adherent CPAP users and nonusers did not change the conclusion. The overall estimate of effect was imprecise.

Figure 7. Meta-analysis of CPAP versus No CPAP: Atrial fibrillation incidence
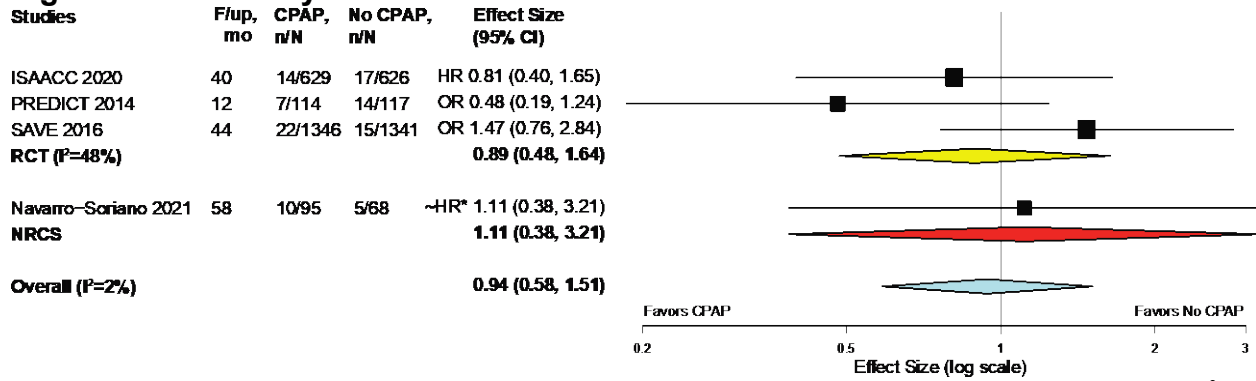
Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Effect Size (95% CI)
ISAACC 2020	40	14/629	17/626	HR 0.81 (0.40, 1.65)
PREDICT 2014	12	7/114	14/117	OR 0.48 (0.19, 1.24)
SAVE 2016	44	22/1346	15/1341	OR 1.47 (0.76, 2.84)
RCT (I²=48%)				0.89 (0.48, 1.64)
Navarro–Soriano 2021	58	10/95	5/68	~HR* 1.11 (0.38, 3.21)
NRCS				1.11 (0.38, 3.21)
Overall (I²=2%)				0.94 (0.58, 1.51)

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, HR = hazard ratio, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, NRCS = nonrandomized comparative study, OR = odds ratio, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

* Approximate hazard ratio, which was reported for no CPAP versus CPAP.

Table 12. CPAP versus no CPAP: Atrial fibrillation

Study PMID	Design	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
ISAACC 31839558[103]	RCT	ITT (No)	40	CPAP No CPAP	38% overall	14/629 (2.2) 17/626 (2.7)	HR 0.81 (0.40, 1.65) [C] NNT 204 (NNH 81, NNT 45)
PREDICT 25172769[109]	RCT	ITT (No)	12	CPAP No CPAP	35% at 1 yr	7/114 (6.1) 14/117 (12.0)	OR 0.48 (0.19, 1.24) [D] NNT 17 (NNH 66, NNT 8)
SAVE 27571048[12]	RCT	ITT (No)	44	CPAP No CPAP	42% overall	22/1346 (1.6) 15/1341 (1.1)	OR 1.47 (0.76, 2.84) [D] NNH 194 (NNH 72, NNT 274) [Adherent analysis: HR 1.84 (0.74, 4.55)]
Navarro-Soriano 2021, 32029279[129]	NRCS	Adherent vs. nonuse / nonadherent (No)	58	CPAP adherent No CPAP	100%	10/95 (10.5) 5/68 (7.4)	*No CPAP vs. CPAP* HR 0.9 (0.3, 2.6)

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, HR = hazard ratio, ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, PREDICT = trial with undefined acronym, RCT = randomized controlled trial, Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.

[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).

[C] Atrial fibrillation or other arrhythmias.

[D] Calculated.

Composite Cardiovascular Outcomes

Twelve studies (7 RCTs and 5 NRCSs) comparing CPAP and no CPAP reported on various composite CV outcomes (Tables 13 and 14).[12, 99, 102, 103, 106, 107, 109, 112, 128, 129, 133-135] Studies evaluated a wide range of followup times (from 1 to 20 years). The studies included patients with OSA but with different diagnostic methods, including specific AHI thresholds or database classification (e.g., International Statistical Classification of Diseases and Related Health Problems [ICD] 9 or 10 codes).

Randomized Controlled Trials (Composite Cardiovascular Outcomes)

Seven RCTs evaluated 11 unique (though overlapping) composite CV outcomes, as described in Table 13, across 1 to about 5 years of followup. Participant eligibility criteria varied across studies, including studies restricted to patients with known CVD (or CeVD) and studies of generally healthy patients (other than OSA).

Four of the trials were deemed to be at moderate risk of bias for lack of blinding of patients and their clinicians, but blinding of outcome assessors (Huang 2015, ISAACC, PREDICT, SAVE). Three trials were at high risk of bias for failure to adjust for baseline differences in AHI (Barbé 2012), high dropout rate (MOSAIC), and high crossover rate (RICCADSA).

Three of the RCTs based their power calculations on composite CV outcomes (SAVE, RICCADSA, ISAACC); however, RICCADSA was ultimately underpowered. None of these studies found a statistically significant or, per their power calculations, a clinically meaningful effect. The RICCADSA trial determined its sample size on a predicted difference in risk of composite CV mortality, acute MI, and new revascularization. Prior to the study they predicted about 25 percent of participants would be nonadherent with CPAP and that 25 percent of participants assigned to no CPAP would have the outcome within 3 years, compared with 10 percent of those using CPAP. A blinded interim analysis found that, in fact, 40 percent of participants were nonadherent and that 21 percent had an incident outcome. The trial was thus enlarged (from a goal of 200 to 242 patient) and the followup was extended to at least 2 years based on a predicted risk reduction from 25 to 12 percent. Ultimately, the authors noted that the study remained underpowered (i.e., that the study found a statistically nonsignificant difference in the primary outcome).

The SAVE trial originally planned to recruit 5000 participants based on a 20 percent relative reduction (from an annual 6.0% to 4.8%) in the risk of composite CV death, nonfatal MI, nonfatal stroke, and any hospitalization for unstable angina, CHF, or TIA. However, due in part to recruitment difficulties and a blinded interim analysis that found better-than-expected CPAP adherence and an annual event rate of 6.86 percent, they revised their sample size to 2500, which they predicted would have 90 percent power to detect a 25 percent lower primary outcome rate over a mean of 4.5 years.

The ISAACC trial met its recruitment goal of 1242 participants for analysis. Their power calculation assumed a 25 percent relative reduction in the risk of a new CV event in a population of patients hospitalized with acute coronary syndrome. They assumed that more than 12 to 20 percent of patients would have new CV event after 1 year. Their primary outcome was a composite of CV death; nonfatal MI or stroke; and hospital admission for CHF, unstable angina or TIA.

Six of the seven RCTs were analyzed on an ITT basis, comparing those prescribed CPAP with those not prescribed CPAP. In contrast, Huang 2015 excluded the 9.5 percent of

participants with poor adherence (mean CPAP use <4 hours/night). Among the trials reporting ITT analyses, between 36 and 64 percent of those assigned to CPAP were adherent (≥4 hours per night). Three of the ITT trials also reported analyses among adherent users.

Summarization across the studies is hampered by the variable definitions of composite outcomes (which CV and related events were included) and the variability in the underlying risk of CV events and death, based on prior history of CVD, CeVD, and on age (Table 13). Nevertheless, excluding the small "as-treated" RCT (Huang 2015), the estimates of the effect of CPAP on composite CV events ranged from an OR of 0.42 to a HR of 1.10. NNT ranged from 10 to a NNH of 63.

The one "as-treated" RCT had a near-imprecise OR of 0.18 (95% CI 0.02 to 1.65). The three ITT RCTs that also reported sensitivity analyses among just CPAP adherers (SAVE, Barbé 2012, RICCADSA) all found nominally stronger HRs in the CPAP adherer analysis, but the difference between analyses varied across studies (Table 13). In SAVE, CPAP-adherer based analyses of the primary outcome and composite ischemic CV events shifted the direction of effects to favor CPAP (in comparison with ITT analyses), but the analyses remained statistically nonsignificant (P = 0.13 and 0.17, respectively). Similarly, for one of its alternative (not primary) composite CV outcomes—CV death, myocardial infarction, stroke—they reported a stronger, but still nonsignificant effect (HR 0.69, P = 0.08 [not accounting for multiple testing]). Barbé 2012 found no difference in its estimates of rate ratios from its ITT and adherers analyses; their comparison of rate ratios among adherers and among nonadherers were not statistically significantly different from each other. In contrast, RICCADSA reported a large shift in HR between the ITT analysis (HR 0.62, 95% CI 0.34 to 1.13) and the multivariable "as-treated" (adherent vs. nonadherent and nonusers) analysis (HR 0.29, 95% CI 0.10 to 0.86).

A rough comparison between ITT and adherent/"as-treated" analyses among the RCTs found no significant difference in effect between analyses (P = 0.53).

All but one ITT estimate of effect were statistically nonsignificant (and the MOSAIC trial found a "just-significant" effect with a P-value of 0.049 for 2 years but a P-value of 0.52 for 5 years). The studies' effects sizes did not clearly differ from each other based on whether included participants had or did not have a history of CVD. There was no correlation between adherence rate and effect size.

Adjusted Nonrandomized Comparative Studies (Composite Cardiovascular Outcomes)

Five NRCSs evaluated six distinct (but overlapping) composite CV outcomes (also distinct from the RCTs). One NRCS (Wu 2015 [Beijing]) evaluated a population with recent coronary revascularization and one (Navarro-Soriano 2021) evaluated patients with resistant HTN (Table 14), one study (Sheth 2021) evaluated patients with type 2 DM, and the other two NRCSs (Myllylä 2019 and Schipper 2017) evaluated otherwise healthy participants (other than OSA). Outcomes were assessed at approximately 5 to 6 years in three of the studies, but at 2.5 years in Sheth 2021 and about 9 years in Myllylä 2019. All NRCSs were deemed to be at high risk of bias for failure to use propensity score matching (or equivalent methodology).

The five NRCSs all primarily evaluated adherent CPAP users (although Myllylä 2019 reported evaluating long-term CPAP treatment, only implying these patients were adherent). Three of the NRCSs compared adherent users with nonusers (Myllylä 2019) or a combination of nonusers and nonadherent users (Navarro-Soriano 2021 and Wu 2015 [Beijing]). Two of the NRCSs compared adherent and nonadherent users (Schipper 2017 and Sheth 2021).

The NRCSs were adjusted for potential confounders and/or patient characteristics. The specific adjustments made for each NRCS are presented in the footnotes in Table 14. All studies adjusted for participant age (although not by uniform categories), and most adjusted for gender. All adjusted for CVD or risk factors (e.g., BMI, DM, dyslipidemia, HTN, smoking). Two studies (Myllylä 2019 and Schipper 2017) adjusted for AHI.

Like all other NRCSs, none was explicitly powered for any outcome, including the composite CV outcome. All reported analyses comparing adherent CPAP users with a combination of nonusers and nonadherent (<4 hours/night) participants. Given that the studies were analyses of adherent users, these NRCSs may have been particularly subject to uncontrolled biases related to self-selection of CPAP use and adherence.

The NRCSs all found adjusted HRs (or OR) that favored CPAP use in direction of effect. Inverting the reported HR for the two studies that compared *no CPAP use* with *CPAP use* (in order to estimate the HR for the typical direction of analysis), the range of adjusted HRs across the NRCSs was 0.37 to 0.83. The three NRCSs conducted in patients with preexisting CVD (Wu 2015 [Beijing]) or CV risk factors (Navarro-Soriano 2021 and Sheth 20211) were not statistically significant, while two NRCSs that did not restrict eligibility based on prior CVD history were.

Heterogeneity of Treatment Effect (Composite Cardiovascular Outcomes)

None of the studies reported subgroup analyses and no explanations for differences in findings across studies were evident. SAVE reported no significant association in *post hoc* CPAP dose-response analyses and CV end points.

Applicability (Composite Cardiovascular Outcomes)

The studies evaluated a spectrum of adult patients with at least "moderately" severe OSA (AHI ≥15), both those with and without a history of CVD. Only the PREDICT trial (which found a nearly highly imprecise effect of CPAP on composite AFib, angina, MI, peripheral vascular disease, stroke, and TIA) was conducted solely in a population that would be eligible for Medicare enrollment (based on age).

Summary of Effect of CPAP on Composite Cardiovascular Outcomes

Seven RCTs do not provide evidence that CPAP affects the risk of composite CV outcomes, with low SoE (Table 4). However, each trial evaluated a unique specific outcome, most provided imprecise estimates of effect, and in aggregate, the trials were at high risk of bias. Across RCTs, the "as-treated" analyses of CPAP users found stronger effects than the ITT analyses of those prescribed CPAP, but there was no significant difference between the sets of analyses. In addition, these "as-treated" analyses were *post hoc*.

In contrast, two NRCSs conducted in otherwise healthy adults found strong associations between CPAP use (vs. nonuse/nonadherence) and reduced risk of (variable) composite CV outcomes. However, three other NRCS, conducted in patients with CVD or risk factors, found no significant association, similar to the RCTs. Given the differences across all studies (RCTs and NRCSs), particularly related to differences in evaluated composite CV outcomes, we were unable to discern a reason why some RCTs and NRCSs found significant effects, while most did not. Particularly as analyses of adherent users, the NRCSs may have been subject to uncontrolled biases related to self-selection of CPAP use and adherence.

Overall, we concluded that studies (both RCTs and NRCSs) have not provided evidence that CPAP affects the risk of composite CV outcomes. This conclusion does not imply that CPAP has been proven to be ineffective to reduce CV outcomes. At best, we have limited confidence in the summary estimates. Additional evidence for a consistently defined composite CV outcome is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

Table 13. CPAP versus no CPAP: Composite cardiovascular outcomes, randomized controlled trials

Study PMID	Population [A]	Outcome	Analysis (Powered? [B])	Followup Duration CPAP Adherence	CPAP n/N (%)	No CPAP n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [C]
Barbé 2012 22618923[99]	No CVD ≤70 yo	Arrhythmia, CV death, HF, MI, stroke, TIA, UA	ITT (No)	4 yr 64% overall	28/357 (7.8)	31/366 (8.5)	Rate ratio 0.87 (0.52, 1.45) NNT 160 (NNH 30, NNT 22) [Adherers analysis: HR 0.80 (0.45, 1.43)]
Huang 2015 25125635[102]	CVD 45-75 yo	CeV death, cor revasc, CV death, HF hosp, MI, stroke	As-treated (No)	3 yr 100% (as-treated)	1/36 (2.8)	5/37 (13.5)	OR 0.18 (0.02, 1.65) [D] NNT 9 (NNH 66, NNT 4)
ISAACC 31839558[103]	Admitted to hospital for acute coronary syndrome	CV death or nonfatal events (MI, stroke, HF hospitalization, UA hospitalization, TIA hospitalization)	ITT (Yes)	40 mo 38% overall	98/629 (15.6)	108/626 (17.3)	HR 0.89 (0.68, 1.17) NNT 60 (NNH 41, NNT 17)
		CV death, MI, stroke	ITT (No)	40 mo 36% overall	57/629 (9.1)	50/626 (8.0)	HR 1.13 (0.77, 1.65) NNH 93 (NNH 24, NNT 50)
MOSAIC 23111478[107] 32928788[106]	General 45-75 yo	AFib, angina, cor revasc, CV death, DM, HTN, MI, PVD, stroke, TIA	ITT (No)	2 yr 38% at 6 mo	8/94 (8.5)	17/94 (18.1)	**OR 0.42 (0.18, 0.996) [D]** NNT 10 (NNH 2763, NNT 5)
	General 45-75 yo	Angina, MI, HTN, PVD, AFib, HF, stroke, TIA, cor revasc	ITT (No)	5 yr 40% at 5 yr	25/94 (26.6)	32/97 (33.0)	HR 0.82 (0.45, 1.50) NNT 16 (NNH 15, NNT 5)
PREDICT 25172769[109]	General ≥65 yo	AFib, angina, MI, PVD, stroke, TIA	ITT (No)	1 yr 35% at 1 yr	14/114 (12.3)	17/117 (14.5)	HR 0.87 (0.40, 1.88) NNT 44 (NNH 15, NNT 9)
RICCADSA 26914592[112]	CVD [E] Any age	Cor revasc, CV death, MI, stroke death	ITT (Yes, under-powered)	4.7 yr 60% at 1 yr	22/122 (18.1)	27/122 (22.1)	HR 0.62 (0.34, 1.13) NNT 24 (NNH 17, NNT 7) [**As-treated: HR 0.29 (0.10, 0.86)**]
SAVE 27571048[12]	CVD 45-75 yo	CV death, HF hosp, MI, stroke hosp, TIA, UA	ITT (Yes)	3.7 yr 42% overall	229/1346 (17.0)	207/1341 (15.4)	HR 1.10 (0.91, 1.32) NNH 63 (NNH 23, NNT 83) [Adherers analysis: HR 0.80 (0.60, 1.07)]
		CV death, MI, stroke	ITT (No)	3.7 yr 42% overall	117/1346 (8.7)	120/1341 (8.9)	HR 0.96 (0.74, 1.23) NNT 390 (NNH 53, NNT 42) [HR 0.69 (0.46, 1.04)]
		Angina hosp, CV death, MI, stroke (ischemic), TIA hosp	ITT (No)	3.7 yr 42% overall	207/1346 (15.4)	191/1341 (14.2)	HR 1.07 (0.88, 1.31) NNH 88 (NNH 26, NNT 65) [HR 0.81 (0.59, 1.10)]

Statistically significant results are in bold font.

Abbreviations: AFib = atrial fibrillation, CeV = cerebrovascular, CI = confidence interval, cor revasc = coronary revascularization, CPAP = continuous positive airway pressure (device), CV = cardiovascular, DM = diabetes mellitus, HF = heart failure, hosp = hospitalization, HR = hazard ratio, hr/noc = hours per night, HTN = hypertension, ITT = intention-to-treat, MI = myocardial infarction, MOSAIC = Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular trial, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, PREDICT = trial with undefined acronym, PVD = peripheral vascular disease, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial, TIA = transient ischemic attack, UA = unstable angina, yo = years old (of age).

[A] Based on study eligibility criteria. General = general population, no restriction related to CVD.
[B] Explicitly powered for this outcome.
[C] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[D] Calculated.
[E] Coronary revascularization.

Table 14. CPAP versus no CPAP: Composite cardiovascular outcomes, adjusted nonrandomized comparative studies

Study PMID	Population [A]	Outcome	Analysis (Powered? [B])	Followup Duration	CPAP n/N (%)	No CPAP n/N (%)	Effect Size (95% CI)
Myllylä 2019 30848437[128]	General Any age	Angina, CAD (incident), CAD death, CM death, MI, stroke, stroke death, other CV/vascular death [C]	Use vs. nonuse / discontinued (No)	8.7 yr	148/1030 (14.4)	194/1030 (18.8)	**Adj HR 0.64 (0.5, 0.8) [D]**
Navarro-Soriano 2021 32029279[129]	Resistant hypertension	Fatal or non-fatal CVE, including TIA, stroke, AMI, cor revasc	Adherent vs. nonuse / nonadherent (No)	58 mo NR	32/95 (33.7)	32/68 (47.1)	Adj HR (no CPAP vs. CPAP) 1.6 (0.96, 2.7) [E,F]
Schipper 2017 28550476[133]	General Any age	MI, stroke, TIA	Adherent vs. nonadherent (No)	5.9 yr	6/140 (4.3)	11/101 (10.9)	**Adj HR (no CPAP vs. CPAP) 2.66 (1.20, 5.91) [G,H]**
Sheth 2021, 33729910[134]	Patients with type 2 diabetes	CVD (undefined)	Adherent vs. nonadherent (No)	2.5 yr	27/260 (10.4)	34/318 (10.7)	Adj OR 0.88 (0.50, 1.59) [I,J]
Wu 2015 (Beijing) 25412159[135]	CVD [K] Any age	Cor revasc, death, MI, stent thrombosis, stroke	Adherent vs. nonuse / nonadherent (No)	4.8 yr	40/128 (31.3)	59/167 (35.3)	Adj HR (no CPAP vs. CPAP) 1.22 (0.79, 1.87) [L,M]
		Cor revasc, death, MI, stent thrombosis	Adherent vs. nonuse / nonadherent (No)	4.8 yr	30/128 (23.4)	52/167 (31.1)	Adj HR (no CPAP vs. CPAP) 1.52 (0.94, 2.56) [L,N]

Statistically significant results are in bold font.

Abbreviations: Adj HR = adjusted hazard ratio, AHI = apnea-hypopnea index, BMI = body mass index, CAD = coronary artery disease, CI = confidence interval, CM = cardiomyopathy, cor revasc = coronary revascularization, CPAP = continuous positive airway pressure (device), CV = cardiovascular, CVD = cardiovascular disease, MI = myocardial infarction, TIA = transient ischemic attack.

[A] Based on study eligibility criteria. General = general population, no restriction related to CVD.
[B] Explicitly powered for this outcome.
[C] Death due to aortic dissection, aortic aneurysm, arrhythmia, heart valve disease, atherosclerosis, hypertension, or deep vein thrombophlebitis.
[D] Adjusted for gender, age, AHI, BMI, history of CVD, hypertension, type 2 diabetes, and chronic obstructive pulmonary disease.
[E] Adjusted for age, gender, and BMI.
[F] This HR is for *no CPAP versus CPAP*, opposite of the typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.57 (95% CI 0.30, 1.08).
[G] Adjusted for age, AHI, and smoking.
[H] This HR is for *no CPAP versus CPAP*, opposite of the typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.37 (95% CI 0.13, 1.03).
[I] Adjusted for BMI, age, sex, number of diabetes medications used, history of insulin use, number of hypertension medications used, lipid medication use.
[J] Calculated from reported adjusted OR for *no CPAP versus CPAP*.
[K] Coronary revascularization.
[L] Adjusted for age, sex, BMI, clinical presentation (acute CV event), smoking, hypertension, type 2 diabetes, dyslipidemia, history of and severity of specific CVDs, renal replacement therapy, CV-related medications, AHI.
[M] This HR is for *no CPAP versus CPAP*, opposite of the typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.83 (0.51, 1.36).
[N] This HR is for *no CPAP versus CPAP*, opposite of the typical analysis direction. The calculated (unadjusted) OR for CPAP vs. no CPAP was 0.68 (0.40, 1.14).

Other Health Outcomes (CPAP Versus No CPAP)

Table 15 summarizes the evidence from RCTs and adjusted NRCSs regarding the effect of CPAP (versus no CPAP) on other evaluated long-term health outcomes (other than mortality and CV outcomes). In brief, comparative studies provide at most low SoE for all outcomes. We have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding that the findings are stable.

We conclude that the RCTs do not provide evidence of an effect of CPAP on long-term risk of accidents, incident DM, depression or anxiety symptoms, cognitive function, or measures of QoL. Inclusion of adjusted NRCSs does not alter these conclusions. There is insufficient evidence from comparative studies for HTN outcomes, sexual function, or days of work missed.

Table 15. Evidence profile for CPAP versus no CPAP: Other outcomes

Outcome	Study Design	No. Studies (Subjects)	Risk of Bias	Consistency	Precision	Directness	Other	Overall SoE	Conclusion Statements Estimate (95% CI)
Accidents	**RCT**	**2 (2917)**	**Moderate**	Consistent	**Imprecise**	**Direct**	**Clin heterog**	**Low**	**No evidence of an effect on driving accidents [A]**
Hypertension	**RCT**	**2 (423)**	**High/Moderate**	**N/A**	**Precise**	**Direct**	**Sparse [B]**	**Insufficient**	**No conclusion**
	adj NRCS	1 (163)	High	N/A	Precise	Direct	Sparse	not evaluated	No conclusion
	Overall	*3 (586)*	*High*	*N/A*	*Precise*	*Direct*	*Sparse [C]*	*Insufficient*	*No conclusion*
Diabetes	**RCT**	**3 (3639)**	**Moderate**	**Consistent**	**Imprecise**	**Direct**	**Clin heterog**	**Low**	**No evidence of an effect ES 1.02 (0.69, 1.51) NNH 716 (NNH 29, NNT 45)**
	adj NRCS	1 (266)	High	N/A	Precise	Direct	Sparse	not evaluated	No conclusion
	Overall	*3 (2999)*	*Moderate*	*Consistent*	*Precise*	*Direct*	*Clin heterog*	*Low*	*No evidence of an effect ES 0.88 (0.57, 1.35)*
Depression	**RCT**	**4 (2824)**	**Moderate**	**Consistent**	**Precise**	**Indirect**	**Clin heterog**	**Low**	**No clinically significant improvement in depression symptom score**
Anxiety	**RCT**	**4 (2824)**	**Moderate**	**Consistent**	**Precise**	**Indirect**	**Clin heterog**	**Low**	**No clinically significant improvements in anxiety symptom scores**
Cognitive function	**RCT**	**4 (628) [D]**	**Moderate**	**Inconsistent [E]**	**Imprecise**	**Indirect**	**Clin heterog**	**Low**	**No evidence of a clinically significant effect on MMSE or TMT-A or B [F]**
	adj NRCS	1 (126)	Moderate	N/A	Precise	Indirect	Sparse	not evaluated	No conclusion
	Overall	*5 (754)*	*Moderate*	*Inconsistent [G]*	*Imprecise*	*Indirect*	*Clin heterog*	*Low*	*Same as RCT conclusions*
QoL & functional status	**RCT**	**10 (6460) [H]**	**Moderate-High**	**PCS: Inconsistent Others: Consistent**	**Precise**	**Indirect**	**None**	**Low**	**PCS: No effect. MCS, EuroQol-5D, and SAQLI: Small, not clinically significant improvements [I]**
	Adj NRCS	1 (214)	High	N/A	Precise	Indirect	Sparse	not evaluated	No conclusion
	Overall	*11 (6874) [H]*	*Moderate-High*	*Consistent*	*Precise*	*Indirect*	*None*	*Low*	*Same as RCT conclusions*
Sexual function	**RCT**	**1 (43)**	**Moderate**	**N/A**	**Precise**	**Indirect**	**Sparse**	**Insufficient**	**No conclusion**
	adj NRCS	2 (265)	High	Inconsistent	Precise	Indirect	None	not evaluated	No conclusion
	Overall	*3 (308)*	*High*	*Inconsistent*	*Precise*	*Indirect*	*None*	*Insufficient*	*No conclusion*
Days of work missed	**RCT**	**1 (2687)**	**Moderate**	**N/A**	**Precise**	**Direct**	**Sparse**	**Insufficient**	**No conclusion**

Abbreviations: adj NRCS = adjusted nonrandomized comparative studies, CI = confidence interval, Clin heterog = clinically heterogeneous (e.g., different populations, eligibility criteria, specific outcomes), ES = (summary) effect size, MCS = Short Form (SF) 36 Mental Component Summary, N/A = not applicable, NNH = (summary) number-needed-to-harm, NNT = (summary) number-needed-to-treat, PCS = Short Form (SF) 36 Physical Component Summary, QoL = quality of life, RCT = randomized controlled trials, SAQLI = Sleep Apnea Quality of Life Index, SoE = strength of evidence.

Evaluations of RCT evidence base are in bold font. Evaluations of all studies (RCTs and adjusted NRCSs together) are in italic font.

[A] Odds or rate ratios for various specific outcomes ranged from 0.52 (95% CI 0.05 to 5.87) to 0.86 (95% CI 0.57 to 1.32), with NNT ranging from 13 to 218, all statistically nonsignificant.

[B] One trial each reported on incident hypertension and resolution of hypertension.

[C] One study each reported on incident hypertension, resolution of hypertension, and hypertensive crises.

[D] A fifth RCT (n = 1098) evaluated unique measures of cognitive function and thus, provided only insufficient evidence.

[E] Although all effects on MMSE and TMT-B were small and statistically nonsignificant, the net effect on MMSE favored CPAP in direction while the net effect on TMT-B favored no CPAP.

[F] Insufficient evidence for other measures of cognitive function, which were evaluated by only one study each.

[G] Although all effects on MMSE and TMT-B were small and statistically nonsignificant, the net effect on MMSE favored CPAP in direction while the net effect on TMT-B favored no CPAP.

[H] Across all QoL and functional status measures.

[I] Insufficient evidence regarding functional status as measured by FOSQ (one RCT).

Accidents

Two RCTs reported on various outcomes pertaining to accidents. Both the PREDICT and SAVE trials are described above, in the section *Mortality and Cardiovascular Outcomes*. Both trials were deemed to be at moderate risk of bias related to lack of participant blinding and clinician bias; although, outcome assessors were blinded. Neither trial was explicitly powered for accident-related outcomes. Both were evaluated as ITT analyses. Adherence was poor in both trials: about 42 percent, overall, in SAVE and 35 percent at 1 year in PREDICT. As the likelihood of accidents is probably associated with degree of sleepiness, it is worth noting that the SAVE trial excluded patients with extreme daytime sleepiness (ESS >15) and while PREDICT included only patients with excessive daytime sleepiness (ESS ≥9), they excluded professional drivers and people reporting sleepiness while driving. No NRCSs that met eligibility criteria (confounder-adjusted parallel cohorts) reported on accidents.

Across specific outcome measures (Table 16), the PREDICT trial,[109] at 1 year followup, and the SAVE trial,[12] at 3.7 years followup, found no statistically significant differences between groups using CPAP or not. The SAVE trial reported multiple reported specific road accident outcomes, which were all secondary endpoints. One of these, the annual rate of accident-causing injuries, was near significant (P = 0.06), favoring CPAP. The PREDICT trial reported no change in rates of accidents during the 1 year of followup.

Neither trial reported subgroup analyses. The trials did not report on correlation between adherence and accident rates.

Summary of Effect of CPAP on Accidents

With low SoE, two RCTs fail to provide evidence of an effect of CPAP on risk of traffic or driving accidents in the long term (Table 15). This conclusion does not imply that CPAP has been proven to be ineffective to reduce traffic or driving accidents. The trials are at moderate risk of bias and yielded imprecise estimates. The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect. The evidence was insufficient regarding other types of accidents.

Table 16. CPAP versus no CPAP: Accidents

Study PMID Design	Design	Outcome	Analysis (Powered? [A])	Followup Duration (mo)	Arm	Baseline Rate	Final Rate	Effect Size (95% CI) Calculated NNT (95% CI) [B]
PREDICT 25172769[109]	RCT	Driving accidents	ITT (No)	12	CPAP No CPAP	1/73 (1.4) 2/77 (2.6)	2/73 (2.7) 1/77 (1.3)	OR 0.52 (0.05, 5.87) [C] NNT 81 (NNH 31, NNT 18)
.		Home accidents	ITT (No)	12	CPAP No CPAP	12/117 (6.7) 6/113 (5.3)	9/113 (8.0) 18/117 (15.4)	"Treatment effect" 0.49 (0.21,1.18) [E] NNT 13 (NNH 124, NNT 6)
SAVE 27571048[12]	RCT	Traffic accidents, participants	ITT (No)	44.4 (mean)	CPAP No CPAP	NR NR	41/1346 (3.0%) 47/1341 (3.5%)	OR 0.86 (0.57, 1.32) [C] NNT 218 (NNH 113, NNT 55)
.		Traffic accidents, annual rate	ITT (No)	44.4 (mean)	CPAP No CPAP	NR NR	56 (1.1%/yr) 70 (1.4%/yr)	Rate ratio 0.78 (0.55, 1.11) NNT not calculable
.		Accident-causing injury, participants	ITT (No)	44.4 (mean)	CPAP No CPAP	NR NR	99/1346 (7.4%) 118/1341 (8.8%)	OR 0.82 (0.62, 1.09) [C] NNT 69 (NNH 162, NNT 29)
.		Accident-causing injury, annual rate	ITT (No)	44.4 (mean)	CPAP No CPAP	NR NR	219 events/1346 (4.4%/yr) 255 events/1341 (5.2%/yr)	Rate ratio 0.84 (0.70,1.00) [D]

Abbreviations: . = no information, CI = confidence interval, CPAP = continuous positive airway pressure (device), ITT = intention-to-treat, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, PREDICT = trial with undefined acronym, SAVE = Sleep Apnea cardioVascular Endpoints trial.

[A] Explicitly powered for this outcome.

[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).

[C] Calculated.

[D] P = 0.06.

[E] Adjusted for accident history at baseline.

Hypertension

Two RCTs and one NRCS compared the effect of CPAP versus no CPAP on three distinct HTN outcomes: incident HTN (Barbé 2012[99]), reversion to normotension (Huang 2015[102]), and hypertensive crisis[129]). All studies are described above, in the section *Mortality and Cardiovascular Outcomes*. Neither RCT was powered for HTN, although Barbé 2012 was powered for the primary outcome combined incident HTN (among participants with normotension at enrollment) or incident CV events.

The Barbé 2012 RCT reported both an ITT and a CPAP adherers analysis, with 59 percent of analyzed participants adherent with CPAP. The study was deemed to be at high risk of bias due to failure to account for baseline differences in AHI between groups and lack of participant or clinician blinding (although outcome assessors were blinded). The Huang 2015 RCT conducted an "as-treated" comparison of adherent CPAP users and participants not prescribed CPAP; it was deemed to be at moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded). The Navarro-Soriano 2021 NRCS also compared adherent CPAP users versus nonadherent users and nonusers; it was rated as high risk of bias for use of a regression model for adjustment of baseline confounders.

Incident Hypertension

Barbé 2012 reported on the incidence of HTN after a 4-year followup period (Table 17). The outcome was analyzed only among the 48 percent of included participants who did not have HTN at baseline. Almost half of participants developed HTN during the study period. Rates of incident HTN were 19.8 events per 100 person-years in the group prescribed CPAP and 17.6 events per 100 person-years in the no CPAP group, with no statistically significant difference between the two. The rate ratio was 0.89 (95% CI 0.65 to 1.21; NNT = 48). An analysis of just the adherent CPAP users yielded a similar result (see Table 17). The trial did not report subgroup analyses.

Resolution of Hypertension

Huang 2015 reported on HTN resolution (or reversion to normotension) at a median followup of 36 months (Table 17). All participants had uncontrolled HTN (and CAD) to be eligible. Participants were treated with antihypertensive and CVD drugs treatments based on current guidelines during a 3-month run-in period and informed not to change their drug regimen without physician approval. On average, participants were on 3.2 antihypertensive drugs. About half the participants (47%) had some change in their drug regimen during the study, with similar rates of change between the two groups. Rates of HTN resolution were 69.4 percent in the CPAP group and 43.2 percent in the no CPAP group. In a *post hoc* analysis, participants adherent with CPAP were statistically significantly more likely to revert to normotension than nonusers (OR 2.98, 95% CI 1.14 to 7.81; NNT = 4). The trial did not report subgroup analyses.

Hypertensive Crisis

The Navarro-Soriano 2021 NRCS evaluated hypertensive crisis, which was defined as elevation of systolic BP >179 mmHg or diastolic BP >109 mmHg. Patients were followed for a median of 58 months. The study was a secondary analysis of adherent versus nonadherent CPAP users (or nonusers) from an existing RCT that randomized patients to CPAP versus no CPAP. Patients included in the original RCT had at least moderately severe OSA (AHI ≥15) and

resistant HTN. Patients were taking, on average, about three antihypertensive medications. Nonadherent CPAP users were significantly more likely to have a hypertensive crisis event yielding an adjusted HR (for *no CPAP vs CPAP*) of 5.1 (95% CI 2.2 to 11.6). The study did not report subgroup analyses.

Heterogeneity of Treatment Effect (HTN)

The studies did not report subgroup analyses for any of the analyzed outcomes, which were addressed by a single study each.

Applicability (HTN)

The study reporting on incident HTN is most applicable to middle-aged (≤70 years, mean 52 years) otherwise healthy adults (who do not have HTN). The study reporting on resolution of HTN evaluated patients with uncontrolled HTN and CAD who were adherent with CPAP. The study reporting on hypertensive crises evaluated patients with resistant HTN who were adherent with CPAP.

Summary of Effect of CPAP on Hypertension

With only a single RCT evaluating each outcome, there is insufficient evidence to determine the effect of CPAP on risk of incident HTN, reversion to normotension, or hypertensive crises (Table 15).

Table 17. CPAP versus no CPAP: Hypertension

Study PMID	Design	Outcome	Analysis (Powered? [A])	Followup Duration (mo)	Arm	CPAP Adherence	n/N (%)	Effect Size (95% CI) Calculated NNT (95% CI) [B]
Barbé 2012 22618923[99]	RCT	Incident HTN	ITT (No)	48	CPAP (ITT) No CPAP	59% [C]	19.8 (15.3, 24.2) events/100 p-yr [75/167 (44.9)] 22.2 (17.5, 26.9) events/100 p-yr [86/183 (47.0)]	Rate ratio 0.89 (0.65, 1.21) NNT 48 (NNH 12, NNT 8)
	.	.	Adherent analysis	48	CPAP use	100%	17.6 (12.4, 22.8) events/100 p-yr [44/98 (44.9)]	Rate ratio 0.79 (0.55, 1.14)
Huang 2015 25125635[102]	RCT	Resolution of HTN	As-treated (No)	36 (median)	CPAP use No CPAP use	100%	25/36 (69.4) 16/37 (43.2)	**OR 2.98 (1.14, 7.81) [D]** NNT 4 (2, 23)
Navarro-Soriano, 2021, 32029279[129]	NRCS	Hypertensive crisis	Adherent vs. nonuse / nonadherent (No)	58 (median)	CPAP adherent No CPAP use	NR	8/95 (8.4) 23/68 (33.8)	*No CPAP vs. CPAP* **Adj HR 5.1 (2.2, 11.6)**

Statistically significant results are in bold font.

Abbreviations: . = no information, CI = confidence interval, CPAP = continuous positive airway pressure, NNH = number needed to harm, NNT = number needed to treat, OR = odds ratio, p-yr = person-years, PMID = PubMed identifier.

[A] Explicitly powered for this outcome.
[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).
[C] Among analyzed participants without hypertension at study baseline.
[D] Calculated.

Diabetes Mellitus

Four studies (3 RCTs[12, 102, 103] and 1 NRCS[121]) compared the effect of CPAP versus no CPAP on risk of incident type 2 DM. No eligible study reported on reversion to normoglycemia or incidence of other hyperglycemia diagnoses (e.g., impaired glucose tolerance, metabolic syndrome). None of the studies provided details about diagnostic parameters for DM, including how rigorously or systematically participants were monitored for DM or reversion to normoglycemia.

Randomized Controlled Trials (DM)

The three RCTs (Huang 2015, ISAACC, and SAVE) are described above, in the section *Mortality and Cardiovascular Outcomes*. They were all deemed to be at moderate risk of bias related to lack of participant and clinician bias; although, outcome assessors were blinded. None was explicitly powered for DM-related outcomes. All analyses of incident DM were restricted to trial participants without DM at enrollment.

Two RCTs, SAVE and, particularly, Huang 2015, provided imprecise estimates, with no significant difference in risk of incident DM between groups at about 3 to 4 years of followup. The ISAACC trial was also nonsignificant, but incident DM was more common among patients assigned to CPAP. The SAVE trial found similar estimates in its ITT and CPAP adherers analyses (see Table 18). The study did not report an analysis of the association between adherence and risk of DM. Across the three trials, CPAP did not affect the risk of incident DM (summary ES 1.02, 95% CI 0.69 to 1.51), with moderate heterogeneity ($I^2 = 40\%$).

Adjusted Nonrandomized Comparative Study (DM)

The **Botros 2009** NRCS compared US veterans who regularly (undefined) using CPAP (n = 160) with those not regularly using CPAP (n = 106), including not ordered, not used, intermittent use.[121] The NRCS may have been subject to uncontrolled biases related to self-selection of CPAP use and adherence. Potential participants with a diagnosis of DM were excluded. Incident DM was the primary outcome of the study. Included participants had to have at least "moderately severe" OSA with an AHI ≥20. The mean age of the group was 63 years and participants were, on average obese, with a mean BMI of 35. Mean AHI was 55.6 and mean baseline ESS was 8.3. The veterans were 93 percent male. Most participants had HTN (70%). The average fasting blood glucose of the patients enrolled was 100.2 mg/dL. The study was rated as high risk of bias related to inadequate reporting regarding OSA primarily because the study combined patients not prescribed CPAP with those who were nonadherent with CPAP.

The adjusted NRCS found a significantly lower association between regular CPAP use and risk of incident DM at about 3 years of followup.

All Studies (DM)

The summary effect size across the RCTs was nonsignificant. With addition of the adjusted NRCS the summary estimate remained nonsignificant and imprecise (ES 0.88, 95% CI 0.57 to 1.35), with moderate heterogeneity ($I^2 = 57\%$) (Figure 8).

Heterogeneity of Treatment Effect (DM)

No study reported subgroup effects (or interactions with other factors). No difference was apparent between the adherent analyses and the ITT analysis, either within or between studies.

The three studies did not differ significantly differ in their findings (although, the NRCS was statistically significant while the RCTs were not, or in one case, was highly imprecise).
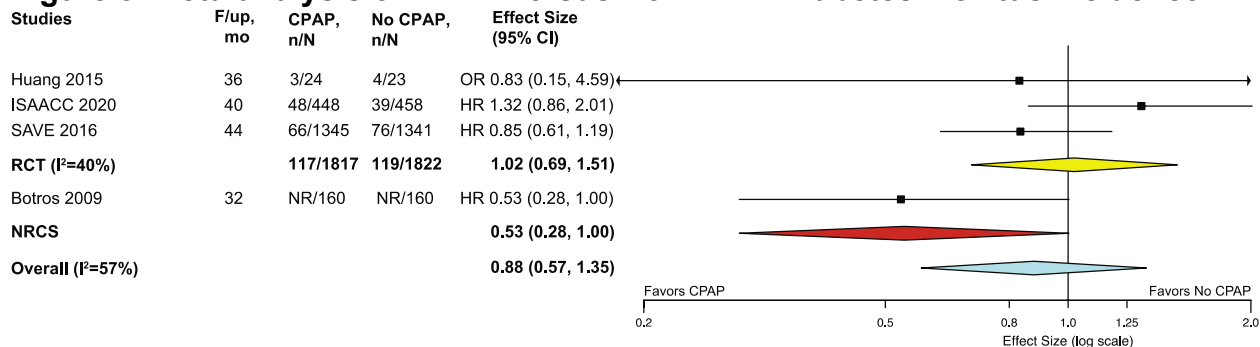
Applicability (DM)

The RCT evidence was mostly applicable to people who would be eligible for enrollment in the SAVE or ISAACC trials, i.e., adults with moderate or severe OSA and preexisting CAD or CeVD, including acute coronary syndrome. The NRCS was conducted among US veterans (93% male) and, thus, may be of more limited applicability to the general population.

Summary of Effect of CPAP on Diabetes

Three RCTs provide low SoE that CPAP does not affect the risk of incident type 2 DM (Table 15). This conclusion does not imply that CPAP has been proven to be ineffective to reduce incident DM. The trials were at moderate risk of bias and provided a null estimate of effect (summary effect size 1.02, 95% CI 0.69 to 1.51). Under an assumption of a control rate of 7.6 percent (by meta-analysis), the summary NNH was 716 (95% CI NNH 29 to NNT 45). Inclusion of one high risk of bias adjusted NRCS does not alter the conclusion. The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of the effect is close to the true effect.

Figure 8. Meta-analysis of CPAP versus No CPAP: Diabetes mellitus incidence

Studies	F/up, mo	CPAP, n/N	No CPAP, n/N	Effect Size (95% CI)
Huang 2015	36	3/24	4/23	OR 0.83 (0.15, 4.59)
ISAACC 2020	40	48/448	39/458	HR 1.32 (0.86, 2.01)
SAVE 2016	44	66/1345	76/1341	HR 0.85 (0.61, 1.19)
RCT (I²=40%)		**117/1817**	**119/1822**	**1.02 (0.69, 1.51)**
Botros 2009	32	NR/160	NR/160	HR 0.53 (0.28, 1.00)
NRCS				**0.53 (0.28, 1.00)**
Overall (I²=57%)				**0.88 (0.57, 1.35)**

Abbreviations: CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, ISAACC = Impact of Sleep Apnea syndrome in the evolution of Acute Coronary syndrome. Effect of intervention with CPAP, NR = not reported, NRCS = nonrandomized comparative study, RCT = randomized controlled trial, SAVE = Sleep Apnea cardioVascular Endpoints trial.

Table 18. CPAP versus no CPAP: Diabetes mellitus

Study PMID	Study Design	Analysis (Powered? [A])	Timepoint (mo)	Arm	CPAP Adherence	n/N (%)	Effect size (95% CI) Calculated NNT (95% CI) [B]
Huang 2015 25125635[102]	RCT	As-treated (No)	36 (median)	CPAP No CPAP	100% (as-treated)	3/24 (12.5) 4/23 (17.4)	OR 0.83 (0.15, 4.78) [C] NNT 20 (NNH 6, NNT 4)
ISAACC, 31839558[103]	RCT	ITT (No)	40	CPAP No CPAP	38% overall	48/448 (10.7) 39/458 (8.5)	HR 1.32 (0.86, 2.01) NNH 45 (NNH 17, NNT 61)
SAVE 27571048[12]	RCT	ITT (No)	44.4 (mean)	CPAP No CPAP	42% overall	66/1345 (4.9) 76/1341 (5.7)	HR 0.85 (0.61, 1.19) NNT 132 (NNH 107, NNT 41) [Adherent analysis: HR 0.77 (0.48, 1.27)]
Botros 2009 19958890[121]	NRCS	Adherent vs. nonuse / nonadherent (No)	32.4 (mean)	CPAP adherent No CPAP	100% (adherent arm)	NR/160 NR/160	**Adj HR 0.53 (0.28, 0.99)**

Abbreviations: Adj = adjusted, CI = confidence interval, CPAP = continuous positive airway pressure, HR = hazard ratio, NNH = number needed to harm, NNT = number needed to treat, NRCS = nonrandomized comparative study, OR = odds ratio, PMID = PubMed identifier, RCT = randomized comparative trial.

[A] Explicitly powered for this outcome.

[B] Calculated only for randomized trials, not for crude rates from nonrandomized comparative studies. NNH in the 95% CI range suggests that one harm (poor outcome) may occur per the reported number of people treated (within the 95% CI).

[C] Calculated, included only participants without diabetes at baseline. Also including the assumption that everyone with diabetes at baseline also had the diagnosis at followup.

Mental Health

Four RCTs reported on the effect of CPAP versus no CPAP on either depression or anxiety symptoms.[93, 109-111, 117] No study addressed diagnoses (or resolution) of any mental health conditions (e.g., incident anxiety disorder) or other mental health conditions. The PREDICT trial, RICCADSA, and SAVE trials are described above, in the section *Mortality and Cardiovascular Outcomes*. Briefly, PREDICT included older adults (≥65 years) with newly diagnosed OSA with sleepiness symptoms, RICCADSA included adults with nonsleepy OSA with a history of coronary revascularization, and SAVE included adults with OSA and a history of CVD or CeVD. PREDICT and SAVE were deemed to be at moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded). RICCADSA was rates as high risk of bias due to lack of blinding and high crossover (20% of participants in the no CPAP arm started CPAP during the study period).

Aarab 2017 was an additional RCT that did not report CV outcomes (and was, thus, not described above). This was primarily a trial comparing CPAP with an oral appliance, which included a placebo (sham) oral appliance group.[93] Here we focus on only the comparison of CPAP versus no CPAP. The study reported an ITT analysis. The RCT included adults with at least "mild" OSA (AHI ≥5) and sleepiness symptoms (ESS ≥10 or ≥2 sleepiness symptoms). Participants (assigned to either CPAP, n = 18, or placebo, n = 19) were, on average, about 52 years old, 70 percent male, had a mean AHI about 20 and mean ESS about 10.7. Participants were mostly obese with a mean BMI of about 31. The study was powered for changes in AHI, not explicitly for any health outcome. Those prescribed CPAP were reported to use their device 83 percent of nights. The study was rated as moderate risk of bias due to lack of participant or clinician blinding; the mental health outcomes were based on self-reported completion of the disease non-specific Symptom Checklist-90-Revised (SCL-90-R).

The studies evaluated depression and anxiety outcomes, but not other mental health conditions. None of the studies incorporated clinician assessment or diagnosis of either depression or anxiety, thus there are no data on incident diagnoses or resolution of either diagnosis. SAVE and RICCADSA used threshold values in mental health symptom scales to define "cases" of depression or anxiety, or of "depressive mood," but these referred only to self-described symptoms, not diagnoses. We omitted these outcomes.

In brief, three RCTs compared treatment with CPAP with no treatment and the fourth RCT, Aarab 2017, compared real and sham CPAP. All studies were evaluated as ITT analyses of CPAP prescription. None of the trials was explicitly powered for a mental health outcome.

Depression Symptoms Scales

All four trials reported on depression outcomes. As noted, Aarab 2017 used the SCL-90-R depression component. The PREDICT and SAVE trials evaluated the Hospital Anxiety and Depression Scale (HADS) test, and RICCADSA evaluated the Zung Self-rating Depression Scale (SDS). The three scales have been validated in a variety of populations, but evidently not in patients with OSA. The scales were designed for cross-sectional, not longitudinal, evaluation, at least in part, of depression symptoms. The interpretation of changes over time in scores, though, is unclear. None of the trials described the validity of the scales in the OSA population or as an assessment of change in symptoms. We found no information on minimal clinically important differences specific to patients with OSA. However, we evaluate this metric based on studies conducted in other populations, as noted below.

The four RCTs all reported changes in depression scores (Table 19). Overall, depression scores decreased (improved) over time in all studies among both those using CPAP or no CPAP. Among analyses of all participants (regardless of baseline diagnosis of depression), only the SAVE trial found a statistically significant relative improvement in depression scores (after 4 years) among those receiving CPAP compared with controls. However, based on a minimal clinically important difference in HADS depression score of 1.9 (derived from a separate analysis of patients with CVD[139]), none of the effects on depression in the SAVE trial were clinically significant. Trials were conducted as ITT analyses. The Aarab 2017 also reported that their per protocol analyses resulted in similar estimates; although, the study had an atypically high adherence rate (83% of nights).

Meta-analysis of the standardized mean differences (SMD) across trials (to account for the different scales used) yielded a statistically significant summary SMD of −0.18 (95% CI −0.42 to −0.87), which may be interpreted as a small effect size favoring CPAP (Figure 9). However, the meta-analysis largely recapitulates the SAVE trial, given its relatively large sample size.

The RICCADSA trial reported on the effect of CPAP on the SDS depression score specifically among participants with symptoms of depression at the start of the study. The trial found a large statistically significant difference in improvement in SDS depression scores between groups of −7.9 (95% CI −11.5 to −4.3). We did not find information about a minimal clinically important difference in nonsurgical populations.

Anxiety Symptom Scale

All four trials reported on anxiety scores. Aarab 2017 used the SCL-90-R anxiety component. The PREDICT and SAVE trials evaluated the HADS test. RICCADSA evaluated the Zung Self-rating Anxiety Scale (SAS). Similar to assessment of depression symptoms, the anxiety symptom scales have been validated in a variety of populations, but evidently not in patients with OSA. The scales were designed for cross-sectional, not longitudinal, evaluation, at least in part, of anxiety symptoms. The interpretation of changes over time in scores, though, is unclear. None of the trials described the validity of the scales in the OSA population or as an assessment of change in symptoms. We found no information on minimal clinically important differences specific to patients with OSA. However, we evaluate this metric based on studies conducted in other populations, as noted below.

The four RCTs all reported changes in anxiety scores (Table 19). Overall, anxiety scores decreased (improved) over time in all studies among both those using CPAP or no CPAP. Only the SAVE trial found a statistically significant relative improvement in anxiety scores (after 4 years) among those receiving CPAP compared with controls. However, based on a minimal clinically important difference in HADS anxiety score of 1.7 (derived from a separate analysis of patients with CVD[139]), none of the effects on anxiety in the SAVE trial were clinically significant. RICCADSA also reported that, among patients with *no anxiety at baseline* (n =171, total), 11 in the CPAP group, and 7 in the no-CPAP group demonstrated SAS score ≥45 (defined as having a clinically significant anxiety) at 12 month-follow-up, although complete data were not reported, the difference was nonsignificant.[111] As for anxiety severity, all trials were conducted as ITT analyses. Aarab 2017 and RICCADSA reported that their per protocol analyses resulted in similar estimates.

Meta-analysis of the SMD across trials (to account for the different scales used) yielded a statistically significant summary SMD of −0.10 (95% CI −0.17 to −0.03), which may be

interpreted as a small effect size favoring CPAP (Figure 10). However, the meta-analysis largely recapitulates the SAVE trial, given its relatively large sample size.

Heterogeneity of Treatment Effect (Mental Health)

None of the studies reported subgroup analyses (except for the subgroup analyses of patients with depression symptoms at baseline or no anxiety at baseline, described above). The studies yielded statistically similar results. No differences in effects were apparent across studies.

Applicability (Mental Health)

Given the large relative sample size of the SAVE trial, the findings are most applicable to patients who would meet eligibility criteria for the SAVE trial (adults with severe OSA, with ODI ≥12, and preexisting CAD or CeVD).
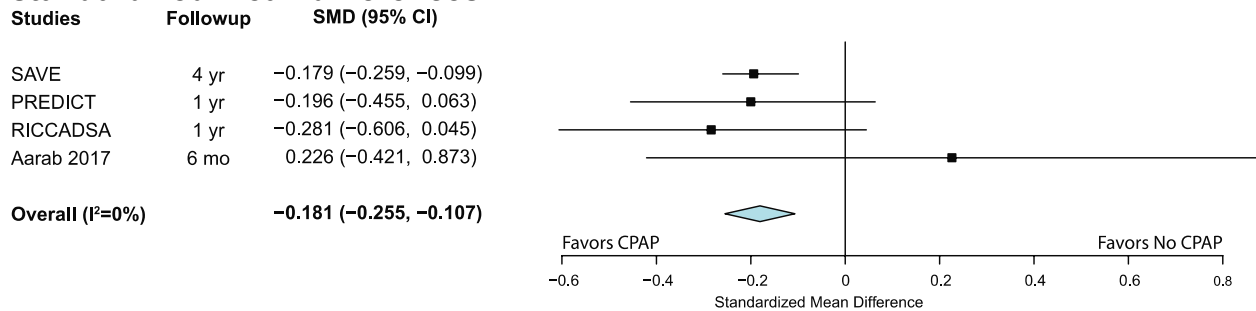
Summary of Effect of CPAP on Mental Health Symptoms

There is no evidence regarding the effect of CPAP on incident depression or anxiety.

Four RCTs provide low SoE that CPAP does not affect depression or anxiety symptom scores by a clinically significant degree (in all patients regardless of baseline symptoms) (Table 15). Although the trials suggest small improvement with CPAP, estimates of effect were less than clinically significant thresholds in all studies. In addition, the measures of symptoms have not been validated in an OSA population or for their interpretation for assessing longitudinal change in scores with treatment; thus, the outcomes are considered "indirect." The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

A single trial provides insufficient evidence regarding the effect of CPAP on depression scores specifically among patients with baseline depression symptoms.

Figure 9. Meta-analysis of CPAP versus No CPAP: Depression scores, standardized mean differences

Studies	Followup	SMD (95% CI)
SAVE	4 yr	−0.179 (−0.259, −0.099)
PREDICT	1 yr	−0.196 (−0.455, 0.063)
RICCADSA	1 yr	−0.281 (−0.606, 0.045)
Aarab 2017	6 mo	0.226 (−0.421, 0.873)
Overall (I²=0%)		−0.181 (−0.255, −0.107)

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, PREDICT = trial with undefined acronym, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial, SMD = standardized mean difference.

93

Figure 10. Meta-analysis of CPAP versus No CPAP: Anxiety scores, standardized mean differences

Studies	Followup	SMD (95% CI)
SAVE	4 yr	−0.113 (−0.193, −0.033)
PREDICT	1 yr	−0.055 (−0.313, 0.204)
RICCADSA	1 yr	0 (−0.272, 0.272)
Aarab 2017	6 mo	−0.022 (−0.666, 0.623)
Overall (I²=0%)		**-0.099 (-0.172, -0.026)**
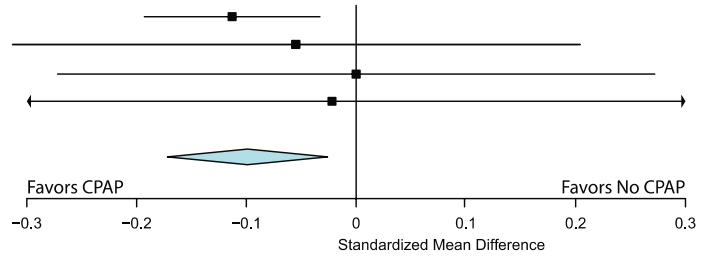
Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, PREDICT = trial with undefined acronym, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial, SMD = standardized mean difference.

94

Table 19. CPAP versus no CPAP RCTs: Depression and anxiety scores

Study PMID [A]	Outcome	Scale (Direction) [MCID]	Followup Duration	Arm	CPAP Adherence	N Analyzed (Baseline / Followup)	Mean Diff (SD), Within-Arm	Net Diff (95% CI), Between-Arm
Aarab 2017 28083705[93]	SCL−90−R: Depression	NR (lower better) [NR]	6 mo	CPAP No CPAP	83% of nights [B]	18/18 19/19	−3.1 (15.9) −6.7 (15.3)	3.6 (−6.5, 13.7) [C]
PREDICT 25172769[109]	HADS: depression component	0-21 (lower better) [1.9]	1 yr	CPAP No CPAP	35% at 1 yr	123/114 124/116	−0.7 (3.0) −0.2 (3.1)	−0.4 (−1.0, 0.3)
RICCADSA 30130421[110] 33341644[111]	Zung SDS: depression (all participants)	25-125 (lower better) [NR]	1 yr	CPAP No CPAP	60%	69/69 78/78	−3.2 (8.9) −0.6 (9.5)	−2.6 (−5.6, 0.4) [C]
	Zung SDS: depression (depression at baseline)	25-125 (lower better) [NR]	1 yr	CPAP No CPAP	60%	30/30 26/26	−9.2 (6.8) −1.3 (7.0)	**−7.9 (−11.5, −4.3) [C]**
SAVE 31312807[117]	HADS: depression component	0-21 (lower better) [1.9]	4 yr	CPAP No CPAP	42% overall	1341/1220 1336/1190	−0.8 (4.0) −0.1 (3.8)	**−0.8 (−1.0, −0.5) [D]**
Aarab 2017 28083705[93]	SCL−90−R: anxiety	NR (lower better) [NR]	6 mo	CPAP No CPAP	83% of nights [B]	1819/1819	−1.7 (9.4) −1.5 (8.6)	−0.2 (−0.9, 0.5)
PREDICT 25172769[109]	HADS: anxiety component	0-21 (lower better) [1.9]	1 yr	CPAP No CPAP	35%	123/114 124/117	−1.2 (3.7) −1.0 (3.6)	−0.2 (−0.9, 0.5)
RICCADSA 30130421[110] 33341644[111]	Zung SAS: anxiety (all participants)	25-100 (lower better) [NR]	1 yr	CPAP No CPAP	60%	103/103 105/105	-1.2 (6.6) -1.2 (8.5)	0 (-2.1, 2.1) [C]
	Zung SAS: anxiety (no anxiety at baseline)	25-100 (lower better) [NR]	1 yr	CPAP No CPAP	60%	11/NR 7/NR	NR NR	Reported only as NS
SAVE 27571048[12]	HADS: anxiety component	0-21 (lower better) [1.7]	4 yr	CPAP No CPAP	42% overall	1341/1220 1336/1190	−0.8 (3.6) −0.4 (3.5)	**−0.4 (−0.6, −0.2) [D]**

Statistically significant results are in bold font.

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, HADS = Hospital Anxiety and Depression Scale, MCID = minimal clinically important difference, Mean Diff = mean difference, Net Diff = net difference (difference-in-difference), NR = not reported, OR = odds ratio, PMID = PubMed identifier, PREDICT = trial with undefined acronym, RCTs = randomized controlled trials, RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAS = (Zung) Self-rating Anxiety Scale, SAVE = Sleep Apnea cardioVascular Endpoints trial, SCL−90−R = Symptom Checklist-90-Revised, SD = standard deviation, SDS = (Zung) Self-rating Depression Scale.

[A] All studies were randomized controlled trials, analyzed by intention-to-treat, and were not explicitly powered for mental health outcomes.
[B] Self-report. Study included frequent visits (initially about every 2 weeks)
[C] Calculated.
[D] Adjusted for baseline score.

Cognitive (Executive) Function

Six studies (5 RCTs[97, 104, 108, 109, 119] and 1 NRCS[124]) reported on cognitive executive function. No study evaluated incident dementia or other conditions related to cognitive function.

The PREDICT trial is described in the *Mortality and Cardiovascular Outcomes* section above.[109] In brief, this 1-year trial included older adults (≥65 years) with newly diagnosed OSA (ODI ≥7.5) and sleepiness symptoms (ESS ≥9). The study was rated to be at moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded).

The remaining studies did not address already-covered outcomes and are, thus, described here. The **APPLES** trial (Apnea Positive Pressure Long-term Efficacy Study) randomly allocated participants to active (n = 556) or sham (n = 542) CPAP for 6 months.[97] The study reported an ITT analysis. Participants were included if they had an AHI >10. The mean baseline AHI was 40 among participants, who were mostly (65%) male. The mean age of participants was 51.5 years and was similar across groups. Patients were mostly obese, with a mean BMI of 32.3. At 6 months, 42 percent of CPAP users were adherent (≥4 hours/night, 70% of nights) and mean CPAP usage was 4. 7 (SD 2.1) hours per night. This double-blind trial, which used sham CPAP, was deemed to be at low risk of bias. The study was explicitly adequately powered for neurocognitive outcomes, specifically for a difference in the Pathfinder Number Test (based on prior pilot studies) since it required the largest sample size. However, the trial reported on unique cognitive measures among studies reporting on cognitive function.

Monasterio 2001 randomized participants with OSA to CPAP (n = 66) or no CPAP (n = 59) and followed them for 6 months.[104] The study reported an ITT analysis. Participants were included if they had "moderately severe" OSA (AHI between 10 and 30; thus excluding patients with "severe" OSA), without severe daytime sleepiness. The study also excluded patients with "notable cardiovascular disease," not otherwise defined. The mean baseline AHI was about 20.5 and mean baseline ESS was about 12.5. The study sample was 86 percent male and participants were on average 53.5 years old. Patients were mostly obese, with a mean BMI of 29.4. The trial was powered for a change in ESS, not explicitly for any health outcome. At 6 months, mean CPAP usage was 4.8 hours per night. The study reported that 8 percent of participants abandoned CPAP use and another 14 percent withdrew from the study (for unreported reasons); 62 percent of study participants chose to continue CPAP after study termination. The study was rated as moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded).

Pelletier-Fleury 2004 randomized adults (≤70 years) with OSA (AHI ≥10) to CPAP (n = 82) or no CPAP (n = 89).[108] Participants who stopped CPAP use before 6 months were excluded, thus, an "as-treated" analysis was conducted. About two-thirds (63%) of patients had severe OSA (AHI ≥30), with similar overall mean AHI between groups (AHI 52). Mean BMI (30.0), mean age (53), and percent male (83%) were similar between groups. On average, at 6 months CPAP users with less severe OSA used their devices 4.8 hours per night and those with severe OSA (AHI ≥30) used their devices 5.5 hours per night. The trial did not report a power analysis. The study was rated as high risk of bias, due to lack of information on allocation concealment, blinding, and about 25 percent loss to followup.

Wu 2016 (Yangzhou) randomized participants to CPAP (n = 68) or no CPAP (n = 68) for 6 months.[119] Participants were included if they had an AHI ≥15. The study reported an ITT analysis. The participants had particularly "severe" OSA, with a mean baseline AHI was 61.0; 90 percent were male. The study excluded patients with CVD or CV risk factors (HTN, DM,

dyslipidemia) or "severe visceral diseases" (not otherwise defined), all of which the study states could affect cognitive function, in addition to excluding those with a BMI ≥35. Nevertheless, patients were mostly obese, with a mean BMI of 28.0. The mean age of participants was 49.6 years (range 30 to 65 years). The study did not report power calculations. The study did not report adherence information. Overall, the study was rated as high risk of bias due to lack of information about randomization, allocation, concealment, and blinding, and because of the apparent mismatch between AHI eligibility criteria (≥15) and mean AHI (61).

The NRCS by **Crawford-Achour 2015** reported an adjusted comparison of older adults (≥65 years) who used CPAP (n = 33) or did not use CPAP (n = 93) for 10 years.[124] The study conducted an "as-prescribed" analysis based on treatment/prescription with CPAP versus no treatment, and all participants were analyzed. Participants were included if they had "severe" OSA (AHI >30). The two arms had significantly different mean baseline AHI (CPAP 49.0, no CPAP 40.7) and ESS (CPAP 7.9, no CPAP 5.8). The two arms were similar in terms of age (mean 75 years) and gender (males 61%). The mean BMI was significantly higher in the CPAP arm (27.8) than the no CPAP arm (26.7). The evaluations of association of CPAP use and cognitive outcomes were adjusted for BMI, AHI. The study was rated as high risk of bias for being an (unblinded) nonrandomized study that used a regression model for adjustment of baseline confounders.

In brief, five studies compared treatment with CPAP with no treatment and the APPLES trial compared real and sham CPAP. Pelletier-Fleury 2004 evaluated only adherent patients, but other studies were evaluated as ITT or "as-prescribed" analyses of CPAP prescription. Only the APPLES trial was explicitly powered for a cognitive function outcomes (the NRCS did not report a power analysis).

Effect of CPAP on Cognitive Function

The studies reported on a wide variety of tests and subtests (full results of all scales are presented in Appendix Table D-1). Here, we discuss the two measures evaluated by more than one study, the Mini-Mental State Examination (MMSE) and Trail Making Test B (TMT-B). None of the studies reported what they considered to be minimal clinically important differences for these tests, limiting conclusions about clinical significance that could be drawn from the studies.

Mini-Mental State Examination

The Wu 2016 (Yangzhou)[119] and PREDICT 2014[109] RCTs and the NRCS (Crawford-Achour 2015[124]) evaluated participants with MMSE at 6 months, 12 months, and 10 years, respectively. The MMSE is used to test cognitive function and screen for cognitive loss across the domains of orientation, attention, calculation, recall, language and motor skills. Each correct answer is worth a point, for a maximum of 30 points with higher scores being indicative of better cognitive function.[140] Based on a study of a mixed population of older adults with and without Alzheimer dementia, a 1 to 3 point decrease in MMSE score was indicative of a meaningful clinical decline in cognitive function.[141] However, it is not clear whether change in MMSE scores is a validated measure of cognitive decline in cognitively normal adults.[142] As expected, study participants had normal cognitive function, with average MMSE of 28 to 29 of 30 points.
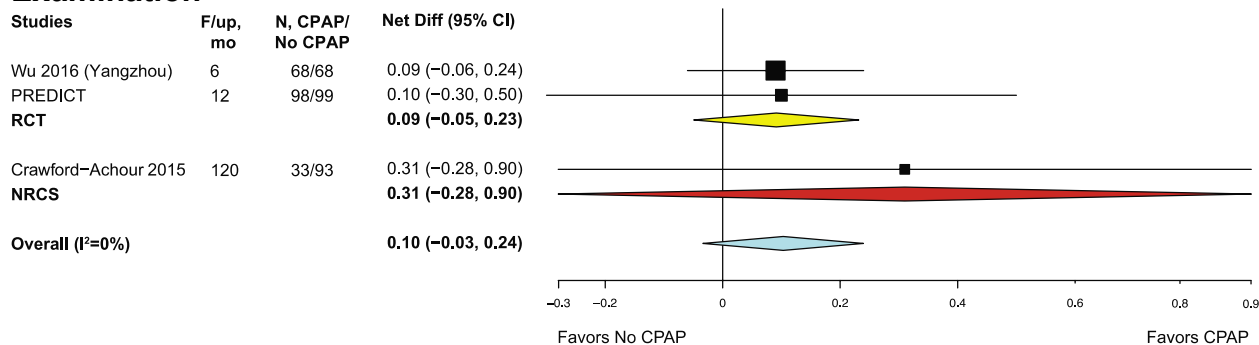
In all study arms (with or without CPAP), MMSE scores changed by 0.34 points or less over time (mostly "improved"). The RCTs found no statistically significant difference in changes in MMSE at 6 or 12 months, with net differences of less than 0.1 point. The NRCS reported no significant differences between baseline and 10 year followup scores across groups in their

adjusted analysis (P = 0.30). Based on the (unadjusted) scores reported and the adjusted P-value, we calculated a net difference of 0.31 (95% CI −0.28 to 0.90).

Neither of the RCTs reported on adherence rates. The NRCS reported 100 percent adherence (≥4 hr/night), but did not provide an explanation for the unusually high adherence rate except that adherence was self-reported.

Meta-analysis of the two RCTs (Figure 11) yielded a nonsignificant difference in changes in MMSE with or without CPAP at 6 or 12 months, favoring CPAP in direction (summary net difference 0.09 points, 95% CI −0.05 to 0.23). However, the difference was considerably less than the MCID of 1 to 3 points. Inclusion of the 10-year followup NRCS did not substantively change the summary net difference of 0.10 (95% CI, −0.03, 0.24). A sensitivity analysis showing the change in results as the analysis "borrows strength" from the NRCS evidence is shown in Appendix D, Figure D-6. The figure shows that that adding the imprecise NRCS evidence (to RCT-only evidence) has very little effect on the summary net difference or its precision.

Figure 11. Meta-analysis of CPAP versus No CPAP: Mini-Mental State Examination

Studies	F/up, mo	N, CPAP/ No CPAP	Net Diff (95% CI)
Wu 2016 (Yangzhou)	6	68/68	0.09 (−0.06, 0.24)
PREDICT	12	98/99	0.10 (−0.30, 0.50)
RCT			**0.09 (−0.05, 0.23)**
Crawford−Achour 2015	120	33/93	0.31 (−0.28, 0.90)
NRCS			**0.31 (−0.28, 0.90)**
Overall (I²=0%)			**0.10 (−0.03, 0.24)**

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, F/up = followup, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, NA = not applicable, Net Diff = net difference.

Trail Making Test A and B

There are two parts of the Trail Making Test (TMT), which consists of 25 circles distributed over a sheet of paper. In Part A, the circles are numbered 1 to 25, and the patients should draw lines to connect the numbers in ascending order. which involves connecting 25 dots as quickly as possible in ascending order. Similarly, in Part B, the patients draw lines to connect circles but, the circles alternate between numbers and letters (i.e., 1-A-2-B-3-C, …). It is a neurophysiological test of visual attention and task switching. Results are reported as the number of seconds required to complete the task, so lower scores indicate better cognitive function.[143] Based on a study of patients with CVD (advanced CHF), a 32-second increase in time to complete the task was provided as a meaningful cognitive decline.[144]

The Monasterio 2001 RCT and the NRCS (Crawford-Achour 2015) evaluated participants on the Trail Making Test A (TMT-A).

In the Monasterio 2001 RCT, participants in both the CPAP and no CPAP groups had improvements in the TMT-A scores after 6 months. The calculated net difference between groups were nonsignificant (net difference 3, 95% CI −4.6 to 10.6) and favoring no CPAP. In the NRCS, TMT-A scores increased (worsened) by 5 to 10 seconds in the two groups at 10 year followup (adjusted P = 0.09). Based on the (unadjusted) scores reported and the adjusted P-value, we calculated a net difference of 5.5 seconds (95% CI −0.8 to 11.1), nonsignificantly favoring no CPAP.
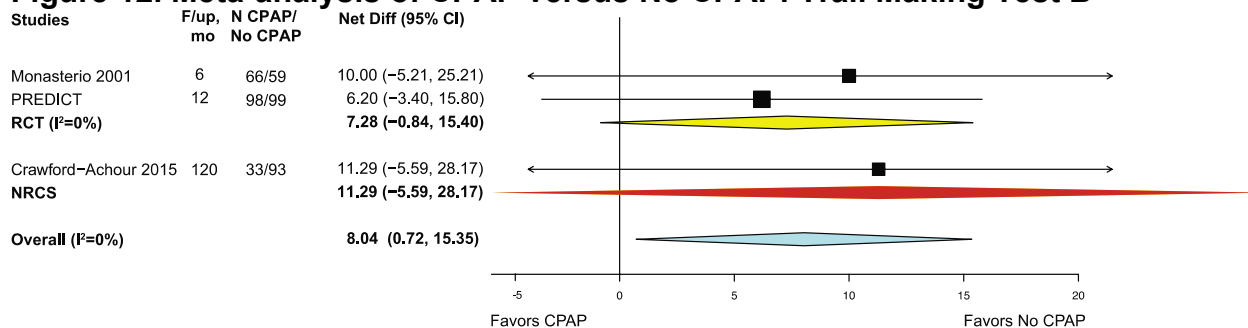
The Monasterio 2001 and PREDICT 2014 RCTs and the NRCS (Crawford-Achour 2015) evaluated participants on the Trail Making Test B (TMT-B).

In the RCTs, participants in both the CPAP and no CPAP groups had improvements in the TMT-B scores after 6 or 12 months. The difference between groups were nonsignificant and favored the no CPAP group (PREDICT: net difference 6.2 seconds, 95% CI −3.4 to 15.8, 12 months; Monasterio 2001: net difference 10, 95% CI −5 to 25, 6 months). In the NRCS, TMT-B scores increased (worsened) by 10 to 20 seconds in the two groups upon 10-year followup (adjusted P = 0.19). Based on the (unadjusted) scores reported and the adjusted P-value, we calculated a net difference of 11.3 seconds (95% CI −5.6 to 28.2), nonsignificantly favoring no CPAP.

Meta-analysis of the two RCTs (Figure 12) yielded a nonsignificant difference in changes in TMT-B with or without CPAP at 6 or 12 months, nonsignificantly favoring no CPAP (net difference 7.3 seconds, 95% CI −0.8 to 15.4). However, the difference was considerably less than the MCID of 32 seconds. Inclusion of the 10-year followup NRCS minimally altered the summary estimate to a statistically significant net difference of 8.0 (95% CI, 0.7, 15.4), favoring no CPAP. A sensitivity analysis showing the change in results as the analysis borrows strength from the NRCS evidence is shown in Appendix D, Figure D-7. The figure shows that that adding NRCS evidence (to RCT-only evidence) has very little effect on the summary net difference but increases precision to the point of significance.

Pelletier-Fleury 2004[108] investigated attention and concentration by adopting an incompletely validated parameter that evaluated the ratio of TMT-B to TMT-A. The ratio eliminated general speed variance (assessed by TMT-A time) to focus on the time required for shifting (from letters to numbers and from numbers to letters). The RCT reported the outcomes for participants in AHI ≥30 and AHI<30 subgroups. The participants in both the AHI subgroups had improvements in the scores after 6 months. However, in both subgroups the net difference between CPAP and no CPAP was 0 seconds (AHI ≥30 P = 0.61; AHI <30 P = 0.93).

Figure 12. Meta-analysis of CPAP versus No CPAP: Trail Making Test B

Studies	F/up, mo	N CPAP/ No CPAP	Net Diff (95% CI)
Monasterio 2001	6	66/59	10.00 (−5.21, 25.21)
PREDICT	12	98/99	6.20 (−3.40, 15.80)
RCT (I²=0%)			7.28 (−0.84, 15.40)
Crawford−Achour 2015	120	33/93	11.29 (−5.59, 28.17)
NRCS			11.29 (−5.59, 28.17)
Overall (I²=0%)			8.04 (0.72, 15.35)

Favors CPAP   Favors No CPAP

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, F/up = followup, I² = measure of statistical heterogeneity ranging from 0% (none) to 100%, NA = not applicable, Net Diff = net difference, NRCS = nonrandomized comparative studies, PREDICT = trial with undefined acronym, RCT = randomized controlled trials.

Other Cognitive Tests

The results of all test and subtests are in Appendix D Table 1. Across six studies, each of the other cognitive tests or subtests was evaluated in only a single study. Most analyses (35 of 42) of the net differences between CPAP and no CPAP groups were not statistically significant.

Heterogeneity of Treatment Effect and Applicability

Subgroup analyses were not conducted regarding effects of CPAP on MMSE or TMT-B. Study findings were statistically similar for these outcomes, thus no clear differences by population characteristics could be gleaned across studies.

One RCT reported subgroup analyses by severity of OSA (APPLES).[97, 98] OSA was defined as mild (AHI 10 to 15), moderate (AHI 15.1 to 30), and severe (AHI >30). In no subgroup did they find a significant net difference for the Pathfinder Number Test Total Time, Buschke Selective Reminding Test Sum Recall, or Sustained Working Memory Test Overall Mid-Day Index scales at 6 months.

The studies reporting on cognitive function, which mostly reported similar (nonsignificant differences), included a broad range of adult patients with OSA across age, AHI severity, history of CVD, and other characteristics. The PREDICT trial and the NRCS by Crawford-Achour 2015 were arguably most applicable to Medicare-eligible patients based on age criteria (≥65 years).

Summary of Effect of CPAP on Cognitive Function

Studies evaluated a wide range of cognitive tests, but with little consistency. It is not clear that the two tests—evaluated by at most three studies each (2 RCTs and 1 NRCS)—are ideal to evaluate changes in cognitive function in cognitively normal participants with OSA. Only the NRCS evaluated patients over a long enough period of time over which one might expect to see cognitive changes (10 years); the two RCTs were relatively short-term (6 and 12 months) in duration.

Three RCTs provide low SoE that CPAP does not have a clinically significant effect on executive cognitive function as measured by the MMSE or TMT-B tests (Table 15). Changes in cognitive function test scores were small and did not differ between groups. The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to any true effect.

There is insufficient evidence regarding the effect of CPAP on other measures of cognitive function. No study evaluated the risk of dementia.

Quality of Life and Functional Status Outcomes

Eight RCTs[12, 97, 101, 103, 105, 109, 113, 118] and one NRCS[120] reported on QoL based on a variety of measures. A ninth RCT reported on functional status outcomes, based on the Functional Outcomes of Sleep Questionnaire (FOSQ).[104] We included all QoL and functional status outcome measures that were reported in eligible studies regardless of whether they have been validated in patients with OSA or whether there is a claim that the QoL measure is in fact a health-related QoL measure.[66]

Five of the RCTs (ISAACC, MOSAIC, PREDICT, RICCADSA, SAVE) reported CVD outcomes and are described in the section, above, *Mortality and Cardiovascular Outcomes*. In brief, ISAACC included patients hospitalized for acute coronary syndrome with AHI ≥15 but without excessive daytime sleepiness. MOSAIC included 45- to 75-year-old adults with ODI >7.5 and ESS >9; the study was deemed high risk of bias due to lack of blinding and high dropout rate at 2 year followup. In PREDICT, participants were at least 65 years of age with newly diagnosed OSA (ODI ≥7.5, ESS ≥9); the study was deemed moderate risk of bias due to lack of participant or clinician blinding. RICCADSA randomized participants with OSA (AHI ≥15 and ESS <10) and CAD with coronary revascularization; the study was rated as high risk of

bias due to lack of blinding and high crossover. SAVE randomized participants with history of CVD or CeVD and OSA (ODI ≥12); the study was deemed moderate risk of bias due to lack of participant or clinician blinding.

The APPLES trial and Monasterio 2001 also reported cognitive function outcomes and are described in the section, above, *Cognitive (Executive) Function*.[104] Briefly, the APPLES RCT included participants with AHI >10 and the Monasterio 2001 RCT included participants with AHI between 10 and 30 without severe daytime sleepiness. APPLES was deemed low risk of bias (the study used a sham CPAP as a control) and Monasterio 2001 was deemed moderate risk of bias due to lack of participant or clinician blinding.

The remaining three studies did not report previously summarized outcomes. **Shaw 2016** randomized participants with OSA (ODI ≥15) and type 2 DM to CPAP (n = 151) or no CPAP (n = 147) and followed them for 6 months.[118] The trial was conducted and analyzed with an ITT approach. The mean baseline AHI was about 27 and mean baseline ODI was about 23. The study sample was 64 percent male and their mean age was 62.3 years. Patients were mostly obese, with a mean BMI of 33.0. The study was powered to detect a change in hemoglobin A1c, but not explicitly QoL outcomes. At 6 months, 61 percent of CPAP users were adherent (≥4 hours per night, 70% of nights) and mean CPAP usage was 4.9 hours per night. The study was rated as high risk of bias due to poor allocation concealment, lack of blinding (including outcome assessors), and higher withdrawal rate in the CPAP group (21%) than the no CPAP group (7%), largely due to intolerance of treatment.

The **BestAIR** (Best Apnea Interventions for Research) trial randomized patients with OSA (AHI ≥10) and CVD to CPAP (n = 83) or sham CPAP (n = 86) and followed for 6 or 12 months, depending on when they were recruited.[100, 101] The trial conducted an ITT analysis. The mean baseline AHI was 26.2 in the CPAP group and 32.0 in the sham CPAP group. The mean baseline ESS was about 8.3. The study sample was 65 percent male. The mean age of participants was 63.8 years. Patients were mostly obese, with a mean BMI of 31.7. The study was powered for changes in vascular markers and was described (*post hoc*) as underpowered for QoL outcomes. At 6 months, 43 percent of CPAP users were adherent (≥4 hours per night, 70% of nights). While the study was fully blinded, it was rated at moderate risk of bias due to differential followup times.

Bjornsdottir 2015 reported a NRCS that compared CPAP use to no use based on adherence over 2 years.[120] CPAP users were those who used CPAP for more than 20 of the prior 28 days for at least 4 hours/night on average (based on objective data) or more than 5 nights/week for at least 60 percent of the night by questionnaire (n = 348). CPAP nonusers were those who had returned their CPAP device within 1 year of therapy initiation and did not undergo upper airway surgery and were not using mandibular devices (n = 214). The study included adults with OSA (AHI ≥15) all of whom were prescribed CPAP. Mean AHI at baseline was not reported. Mean ESS for all participants was 11.7. The mean age of participants was 54.6 years and 81 percent were men. Most participants were obese, with a mean BMI of 33.4. The study was rated as high risk of bias as participants were categorized based on their decision to use or forego CPAP (thus participant selection bias based on participant characteristics observed after the start of intervention); although the study used a propensity score-model to adjust for baseline confounders.

In brief, seven studies compared treatment with CPAP with no treatment and BestAIR compared real and sham CPAP. All but one study (Bjornsdottir 2015) were evaluated as ITT or "as-prescribed" analyses of CPAP prescription. Bjornsdottir 2015 compared adherent users with nonusers. None of the studies was explicitly powered for QoL or functional status outcomes.

As will be described for each measure, the measures mostly have not been validated in adults with OSA. In addition, the QoL measures might not qualify as health-related QoL measures.[66]

Effect of CPAP on Quality of Life and Functional Status

Appendix Table D-3 includes the results from each study.

Short Form

The non-disease-specific Short Form (12 and 36 component versions) is a widely-used measure of QoL composed of set of generic questions that relate to vitality, physical functioning, bodily pain, physical role functioning, general health perceptions, emotional role functioning, social role functioning, and mental health). Scores are calculated for a Physical Component Summary and a Mental Component Summary. Scores range from 0 to 100 with *higher* scores indicating better QoL. The measure, and various versions and subscores (such as the Physical Component Summary and the Mental Component Summary) has been validated and found reliable in populations with numerous health conditions, but we did not find a validation study specifically in adults with OSA. An evaluation of the Medical Expenditure Panel Survey of 1040 older adults (mean age 74) who responded to a survey found the Physical and Mental Component Summaries to each be valid and reliable in this population.[145] These subscores are designed to have the same mean, SD, and thus MCID as the total SF-36 score. Thus, the MCID is about 4 to 7 points in a range of populations.[61, 146]

Physical Component Summary

Four RCTs evaluated QoL based on the SF-36 Physical Component Summary at either 6 months (MOSAIC), 1 year (BestAIR and RICCADSA) or a mean of 3.7 years (SAVE). The trials evaluated change in the measure in most randomized participants (BestAIR 100%, SAVE 89%, RICCADSA 84%, MOSAIC 83%). The four trials were inconsistent in that SAVE, MOSAIC, and BestAIR reported a significantly higher net improvement with CPAP,[12, 100, 101, 105] while RICCADSA reported a nonsignificantly worse improvement with CPAP.[113] Meta-analysis of these RCTs yielded a nonsignificant net difference of 1.2 point (95% CI −0.5, 3.0), nonsignificantly favoring CPAP, with a large degree of heterogeneity across studies (Figure 13).

Adherence with CPAP among the four RCTs was 38 percent (MOSAIC), 42 percent (SAVE), 52 percent (BestAIR), and 60 percent (RICCADSA). SAVE did not report an analysis among CPAP adherers. BestAIR found no correlation between CPAP adherence the QoL outcome (P = 0.89); changes in score were not significantly different between low and higher CPAP adherence. RICCADSA also reported no significant difference based on adherence.

The NRCS (Bjornsdottir 2015) reported results based on the shorter SF-12 comparing CPAP users with nonusers. The mean difference in the Physical Component Score was 3.42 (standard error [SE] 0.53) for the adherent CPAP group and 1.79 (SE 0.66) for the no CPAP group, with a nearly significant difference in an adjusted ANCOVA model (P = 0.06) at 2 years; although this P-value does not account for multiple comparisons testing. Based on the (unadjusted) scores reported and the adjusted P-value, we calculated a net difference of 1.6 (95% CI −0.1 to 3.3).

Figure 13. Meta-analysis of CPAP versus no CPAP: SF-36 Physical Component Summary

Studies	F/up, mo	N, CPAP/No CPAP	Net Diff (95% CI)
BestAIR 2017	12	83/86	3.4 (1.5, 5.3)
MOSAIC 2012	6	165/158	1.8 (0, 3.6)
RICCADSA 2016	12	102/104	−1.6 (−4.2, 1.0)
SAVE 2016	44	1218/1189	0.9 (0.3, 1.5)
Overall (I^2 = 79%)			1.2 (−0.5, 3.0)
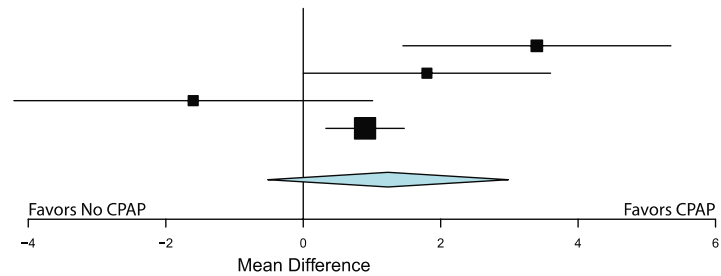
Favors No CPAP — Favors CPAP

Mean Difference: −4, −2, 0, 2, 4, 6

Abbreviations: BestAIR = Best Apnea Interventions for Research, CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, MOSAIC = Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular trial, Net Diff = net difference (difference-in-difference), RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial, SF = Short Form.

Mental Component Summary

Five RCTs evaluated QoL based on the SF-36 Mental Component Summary at 6 months (MOSAIC and Shaw 2016),[105, 118] 1 year (BestAIR and RICCADSA),[100, 101, 113] or a mean of 3.7 years (SAVE).[12] The trials evaluated change in the measure in most randomized participants (BestAIR and Shaw 2016 100%, SAVE 89%, RICCADSA 84%, MOSAIC 83%). The five trials were mostly consistent, with net differences ranging from 0.2 to 2.6. Three reported significantly higher net improvements with CPAP. By meta-analysis, the five RCTs yielded a statistically significant 1.4 point (95% CI 0.8 to 2.1) higher net improvement for CPAP when compared to no CPAP (Figure 14). The difference is likely not clinically significant (<4 points).

Adherence with CPAP among the five RCTs was 38 percent (MOSAIC), 42 percent (SAVE), 52 percent (BestAIR), and 60 percent (RICCADSA); in Shaw 2016, mean CPAP usage was 4.9 hours per night at 6 months. SAVE did not report an analysis among CPAP adherers. The MOSAIC trial reported a significantly larger effect among adherent (>4 hr/night) CPAP users than nonadherent users (4.75 vs. 1.64, P for interaction 0.02). BestAIR found no correlation between CPAP adherence the QoL outcome (P = 0.69); changes in score were not significantly different between low and higher CPAP adherence. RICCADSA also reported no significant difference based on adherence. Shaw 2016 reported a statistically significant difference (with no CPAP) in the Mental Component Score in the CPAP adherent subgroup, but did not report further details including whether the difference was significantly different than among nonadherent users.

The NRCS (Bjornsdottir 2015) also reported results based on the SF-12 comparing CPAP users with nonusers. The mean difference in the Mental Component Score was 2.13 (SE 0.54) for the adherent CPAP group and 2.35 (SE 0.68) for the no CPAP group, with no significant difference in an adjusted ANCOVA model (P = 0.80) at 2 years. Based on the (unadjusted) scores reported and the adjusted P-value, we calculated a net difference of −0.2 (95% CI −1.9 to 1.5).

Figure 14. Meta-analysis of CPAP versus no CPAP: SF-36 Mental Component Summary

Studies	F/up, mo	N, CPAP/No CPAP	Net Diff (95% CI)
BestAIR 2017	12	83/86	0.2 (−2.1, 2.5)
MOSAIC 2012	6	165/158	2.6 (0.9, 4.3)
RICCADSA 2016	12	102/104	2.6 (0.1, 5.1)
SAVE 2016	44	1218/1189	1.2 (0.6, 1.8)
Shaw 2016	6	151/147	1.5 (−0.7, 3.7)
Overall (I²=13%)			1.4 (0.8, 2.1)
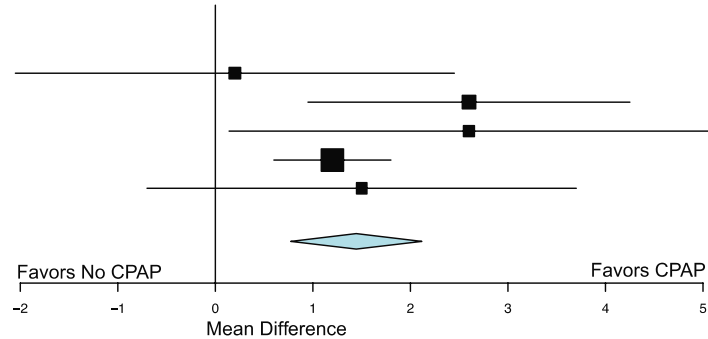
Favors No CPAP — Favors CPAP

Mean Difference (−2, −1, 0, 1, 2, 3, 4, 5)

Abbreviations: BestAIR = Best Apnea Interventions for Research, CPAP = continuous positive airway pressure, CI = confidence interval, F/up = followup, I^2 = measure of statistical heterogeneity ranging from 0% (none) to 100%, MOSAIC = Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular trial, Net Diff = net difference (difference-in-difference), RICCADSA = Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA trial, SAVE = Sleep Apnea cardioVascular Endpoints trial, SF = Short Form.

EuroQol-5D

EuroQol-5D is a widely-used non-disease-specific measure of QoL composed of generic health status questions that relate to mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Scores range from 0 to 1 with *higher* scores indicating better QoL (the measure also allows for negative values that can be considered to indicate being worse than dead). The measure has been validated and found reliable in populations with numerous health conditions, but we did not find a validation study specifically in adults with OSA. The MCID is about 0.18 points across a range of adults with various health conditions.[147]

ISAACC, MOSAIC, and SAVE all reported a small, not clinically significant, net difference of 0.02 on the EuroQol-5D scale, favoring CPAP use over no CPAP use (ISAAC reported a difference of 0.02 at 12, 24, and 36 months, and 0.03 at 48 months). The effect was statistically significant in the relatively large SAVE trial (with 91% of randomized participants analyzed), but not in the smaller MOSAIC trial (with only 55% analyzed). In ISAAC, the difference was just statistically significant at 12 and 24 months, but not 36 or 48 months.

Quality of Well-Being Scale

APPLES[95] reported significant improvements in Quality of Well-Being Scale scores between baseline and the 6-month followup among both sham and CPAP groups. However, the changes were the same in both groups; thus, no difference between groups.

Sleep Apnea Quality of Life Index

The Sleep Apnea Quality of Life Index (SAQLI) is an OSA-specific questionnaire composed of four core domains: daily functioning, social interactions, emotional functioning, and symptoms. The index has been validated against other measures of QoL symptom scores, and physiologic measures in adults with OSA.[148] The index ranges from 1 to 7 with *higher* scores indicating better QoL. The MCID is about 1 to 2 points.[62]

MOSAIC and PREDICT reported statistically significantly higher net improvements on the SAQLI scale, of 0.6 (95% CI 0.4 to 0.8) points[105] and 0.4 (95% CI 0.2, 0.6),[109] which were likely not clinically significant. Both trials analyzed most randomized participants (MOSAIC 84%, PREDICT 86%). Neither trial compared ITT and CPAP adherers analyses. In a comparison between CPAP and sham CPAP, the APPLES trial reported an increased odds of having a

change in SAQLI total score of at least 1 point in the CPAP arm than the sham CPAP arm (adjusted OR 1.66, 95% CI 1.1 to 2.6).[96] Among nonadherent participants (<4 hours/night), there were no changes or difference between groups over 6 months. Among adherent participants, SAQLI scores rose in both groups (CPAP 0.3 points vs. sham 0.1 points), which we calculated as a statistically significant difference (net difference 0.2, 95% CI 0.05 to 0.35). However, CPAP adherence (>4 hours vs. <4 hours) was not predictive of SAQLI change (P = 0.53).

Functional Outcomes of Sleep Questionnaire

FOSQ was designed to assess the impact of disorders of excessive sleepiness on multiple activities of everyday living and the extent to which these abilities are improved by effective treatment. It addresses questions related to activity level, vigilance, intimacy and sexual relationships, general productivity, social outcome, and the perceived difficulty of performing a given activity. It has been validated against other QoL measures in CPAP-treated patients with OSA;[68] although we found no studies that have validated it against established OSA-related health-related QoL outcomes. The scale ranges from 0 to 120 points, with *higher* scores indicating better functional status. The MCID is about 17 to 20 points.[149]

Monasterio 2001 evaluated functional status using a Spanish version of the original FOSQ-30.[104] This version of FOSQ had been validated in 39 patients with OSA (mean AHI 57.3, SD 21.7) and was found to be reliable and to correlate with self-determined assessments of health status ("regular" or "poor" health"[150]). Of note, it is unclear how many of the randomized participants were evaluated with FOSQ at 6 months. The RCT found a nonsignificant difference in change in FOSQ at 6 months, favoring CPAP in direction with a net difference 3.0 points (95% CI −3.6 to 9.6). The difference is not likely to be clinically significant. The study reported no significant difference in outcome between the CPAP adherers (64%) and nonadherent patients.

Heterogeneity of Treatment Effect and Applicability (QoL and Functional Status)

A single study evaluated the effect of CPAP on QoL in subgroups of participants; however, the study did not analyze whether the subgroup effects were significantly different from each other and the study did not account for multiple (subgroup) comparisons in their conclusions. The Bjornsdottir 2015 NRCS, of adherent users and nonusers, reported no significant effect of CPAP on SF-12 scores among participants with sleepiness symptoms (ESS ≥10), without sleepiness symptoms (ESS <10), or who used antidepressants.[120] They did find a small effect of CPAP on the Physical Component Score among morbidly obese participants (BMI ≥35, net difference 3.6, uncorrected P = 0.02), which was not seen in participants in other weight categories.

While the three trials that reported on SF-36 Physical Component Summary scores yielded heterogeneous results, it is not clear what differences across studies may have accounted for the different findings. Studies did not find significant differences in effect based on adherence with CPAP, which may suggest a true lack of effect or may indicate insufficient power to evaluation the association.

Studies were heterogeneous regarding whether the effect of CPAP on the SF-36 Mental Component Score was associated with CPAP adherence. Two RCTs suggested possible larger effects among CPAP adherent users than nonadherent users, but two found no significant association.

The conclusions may be considered generally applicable to adults with OSA as the studies included a broad range of adult patients with OSA across age, AHI severity, history of CVD, and other characteristics.

Summary of Effect of CPAP on Quality of Life and Functional Status

Four RCTs provide low SoE that CPAP does not have a statistical or clinically significant effect on the Physical Component Score of the SF-36 (Table 15). Although imprecise, all estimates of effect were less than the clinically significant threshold of about 4 to 7 points (summary net difference 1.2 (95% CI −0.5 to 3.0). Studies found no association between CPAP adherence and change in the Physical Component Score. The measure has not been validated in an OSA population.

Five RCTs provide low SoE that CPAP does not have a clinically significant effect on the Mental Component Score of the SF-36. Although the trials suggest a small improvement with CPAP, all estimates of effect were less than the clinically significant threshold of about 4 to 7 points (summary net difference 1.4, 95% CI 0.8 to 2.1). Studies were inconsistent whether adherence with CPAP was associated with the Mental Component Score. The measure has not been validated in an OSA population.

Four RCTs provide low SoE that CPAP does not have clinically significant effects on either the EuroQoL-5D (3 RCTs) or the SAQLI (2 studies) measures of QoL. The measures have not been validated in an OSA population. A fifth study found no difference in the Quality of Well-Being Scale with or without CPAP use.

Thus, overall, there is low SoE that there is no clinically significant effect of CPAP on long-term changes in QoL.

The low SoE for the various specific measures suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimates of effect are close to the true effect.

A single RCT provided insufficient evidence regarding the effect of CPAP on functional status as measured by the FOSQ.

Sexual Function

Three studies (1 RCT[93, 94] and 2 NRCSs[122, 151]) reported on sexual function outcomes. None reported on incident diagnoses (or resolution) such as erectile dysfunction or anorgasmia.

The Aarab 2017 RCT was described above, under *Mental Health*. In brief, the RCT included adults with OSA and sleepiness symptoms (ESS ≥10 or ≥2 sleepiness symptoms) with at least "mild" OSA (AHI ≥5). The participants were 70 percent male. The study was rated as moderate risk of bias due to lack of participant or clinician blinding; the sexual health outcomes were based on self-report. Sexual function outcomes were reported in a separate article.[94]

Budweiser 2013 was an observational study derived from a cohort of men who had had diagnostic polysomnography for suspected OSA.[122] Although reporting was unclear, the study apparently was a comparison of actual users versus nonusers (including nonadherent) with CPAP. Among these 401 men, only 91 responded to a sexual function survey at a median followup time of 3 years; OSA (AHI ≥5) was diagnosed in 83 of them (the subsample of analyzed men had statistically similar characteristics as the total cohort). The men were on average about 55 years old. CPAP users (n = 56) had significantly higher mean AHI (28.1) than nonusers (14.7, n = 35) and were significantly heavier (BMI 32.6 vs. 29.7, respectively). The

study was rated as high risk of bias because of inadequate description of assignment to CPAP, the likely combination of patients not prescribed CPAP with those who were nonadherent with CPAP, and the inadequate reporting of the adjusted analysis results.

Jara 2018 was an observational study that compared adherent CPAP use to nonadherent use.[151] The study included participants with an AHI ≥5 who used CPAP more than 4 hours per night (72 CPAP users) or not (110 CPAP nonusers) over the course of 1 year. Median AHI was higher among CPAP users (37.8) than among nonusers (20.5). The mean age of participants was 47.2, but CPAP users were older; 63 percent were men. Most participants were obese, with a mean BMI of 31.9, which was similar between groups. The study was rated as high risk of bias for evaluating CPAP based on adherence (rather than assignment/prescription).

In brief, three studies compared treatment with CPAP with no treatment. The single RCT was evaluated as an ITT analysis. The two NRCSs compared adherent users of CPAP with nonadherent/nonusers of CPAP. None of the studies was explicitly powered for sexual function.

Effect of CPAP on Sexual Function

Each of the three studies evaluated a different measure of sexual function based on different questionnaires.

The single RCT, Aarab 2017 (reported in Nikolopoulou 2017[94]) used the sexual/social dissatisfaction measure on the Sleep Disorders Questionnaire (SDQ; range 0 to 5 with *lower* scores indicating better sexual function).[152] The validity of this specific submeasure is unclear. The RCT found no net difference between CPAP and placebo in change in SDQ among men and women, combined, after 6 months (net difference 0; 95% CI −0.61 to 0.61). The study reported that their per protocol analysis (with adherence on 83% of nights) was similar to the ITT analysis.

The Budweiser 2013 NRCS reported multiple domains of sexual function based on the International Index of Erectile Function (IIEF-15) questionnaire (range 5 to 75 with *higher* scores indicating better sexual function).[153] The validity of the specific domains is unclear. Here we include only the measures for which they conducted multivariable analyses: erectile function and overall sexual function. At baseline (at the time of initial assessment), CPAP users and nonusers had statistically similar ratings on the two scores, although both were higher (better) for those who went on to use CPAP. The median IIEF-15 summary scores decreased (worsened) somewhat for all participants over a median of 36.5 months. In multivariable ANOVA analysis, adherent CPAP users had a somewhat improved sexual function summary score compared with nonusers, but the difference between CPAP users and nonusers was not statistically significant (P = 0.15, specific data not reported). In the multivariable analysis, erectile function was not associated with CPAP use (data not reported).

The Jara 2018 NRCS[151] created a sexual function measure based on two sex-specific questions from the Symptoms of Nocturnal Obstruction and Related Events (SNORE-25) instrument.[45] The validity of this *de novo* specific submeasure is unclear. The study authors converted the measure values into a standardized effect size such that a clinically important improvement as ≥|0.2| and a large effect size as ≥|0.80|; reductions in scores indicate improvements in sexual function. The study found that adherent CPAP users had improved sexual function after 12 months (change in score −0.7) units compared with a marginal improvement in the nonadherent group (−0.1) units. After multivariable adjustment, the net *reduction* was a statistically and clinically significant 0.49 units (95% CI 0.09 to 0.89), favoring CPAP use.[151]

Heterogeneity of Treatment Effect and Applicability (Sexual Function)

The two NRCSs provided some information about subgroup differences in association with CPAP use. In Budweiser 2013, in contrast with the overall group, in the subgroup of men with moderate to severe erectile dysfunction, CPAP users (n = 21) experienced greater improvement in the summary score than CPAP nonusers (n = 18, P = 0.014), but this analysis does not appear to have been adjusted.[122]

The Jara 2018 NRCS reported subgroup multivariable-adjusted results separately for men and women.[151] For men, effect size for CPAP was not statistically or clinically significant (effect size 0.16, 95% CI −0.26, 0.58), while it was large and statistically significant for women, favoring CPAP (effect size 1.34, 95% CI 0.50 to 2.18). However, whether the effect was significantly different in women than men was not analyzed.

Further comparison across studies to elucidate heterogeneity of treatment effect is limited due to different study designs and different outcome measures evaluated. No conclusions can be made about potential differences in effect based on factors such as severity of OSA (e.g., by AHI), definition of OSA or of apnea or hypopnea.

The three studies had relatively inclusive eligibility criteria, with, generally, a requirement that patients had an AHI ≥5, although on average an AHI of about 20. None of the studies was specifically applicable to Medicare-eligible patients, based on age.

Summary of Effect of CPAP on Sexual Function

There is insufficient evidence regarding the effect of CPAP on sexual function. Three studies each reported on unique measures.

Other Sequelae of Sleep Deprivation

The only other sequela of sleep deprivation that was reported in eligible long-term studies was days missed from work because of poor health. The SAVE trials reported the outcome, over the mean followup of 3.7 years. They reported similar numbers of participants missing work among those on CPAP (22.7%) and not on CPAP (23.6%). However, the number of missed days was higher in the no CPAP group, resulting in a significantly higher annual rate: 130 (CPAP) versus 159 (no CPAP) days per 100 participants per year; and a rate ratio of 0.82 (95% CI 0.80 to 0.85), favoring CPAP use. Of note, this endpoint was not prespecified and there was no power calculation for it. They study did not report a subgroup analysis for this outcome.

The SAVE trial is primarily applicable to adults with a history of CAD or CeVD with moderate to severe OSA (as defined by the study).

In summary, there is an insufficient SoE that CPAP decreases days of work missed due to poor health compared to CPAP nonuse, based on a single RCT.

Key Question 1: CPAP Versus Other Active Treatments

Key Points

- Data on long-term health outcomes comparing CPAP with other active treatments are relatively sparse. None was designed as a noninferiority or equivalence trial regarding health outcomes.

- There is low SoE that depression and anxiety symptoms remain similar between patients receiving either CPAP or MAD. The comparative effects on QoL, functional status, and sexual function are insufficient; other long-term health outcomes have not been reported.
- There is low SoE that functional status remains similar between patients prescribed fixed CPAP or autoCPAP. Other long-term health outcomes have not been reported.

Seven studies that reported long-term health outcomes compared CPAP to other active treatments (4 studies) or different CPAP devices (3 studies) that are approved or cleared for use in the U.S. None of the studies evaluated long-term (≥1 year) CVD outcomes, death, or accident/trauma rates. None was designed or analyzed as a noninferiority or equivalence trial regarding long-term health outcomes.

Table 20 summarizes the evidence from RCTs regarding the comparative effect of CPAP versus other active treatments on long-term health outcomes. No eligible NRCSs were found. In brief, comparative studies provide at most low SoE for all comparisons and outcomes. The low SoE for the conclusions suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding that the findings are stable or that the estimate of effect is close to the true effect. .

We conclude that the RCTs do not provide evidence of a difference in effect between auto-versus fixed CPAP on functional status, or between CPAP and MAD on depression and anxiety symptoms, QoL, or functional status. There is insufficient or no comparative evidence for other comparisons and outcomes.

Table 20. Evidence profile for CPAP versus mandibular advancement devices or between types of CPAP

Outcome	Study Design	No. Studies (Subjects)	Risk of Bias	Consistency	Precision	Directness	Other	Overall SoE	Conclusion Statements
Death and CV outcomes	.	0	.	.	.	.	.	.	None
Accidents	.	0	.	.	.	.	.	.	None
Hypertension	.	0	.	.	.	.	.	.	None
Diabetes	.	0	.	.	.	.	.	.	None
Depression (CPAP vs. MAD)	**RCT**	**2 (123)**	**Moderate**	**Consistent**	**Imprecise**	**Indirect**	**None**	**Low**	**No evidence of difference**
Anxiety (CPAP vs. MAD)	**RCT**	**2 (123)**	**Moderate**	**Consistent**	**Imprecise**	**Indirect**	**None**	**Low**	**No evidence of difference**
Cognitive function	.	0	.	.	.	.	.	.	None
QoL & functional status (CPAP vs. MAD)	**RCT**	**4 (239)**	**Moderate**	**Consistent**	**Imprecise**	**Indirect**	**None**	**Low**	**No evidence of differences**
Functional status (auto vs. fixed CPAP)	**RCT**	**3 (352)**	**Moderate**	**Consistent**	**Imprecise**	**Direct**	**None**	**Low**	**No evidence of differences**
Sexual function (CPAP vs. MAD)	**RCT**	**1 (38)**	**Moderate**	**N/A**	**Imprecise**	**Indirect**	**Single study**	**Insufficient**	**No conclusion**
Days of work missed	.	0	.	.	.	.	.	.	None

Abbreviations: . = no information, adj NRCS = adjusted nonrandomized comparative studies, CPAP = continuous positive airway pressure (device), CV = cardiovascular, MAD = mandibular advancement device, QoL = quality of life, RCT = randomized controlled trials, SoE = strength of evidence.

Evaluations of RCT evidence base are in bold font.

Note that where each included study is described for the first time (in detail), we have put the study or author/name in bold font to assist the reader to find the relevant description when the study is included in subsequent sections.

CPAP Versus MAD

Four RCTs compared CPAP with mandibular advancement devices (MAD),[93, 154-156] evaluating depression and anxiety, QoL, functional status, and sexual function (Table 21).

Mental Health (CPAP Versus MAD)

Two RCTs evaluated anxiety and depression (Aarab 2017 and de Vries 2019). Aarab 2017 is described under *CPAP Versus No CPAP /Mental Health*. Here we focus on their comparison of CPAP versus MAD. Briefly, it included adults with at least "mild" OSA (AHI ≥5) and sleepiness symptoms (ESS ≥10 or ≥2 sleepiness symptoms).[93] Participants (assigned to either CPAP, n = 18, or MAD, n = 20) were, on average, about 52 years old, 74 percent male, had a mean AHI about 21 and mean ESS about 11. Participants were mostly obese with a mean BMI of about 29. The study was powered for changes in AHI, not explicitly for any health outcome. The study was rated as moderate risk of bias due to lack of participant or clinician blinding; the mental health outcomes were based on self-reported completion of the SCL-90-R. SCL-90-R has not been validated in the OSA population.

After 6 months of CPAP or MAD, depression scores changed (improved) by −2.9 and −2.3 points, respectively, and anxiety scores changed (improved) by −1.7 and −2.0, respectively. The differences between groups were statistically and clinically nonsignificant: depression net difference −0.8 (95% CI −9.4, 7.8), anxiety net difference 0.3 (95% CI −4.7 to 5.3).

The **De Vries 2019** RCT compared CPAP (n = 42) to MAD (n = 43) in 85 participants with OSA in an ITT analysis.[154] Participants were included if they had an AHI between 15 and 30. The median baseline AHI was about 20 and the participants were 82 percent male. The mean age of participants was 50.7. The study was powered for changes in AHI, not explicitly for any health outcome. Among monitored patients, 68 percent of participants continued with their assigned treatment (not reported for each device). Those who continued treatment used their devices for a median of 6.8 hours per night with CPAP and 7.4 hours per night with MAD. The study was powered for changes in AHI, but not explicitly for any health outcome. The study was rated as moderate risk of bias due to lack of participant or clinician blinding; the mental health outcomes were based on self-reported completion of the HADS. HADS has not been validated in the OSA population.

After 12 months of CPAP or MAD, median depression scores changed improved) by −1 point in the CPAP group and −2 points in the MAD group. Median anxiety scores changed by −1 point in both groups. None of the changes or differences were clinically or statistically significant.

Neither study reported subgroup analyses.

In summary, there is low SoE that depression and anxiety symptom scores are improved by similar, not clinically significant, degrees after 6 or 12 months of either CPAP or MAD use (Table 20). The measures have not been validated in an OSA population. The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

Quality of Life and Functional Status Outcomes (CPAP Versus MAD)

Two RCTs evaluated various measures of QoL (Berg 2020 and de Vries 2019).[154, 155] In addition, two studies (de Vries 2019 and Uniken Venema 2020) evaluated functional status.[154, 156]

Berg 2020 randomized participants aged 20 to 75 years with moderate OSA (AHI 10 to 29.9) to CPAP (n = 37 assessed) or MAD (n = 41 assessed) for 1 year as an "as-treated" analysis.[155] For patients allocated to CPAP treatment, an autoCPAP device (ResMed, San Diego, CA, USA) was adapted and calibrated by a sleep technician. Participants in the two groups had similar characteristics in terms of age (52 years), gender (male 64%), BMI (31.5), and AHI (median 17.6). Adherence was poor; 51 percent of CPAP users were not adherent (≥4 hours/night, ≥70% of nights) and 12 percent of MAD users were nonadherent. The study was designed to be powered for a 10 percent difference in SF-36 scores, but underpowered, with 69 patients needed in each group. This trial was deemed to be at high risk of bias due to lacking of blinding and a high rate of loss to followup due to quitting treatment (25%).

De Vries 2019, described in the previous section compared the two treatments in 85 patients with moderate OSA (AHI 15 to 30).[154] The study reported SF-36, EuroQol-5D, and FOSQ in CPAP compared to MAD at 1 year. For SF-36, we calculated the estimated Physical and Mental Component Summary scores from the reported mean specific component scores.[157, 158]

Uniken Venema 2020 reported long-term followup of participants from a CPAP versus MAD RCT who were continuing to use the treatment to which they were randomized.[156] At 10 year followup, about 30 percent of original patients participated and were still using CPAP (n = 17) or MAD (n = 14). Patients were included in the original RCT[159] with OSA (AHI ≥5 and excessive daytime sleepiness or at least two other sleep-related symptoms). Analyzed patients were mostly male (94%) with a mean age of 60 years, mean BMI of 32.8, and a mean baseline AHI of 49.2 in the CPAP group and 31.7 in the MAD group. Mean self-reported adherence at 10 years was very high at 6.5 to 7 nights per week and 7 to 8 hours per night of use. This analysis did not define a power calculation. The study was deemed to be at high risk of bias given the selective criteria for who was included in the analyses and a lack of adjustment between groups.

The SF-36 Physical and Mental Component Summaries were reported in (or calculable from) Berg 2020 and de Vries 2019 (Table 21). In both trials, QoL scores improved, on average, with both CPAP and MAD treatment; however, changes in scores did not significantly differ between devices. Differences between devise were less than the MCID of 4 to 7 points.[61]

The de Vries 2019 trial found similar results for the EuroQol-5D analysis, with improvements in both groups at 1 year, but a small, not clinically significant, imprecise difference between groups: net difference −0.8 (95% CI −7.6 to 6.0; MCID ~18).

Two trials reported overall functional status, as measured with the FOSQ (de Vries 2019 and Uniken Venema 2020). Mean functional status improved with both treatments, by clinically significant amounts (MCID ~2), but differences between devices were small and nonsignificant.

None of the trials evaluated subgroup differences.

In summary, based on three small RCTs evaluated QoL (2 trials with SF-36, 2 trials of EuroQol-5D) and functional status (2 trials). Overall, the trials did not provide evidence of differences in effect of CPAP versus MAD on QoL or functional status (Table 20).

Sexual Function (CPAP Versus MAD)

Aarab 2017, described more fully above, included adults with at least "mild" OSA (AHI ≥5) and sleepiness symptoms, and was of moderate risk of bias.[93, 94] The study evaluated SDQ in 18

participants using CPAP and 20 using MAD for 6 months (Table 21).[94] The validity of this specific submeasure is unclear. The RCT found no statistically significant net difference between CPAP and MAD in sexual/social dissatisfaction on the SDQ scale (0.2; 95% CI −0.4 to 0.8).

The single small study provides insufficient evidence to make a conclusion regarding the relative effectiveness of CPAP or MAD on sexual function (Table 20).

Comparison of CPAP Modalities

Three RCTs compared different CPAP modalities and reported on either SF-36 (Meurice 2007) or FOSQ (Bloch 2018 and Kushida 2011) (Table 22).

Quality of Life and Functional Status Outcomes (CPAP Modalities)

SF-36

Meurice 2007 evaluated SF-36. The trial randomized participants to five different CPAP devices (4 different autoCPAP devices and 1 fixed CPAP) for 6 months.[160] Participants had severe OSA (AHI >30). Although this was an RCT, the study conducted an "as-treated" analysis based on actual use of CPAP, with 65 patients with sleep study data out of the 83 patients initially included in this study. The five groups had no differences between groups in mean AHI (52.3, n = 83) or mean ESS (11.3, n = 65). Overall, patients had a mean age of 56 years and a mean BMI of 30.8. The study was minimally powered for an undefined clinically significant change in AHI (requiring 12 participants per group). At 6 months, participants demonstrated good adherence with no difference between groups (mean 5.5 to 7.0 hours per night). The study was rated at high risk of bias due to lack of allocation concealment and blinding, and selective reporting.

Mean SF-36 Physical and Mental Component Scores changed by statistically nonsignificant amounts in all CPAP groups over a 6-month period, with a wide range of changes across groups. Comparisons between each of the four autoCPAP groups and the fixed CPAP group, likewise varied widely, but were also all highly nonsignificant.

The single high risk of bias trial provides insufficient evidence regarding difference in effect on QoL with either fixed or autoCPAP.

FOSQ

Two RCTs evaluated FOSQ in comparisons of autoCPAP and fixed CPAP (Table 22).

Bloch 2018 randomized participants with OSA (AHI ≥10 and ESS ≥8) to either autotitrated CPAP (n = 113) or fixed CPAP (n = 95) for 2 years as an ITT analysis.[161] The study was designed to test equivalence of autotitrated and fixed CPAP for ESS and sleep resistance, but not for health outcomes. The median baseline AHI was 48 and median ESS was 13. The study sample was 87 percent male. The median age was 55.5 years. Patients were mostly obese, with a median BMI of 32.7. The study was designed as and powered for an equivalence trial regarding ESS and sleep resistance time. The study was rated as moderate risk of bias due to lack of participant or clinician blinding (although outcome assessors were blinded). The trial was funded in part by ResMed and the Philips-Respironics Foundation.

Kushida 2011 randomized participants with OSA (AHI ≥15) to autotitrated CPAP (n = 54), or fixed CPAP (n = 57) for 6 months as an ITT analysis.[162] The mean baseline AHI was 38. The study sample was 87 percent male. The mean age of participants was 48.7 years. Patients were mostly obese, with a mean BMI of 34.5. A third randomized group (n = 53) received autotitrated

CPAP for 2 weeks to determine the level at which to fix CPAP pressure for the duration of the study. This comparison group was omitted as being outside the scope of this review. The study did not report an explicit power calculation. This trial was deemed to be low risk of bias for all criteria, including participant, clinician, and outcome assessor blinding. The study was funded by Philips Respironics.

In both studies, functional status scores improved (increased) among those receiving autoCPAP and fixed CPAP. However, the differences between groups at 2 years (Bloch 2018) and 6 months (Kushida 2011) were small (-0.2 and -0.4), and neither statistically significant nor clinically significant (MCID about 2 points), but estimates were imprecise. Based on these two studies, there is low SoE that functional status does not differ in patients prescribed autoCPAP or fixed CPAP (Table 20). The low SoE suggests that we have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.

Table 21. CPAP versus MAD RCTs: Mental health, quality of life, functional status, and sexual function

Study PMID	Outcome	Scale (Direction) [MCID]	Followup Duration	Arm	N Analyzed (Baseline / Followup)	Mean Diff (SD) [Diff of Median], Within-Arm	Net Diff [of Median] (95% CI), Between-Arm
.	**Depression**	.	.	.	.	.	.
Aarab 2017 28083705[93]	SCL−90−R: Depression	NR (lower better) [NR]	6 mo	CPAP MAD	18/18 20/20	−3.1 (15.9) −2.3 (10.3)	−0.8 (−9.4, 7.8)
de Vries 2019 31596213[154]	HADS: Depression component	0-21 (lower better) [1.9]	1 yr	CPAP MAD	39/37 40/29	[−1.0] [−2.0]	[1.0], NS
.	**Anxiety**	.	.	.	.	.	.
Aarab 2017 28083705[93]	SCL−90−R: Anxiety	NR (lower better) [NR]	6 mo	CPAP MAD	18/18 20/20	−1.7 (9.4) −2.0 (4.2)	0.3 (−4.7, 5.3)
de Vries 2019 31596213[154]	HADS: Anxiety component	0-21 (lower better) [1.7]	1 yr	CPAP MAD	39/37 40/29	[−1.0] [−1.0]	[0] NS
.	**QoL/Functional status**	.	.	.	.	.	.
Berg 2020 32665778[155]	SF-36, PCS	0-100 (higher better) [~4-7]	1 yr	CPAP MAD	55/55 49/49	1.7 0.5	1.8 (−0.5, 4.1) [A]
	SF-36, MCS	0-100 (higher better) [~4-7]	1 yr	CPAP MAD	55/55 49/49	2.2 4.4	−2.5 (−6.3, 1.3)
de Vries 2019, 31596213[154]	SF-36, PCS (estimated [B])	0-100 (higher better) [~4-7]	1 yr	CPAP MAD	39/37 40/29	7.2 (51.0) 4.9 (51.8)	2.4 (−20.7, 25.5)
	SF-36, MCS (estimated [B])	0-100 (higher better) [~4-7]	1 yr	CPAP MAD	39/37 40/29	1.9 (45.5) 1.6 (46.1)	0.4 (−20.2, 20.9)
	EuroQol-5D	0-100 (higher better) [~18]	1 yr	CPAP MAD	39/37 40/29	4.0 (12.9) 4.8 (14.7)	−0.8 (−7.6, 6.0)
	FOSQ	5-20 (higher better) [~2]	1 yr	CPAP MAD	39/37 40/29	[2.4] [2.5]	[−0.1], NS
Uniken Venema 2020, 31992403[156]	FOSQ	5-20 (higher better) [~2]	10 yr	CPAP MAD	17/17 14/14	5.1 (3.8, 6.4) 4.3 (2.9, 5.7)	0.8 (−1.1, 2.7)
.	**Sexual function**	.	.	.	.	.	.
Aarab 2017, 28294380[94]	SDQ, sexual/social dissatisfaction	0-5 (lower better), NR	6 mo	CPAP MAD	18/18 20/20	0 (0.9) −0.2 (1.1)	0.2 (−0.4, 0.8)

Abbreviations: . = no information, CPAP = continuous positive airway pressure, FOSQ = Functional Outcomes of Sleep Questionnaire, HADS = Hospital Anxiety and Depression Scale, MAD = mandibular advancement device, MCID = minimal clinically important difference, MCS = Mental Component Summary, Mean Diff = mean difference, Net Diff = net difference (difference-in-difference), NR = not reported, NS = nonsignificant (statistically), PCS = Physical Component Summary, PMID = PubMed identifier, QoL = quality of life, RCTs = randomized controlled trials, SCL−90−R = Symptom Checklist-90-Revised, SD = standard deviation, SDS = Self-rating Depression Scale, SF = Short Form.

[A] Adjusted difference. Analyses of adherent users had smaller, less statistically significant differences.
[B] Calculated from reported specific component scores.

Table 22. Comparison of CPAP devices RCTs: Quality of life and functional status

Study PMID	Outcome	Scale (Direction) [MCID]	Followup Duration	Arm	N Analyzed	Mean Diff (SD), Within-Arm	Net Diff (95% CI), Between-Arm
Meurice 2007 17638595[160]	SF-36, PCS	0-100 (higher better) [~4-7]	6 mo	AutoCPAP (4 devices) Fixed CPAP	51 total 14	−5.0 to 7.5 (all NS) 4.2 (−0.6, 9.0)	−9.2 to 3.3 (all NS)
	SF-36, MCS	0-100 (higher better) [~4-7]	6 mo	AutoCPAP (4 devices) Fixed CPAP	51 total 14	−0.8 to 7.8 (all NS) 1.9 (−2.7, 6.5)	−2.7 to 5.9 (all NS)
Bloch 2018 28982804[161]	FOSQ	5-20 (higher better) [~2]	24 mo	AutoCPAP Fixed CPAP	113 95	3.0 (2.4) 2.6 (2.2)	−0.4 (−0.9, 0.1) [A]
Kushida 2011 21804670[162]	FOSQ	5-20 (higher better) [~2]	6 mo	AutoCPAP Fixed CPAP	54/46 57/47	1.9 (2.8) 3.3 (2.2)	−0.2 (−1.6, 1.1)

Abbreviations: AutoCPAP = auto-adjust CPAP, CPAP = continuous positive airway pressure, FOSQ = Functional Outcomes of Sleep Questionnaire, HADS = Hospital Anxiety and Depression Scale, Mean Diff = mean difference, Net Diff = net difference (difference-in-difference), PMID = PubMed identifier, RCTs = randomized controlled trials.

[A] Adjusted difference.

Key Question 1: CPAP Adverse Events

Key Points

- Eligible comparative studies (RCTs and adjusted NRCSs) that reported long-term outcomes provided insufficient evidence regarding adverse events.
- Based on 854 case reports of adverse events available in the FDA database, leading types of adverse events related to oral and dental health, respiratory system, otolaryngology, odors, allergies and rashes, burns, eye health, aspiration, aerophagia, and miscellaneous other adverse events. Postulated reasons for adverse events mostly related to inadequate humidification, user errors, or device malfunction. No deaths were attributed to CPAP device use. We did not include information on FDA warnings or recalls.

Evidence

Only two RCTs reported adverse events related to CPAP use.[109, 118] None of the NRCSs reported adverse event data. Here we focus on adverse events reasonably attributable to CPAP use, as opposed to adverse outcomes that may occur among people using (or not using) CPAP that some studies classify as adverse events (e.g., stroke). Neither study reported the same adverse events (Table 23). We omitted a third trial of CPAP versus MAD that reported only adverse events attributable to MAD (temporomandibular joint disorder and delta overjet, a measure of tooth misalignment).[159, 163, 164] These data are, however, included in the Appendix tables.

The PREDICT trial prescribed an autotitrating CPAP device (S9 Autoset, ResMed), with humidifiers and choice of interface made on an individual basis (n = 114 on CPAP).[109] Shaw 2013 prescribed a similar autotitrating CPAP device (S8 Autoset Spirit II, ResMed), with pressure settings between 5 and 20 cm H_2O.[118]

Shaw 2016 directly compared rates of reported adverse events between groups, finding no significant differences in rates of pneumonia, headache, epistaxis, musculoskeletal pain, or gastrointestinal distress. In each instance, the adverse event occurred in no or one participant total.

PREDICT reported adverse events probably related to CPAP treatment only for those using CPAP. About one-quarter of participants reported interface related issues (e.g., claustrophobia, dislike of mask, leaking air, red/watery eyes, sore skin, pressure uncomfortable) and about one-third reported upper airway problems (e.g., dry mouth, runny or stuffy nose, sinus problems, nose bleeds). Four participants (3%) each reported abdominal bloating or, separately, anxiety or dyspnea related to CPAP use.

In summary, serious adverse events (that may require prescription treatment), including pneumonia and headache, are uncommon, but minor adverse events, particularly related to pressure or blown dry air, are common (at least for one device used in a particular manner). Important to note, though, is that due to our restrictive eligibility criteria, this systematic review is inadequate for the full assessment of CPAP-related adverse events. Thus, we conclude that the eligible studies provide insufficient evidence regarding adverse events.

Table 23. Adverse Events

Outcome Type	Outcome	Study	CPAP Device	Timepoint	CPAP n/N (%)	No CPAP n/N (%)	Risk Difference (95% CI), %	P Between Groups, Reported
Head and respiratory	Pneumonia	Shaw 2016[118]	AutoCPAP (S8 Autoset Spirit II; ResMed)	6 mo	1/151 (0.7)	0/147 (0)	0.7 (−0.6, 2.0)	1.00
	Headache	Shaw 2016[118]	AutoCPAP (S8 Autoset Spirit II; ResMed)	6 mo	0/151 (0)	1/147 (0.7)	−0.7 (−0.2, 0.7)	0.49
	Upper airway problems [A]	PREDICT[109]	AutoCPAP (S9 Autoset, ResMed)	12 mo	47/140 (33.6)	NR	NR	NR
	Epistaxis	Shaw 2016[118]	AutoCPAP (S8 Autoset Spirit II; ResMed)	6 mo	1/151 (0.7)	0/147 (0)	0.7 (−0.6, 2.0)	1.00
Psycho-logical	Anxiety/dyspnea related to CPAP	PREDICT[109]	AutoCPAP (S9 Autoset, ResMed)	12 mo	4/140 (2.9)	N/A	N/A	N/A
Musculo-skeletal	Pain, musculoskeletal	Shaw 2016[118]	AutoCPAP (S8 Autoset Spirit II; ResMed)	6 mo	1/151 (0.7)	1/147 (0.7)	0 (−1.9, 1.8)	1.00
Gastro-intestinal	Abdominal bloating	PREDICT[109]	AutoCPAP (S9 Autoset, ResMed)	12 mo	4/140 (2.9)	NR		
	Gastrointestinal distress	Shaw 2016[118]	AutoCPAP (S8 Autoset Spirit II; ResMed)	6 mo	1/151 (0.7)	0/147 (0)	0.7 (−0.6, 2.0)	1.00
General	Interface-related issues [B]	PREDICT[109]	AutoCPAP (S9 Autoset, ResMed)	12 mo	33/140 (23.6)	N/A	N/A	N/A

Abbreviations: AutoCPAP = autotitrating CPAP device, CI = confidence interval, CPAP = continuous positive airway pressure (device), N/A = not applicable (adverse event specific to CPAP use), NR = not reported.

[A] E.g., dry mouth, runny or stuffy nose, sinus problems, nose bleeds.
[B] E.g., claustrophobia, dislike of mask, leaking air, red/watery eyes, sore skin, pressure uncomfortable.

FDA MAUDE Database

To supplement evidence on adverse events reported in eligible studies, we searched the FDA Manufacturer and User Facility Device Experience (MAUDE) database for reports of adverse events related to CPAP device use. We queried using the terms "sleep apnea", "continuous positive airway pressure" and "CPAP" limiting results to January 1, 2009 through March 1, 2020. Of 854 records found, we identified 93 that reported adverse events during CPAP use. All were described as case reports.

We categorized reported events into 10 groups, listed in descending order of number of relevant records:

1. *Oral and dental health:* dry mouth, loose teeth, displaced denture, shifting of teeth, bridge loss, tooth decay, receding gums, tooth loss.
    a. These adverse events were noted to possibly be due to inadequate humidification in the oral cavity.
2. *Respiratory system*: sore throat, closed up throat, respiratory distress, difficulty in breathing, bronchitis (in the absence of humidification), pneumothorax, dyspnea, hypercapnia.
3. *Otolaryngology*: pain, ruptured eardrum, perforated nasal septum, burning in nostrils, damaged sense of smell, nasal abscess, throat infection.
4. *Bad odor* (including device malfunctions): soot emissions, headaches (due to odor), smoke in mask, noxious fumes.
5. *Allergies and rashes*: redness around the mask, allergies to neoprene, contact dermatitis, skin abrasions.
6. *Burns* (due to device or accessory malfunction): heated tubing.
7. *Eye health*: ulcers in the eye, vision loss, drying and inflammation of the eye.
8. *Aspiration:* Aspirated vomit.
9. *Aerophagia:* Compressed air in the stomach and bowel leading to bloody emesis and distended stomach.
10. *Miscellaneous*: Examples include fatigue, joint pains, paralysis on the right side of the face.

Within the case reports, the most frequent postulated reasons for the above events included inadequate humidification, user errors, or device malfunction. In the 854 records, there were 144 reported deaths but none was attributed to CPAP device use.

Key Questions 1b and 2: Intermediate and Surrogate Measures

Key Points

- No study has evaluated whether *change* in AHI, ESS, or other sleep and breathing measures are valid intermediate or surrogate measures for long-term outcomes. We did not assess the validity of single measurements of breathing or sleepiness measures (e.g., measured pretreatment) as predictors of outcomes or treatment effect.
- No study explicitly evaluated surrogacy or mediation analyses of the measures.
- None of the clinical event outcomes was reported by a sufficient number of studies that also reported change in breathing measure to allow adequate cross-study evaluation of concordance.
- Among 15 eligible studies that reported both changes in breathing or sleepiness measures (intermediate or surrogate outcomes) and effects on health outcomes in two or more comparable groups of study participants, none explicitly evaluated surrogacy or mediation analyses of intermediate measures. Across-study analyses did not find evidence of possible correlations; all such correlations were highly nonsignificant.
- Most comparisons between change in a given breathing measure and health outcome were informed by only one study.

Fifteen studies met eligibility criteria regarding the correlation between changes in potential intermediate or surrogate measures (AHI, ODI, ESS) and health outcomes in adults with OSA. Eleven studies evaluated changes in AHI, four studies evaluated changes in ODI, and 12 studies evaluated changes in sleepiness scores (i.e., ESS). No study evaluated changes in other potential intermediate or surrogate measures including apnea index, respiratory disturbance index, or respiratory event-related arousals.

Eligible studies had to compare two or more groups (interventions) and report changes in breathing measures and/or ESS over time, as well as health outcomes.

We found no study that formally validated changes in breathing measures as intermediate or surrogate outcomes for health outcomes. No study reported within-study participant-level correlation analyses between intermediate or surrogate outcomes and health outcomes.

Consistent with the findings described above in the *Definitions of Breathing Measures Used Across Studies* section, there was a high degree of inconsistency in the criteria used to define the sleep study measures and multiple instances of incomplete reporting of definitions and criteria.

Among the 11 studies that evaluated changes in AHI, sleep studies were conducted in the laboratory (only) in three studies, in either the lab or at home in four studies, at home (only) in two studies, and not reported in two studies. Eight of the studies cited AASM criteria (1 1997, 2 1999, 2 2007, 2 2012, 1 no date), one study cited the American Thoracic Society Statement from 1994, and one cited the Spanish Sleep Network from 2011. Apnea thresholds were defined as 90 percent in two studies, 100 percent in two studies, and not reported in the other studies. Hypopnea thresholds were defined as 50 percent in three studies, "30 to 90" percent in one study, and not reported in the other studies. Additional oxygen desaturation criteria for hypopnea was set at 3 percent in three studies, 4 percent in two studies, and not reported in the others.

Among the four studies that evaluated changes in ODI, sleep studies were conducted in the laboratory (only) in one study, in either the lab or at home in two studies, at home (only) in one

study. Three of the studies cited AASM criteria (1999, 2007, and 2012), and one cited the Spanish Sleep Network from 2011. Oxygen desaturation criteria was set at 3 percent in one study, 4 percent in another study, and not reported in the other two studies.

Correlation of Changes in Intermediate or Surrogate Measures With Health Outcomes

From the 15 eligible studies, we extracted 50 sets of breathing measure and health outcome pairs. Evaluated breathing measures (intermediate or surrogate outcomes) included AHI (41 comparisons in 11 studies) and ODI (9 comparisons in 4 studies). Studies commonly reported multiple health outcomes and/or multiple breathing measures. However, most comparisons between a given breathing measure and health outcome were informed by only a single study (Appendix Table D-2). We describe the rare instances of reports of both intermediate or surrogate measures and clinical event outcomes, but otherwise describe only those breathing measure–(continuous) outcome pairs for which there were at least three comparisons that allowed metaregression analysis across studies.

Notably, none of the clinical event outcomes (e.g., mortality, stroke) was reported by a sufficient number of studies that also reported change in breathing measure to allow cross-study evaluation of concordance.

Change in Intermediate or Surrogate Measures and Health Outcomes

Only three RCTs reported changes in breathing or sleepiness measures in both randomized groups and reported clinical event outcomes (Table 24). As noted, none reported a patient-level correlation between measures and health outcomes. None of the trials used the same method to estimate AHI, ODI, or other sleep study measures in both randomized groups (the CPAP and the no CPAP groups).

In the SAVE trial,[12] at enrollment, patients, ODI was estimated from a two-channel (oximetry and nasal pressure) home device, but apneas and hypopneas (and thus AHI) were measured only with the nasal pressure channel. The definition of AHI was not reported. Enrolled participants had mean AHI of 29.0 (SD 15.9) in the CPAP group (n = 1346) and 29.6 (SD 16.4) in the no CPAP group (n = 1341). During the trial, AHI was measured in a different manner, by the CPAP device itself, at home while in use. It is unclear whether the definitions of AHI used at enrollment and during the study differed. Over the duration of the trial (mean 44 months), the mean AHI among 1299 participants using the CPAP device was 3.7 (SD 4.3) with a median AHI of 2.5 (IQR 1.8 to 4.1). Followup AHI was not measured in the no CPAP group. Thus, strictly speaking, no patient-level or study-level assessment can be made to assess the validity of change in AHI as a surrogate or intermediate measure of risk of health outcomes. However, under the assumption that AHI did not change substantially among those not using CPAP, it can be noted that the large decrease in AHI with CPAP use (despite problems with the assessment of AHI) did not correspond to the relatively small, nonsignificant effects on their primary outcome (composite CV events), all-cause mortality, or other clinical events. The SAVE trial also reported change in ESS, which similarly favored CPAP (net change −2.5, 95% CI −2.8 to −2.2) but also did not correspond with the lack of significant effect on clinical events. However, no patient-level correlation analysis was conducted between change in ESS and clinical event risk.

Corral 2017 randomized OSA patients to home versus laboratory testing followed by CPAP in all patients (the study was designed as an equivalence study and was not a comparison of CPAP, per se).[165] The study measured AHI among those who had laboratory sleep tests; both

apnea and hypopnea definitions used a 10 second threshold and hypopnea was defined as ODI ≥3 percent or arousal. For those receiving home sleep studies, the definition of AHI did not include arousals. In addition, the home recorded sleep time is higher than the laboratory value. Therefore, home AHI measures may have systematically been less than laboratory AHI measures. This was borne out by pre-CPAP measures: median home AHI was 20.9, while laboratory AHI was 28.5; statistical significance is not reported. Potentially, thus, participants in the home sleep study group may have been less likely to have been prescribed CPAP (since their measures of AHI were lower). Although the definitions of AHI differed in the two groups, AHI was substantial reduced with CPAP to a median of 3.7 for the home sleep study group and 4.6 for the laboratory group. The difference was nonsignificant (P = 0.54). The annualized composite CV event incidence rate was also similar between groups (6.4 [SD 30.7] vs. 7.3 [32.7] events per 100 patients per year; P = 0.77) as were work and traffic accidents at 6 months (8.7% vs. 7.1%; P = 0.68). No analyses or conclusions could be drawn about possible correlation between change in AHI (not to mention any given definition of AHI) and long-term health outcomes. As shown in Table 24, there was also no evidence suggesting a correlation between ESS and health outcomes.

The PREDICT trial reported change in ESS, which was significantly improved in the CPAP group (–4.2 [SD 4.1] compared with the no CPAP group (-2.1 [SD 3.6]; Table 24).[109] While the risk of composite CV events favored CPAP prescription in direction, the effect size was nonsignificant (HR 0.87, 95% CI 0.40 to 1.88). The study did not report an analysis of correlation between change in ESS and risk of composite CV event.

Table 24. Trials reporting both changes in intermediate measures and health outcome effect sizes

Study, PMID	Followup Interval	Arm	Measure	Estimated Net Difference (95% CI)	Event	Relative Difference (95% CI)
Corral 2017, 28636405[165]	6 mo	Home vs. lab sleep study	ESS	0.7 (−0.2, 1.6) Favor (NS) lab sleep study	Accident	1.1% (−3.7, 5.7) Favor (NS) lab sleep study
			AHI [A]	1.4 (−1.2, 4.0) Favor (NS) lab sleep study	CV incidence	−0.9% (−6.9, 5.1) Favor home sleep study
PREDICT, 25172769[109]	1 yr	CPAP vs. no CPAP	ESS	−2.1 (−3.0, −1.2) Favor CPAP	Composite CV events [B]	HR 0.87 (0.40, 1.88) Favor (NS) CPAP
SAVE, 27571048[12]	3.7 yr (mean)	CPAP vs. no CPAP	ESS	−2.5 (−2.8, −2.2) Favor CPAP	Composite CV events [C]	HR 1.10 (0.91, 1.32) Favor (NS) no CPAP
			AHI [D]	−25.3 (−24.1, −22.2), CPAP group	Composite CV events [C]	HR 1.10 (0.91, 1.32) Favor (NS) no CPAP

Abbreviations: CI = confidence interval, CPAP = continuous positive airway pressure, NS = not statistically significant, PMID = PubMed Identifier, PREDICT = trial with undefined acronym, SAVE = Sleep Apnea cardioVascular Endpoints trial.

[A] Measured differently in two groups. See text.
[B] Specific event outcomes were rare and effect estimates were imprecise.
[C] Primary outcome. Other event outcomes also statistically nonsignificant.
[D] Reported only among those using CPAP, while using CPAP. Measured differently at enrollment and during CPAP use. See text.

Change in Breathing Measures and Quality of Life
AHI and SF-36 Physical Component Summary

Four studies reported data on changes in AHI and the SF-36 Physical Component Summary score (Figure 15A).[154, 160, 165, 166] Note that, as described in the *Quality of Life and Functional*

Outcomes section of Key Question 1, the outcome measure has not been validated in patients with sleep apnea.

Across studies there was a moderate "positive" correlation (correlation coefficient = 0.40), consistent with larger improvements (reductions) in AHI correlating with larger improvements (increases) in SF-36 score. However, the correlation was not statistically significant (P = 0.75).

There were three studies that found a "positive" correlation (in the northwest or southeast quadrants of the figure) and one study that found a "negative" correlation (in the northeast or southwest quadrants); the difference between "positive" and "negative" correlations was not significant (P = 0.63, by exact binomial test).

AHI and SF-36 Mental Component Summary

The same four studies reported data on changes in AHI and the SF-36 Mental Component Summary score (Figure 15B).[154, 160, 165, 166] Note that the outcome measure has not been validated in patients with sleep apnea. Across studies there was a moderate "negative" correlation (correlation coefficient = 0.40), consistent with larger improvements (reductions) in AHI correlating with larger worsening (decreases) in SF-36 score. However, the correlation was not statistically significant (P = 0.75). The number of studies with "positive" and "negative" correlations (2 vs. 2) were the same (P = 1.00).
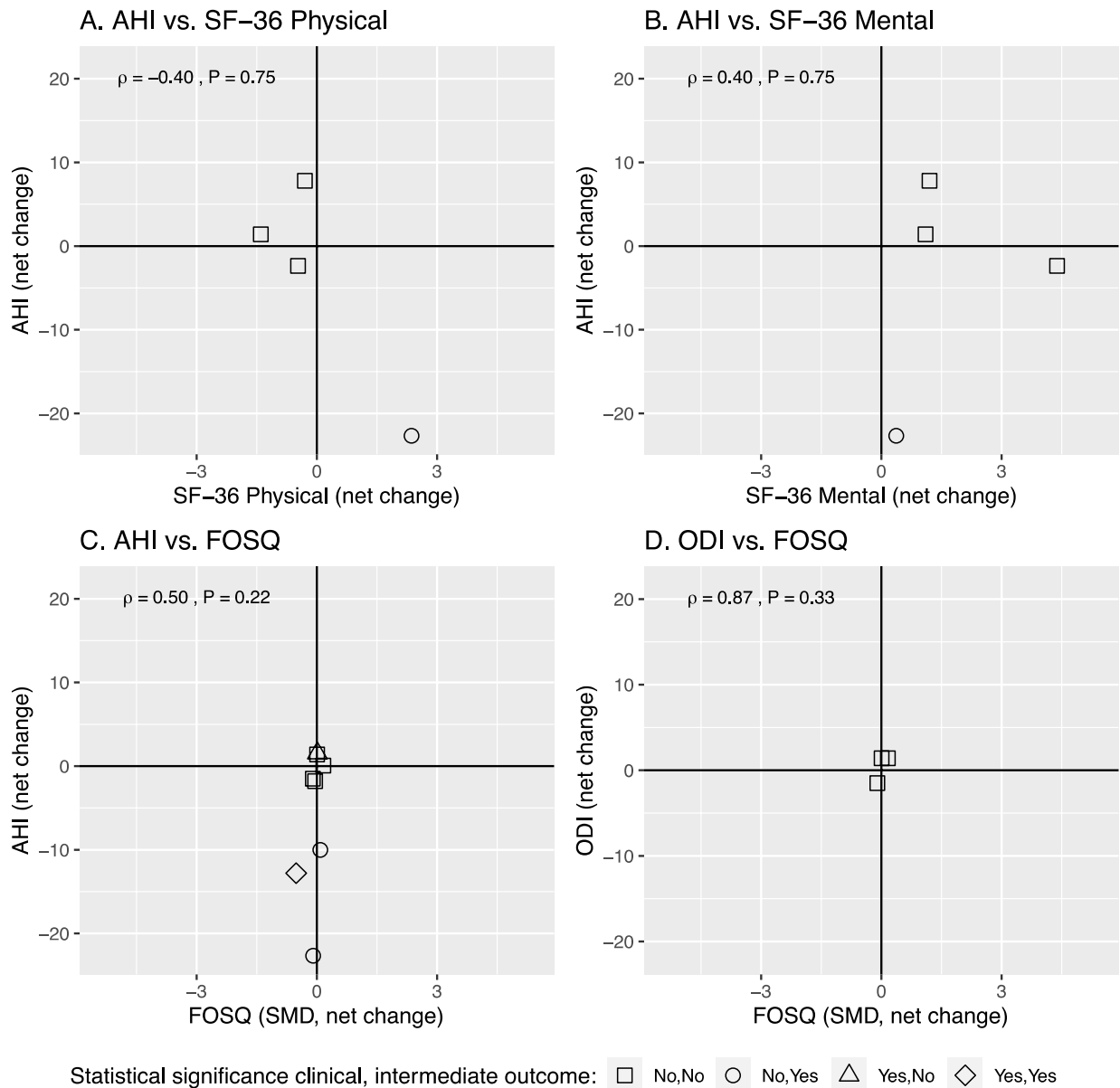
Change in Breathing Measures and Functional Status
AHI and FOSQ

Eight studies reported data to allow a comparison between changes in AHI and the FOSQ measure of functional status (Figure 15C).[104, 154, 161, 162, 165, 167, 168] Note that FOSQ has been validated against other QoL measures in CPAP-treated patients with OSA.[68] Studies tended to be centered around the null for both AHI and FOSQ (finding small and/or nonsignificant net changes in both AHI or FOSQ scores). Across studies, there was a moderate "negative" correlation (correlation coefficient = 0.50), consistent with larger improvements (reductions) in AHI correlating with larger worsening (decreases) in FOSQ score. However, the correlation was not statistically significant (P = 0.22). One study found a "positive" correlation and seven studies found a "negative" correlation; the difference between "positive" and "negative" correlations was near statistically significant (P = 0.070).

ODI and FOSQ

Three studies reported data on ODI and FOSQ (Figure 15D).[161, 165, 167] Across studies there was a very strong "negative" correlation (correlation coefficient = 0.87), consistent with larger improvements (reductions) in ODI correlating with larger worsening (decreases) in FOSQ. However, the correlation was not statistically significant (P = 0.33). The number of studies with "positive" and "negative" correlations (0 vs. 3) did not significantly differ (P = 0.25).

Figure 15. Correlation between changes in breathing measures and health outcomes

A. AHI vs. SF–36 Physical

ρ = −0.40 , P = 0.75

B. AHI vs. SF–36 Mental

ρ = 0.40 , P = 0.75

C. AHI vs. FOSQ

ρ = 0.50 , P = 0.22

D. ODI vs. FOSQ

ρ = 0.87 , P = 0.33

Statistical significance clinical, intermediate outcome: □ No,No ○ No,Yes △ Yes,No ◇ Yes,Yes

Correlation figures between net changes (between groups) of breathing measures versus (continuous) health outcomes. Northwest and southeast quadrants represent improvements (reductions) in breathing measures correlating with improvements in health outcome measures.

Abbreviations: ρ = correlation value, AHI = apnea-hypopnea index, FOSQ = Functional Outcomes of Sleep Questionnaire, ODI = oxygen desaturation index, P = P-value (statistical significance), SF = Short Form, SMD = standardized mean difference.

Correlation of Changes in Sleepiness Measures With Health Outcomes

From the 12 eligible studies, we extracted 85 sets of ESS measurements and health outcome pairs. Studies commonly reported multiple health outcomes. However, most comparisons between a given breathing measure and health outcome were informed by only one study

(Appendix Table D-2). Here we describe only those ESS–outcome pairs for which there were at least three comparisons that allowed regression analysis across studies. Notably, none of the clinical event outcomes (e.g., stroke) was reported by a sufficient number of studies that also reported change in ESS to allow cross-study evaluation of concordance.

Change in ESS and Mental Health
ESS and HADS Depression Score

Three studies reported data on changes in ESS and the HADS depression score (Figure 16A).[12, 109, 154] The deficiencies of the HADS depression scores, in terms of its validity and interpretation, have been addressed in the Mental Health section, above, for CPAP versus no CPAP.

Across studies there was a moderate "positive" correlation (correlation coefficient = −0.50), consistent with larger improvements (reductions) in ESS correlating with larger improvements (decreases) in the HADS score. However, the correlation was not statistically significant (P = 1.00). The number of studies with positive and negative correlations (3 vs. 0) did not significantly differ (P = 0.25).

ESS and HADS Anxiety Score

The same three studies reported data on changes in ESS and the HADS anxiety score (Figure 16B).[12, 109, 154] The same caveats apply as for the depression scores.

Across studies there was a moderate "positive" correlation (correlation coefficient = −0.50), consistent with larger improvements (reductions) in ESS correlating with larger improvements (decreases) in the HADS score. However, the correlation was not statistically significant (P = 1.00). The number of studies with "positive" and "negative" correlations (3 vs. 0) did not significantly differ (P = 0.25).

Change in ESS and Quality of Life
ESS and SF-36 Physical Component Summary

Six studies reported data on changes in ESS and the SF-36 Physical Component Summary score (Figure 16C).[12, 101, 154, 160, 165, 166] Across studies there was a strong "positive" correlation (correlation coefficient = −0.60), consistent with larger improvements (reductions) in ESS correlating with larger improvements (increases) in the SF-36 score. However, the correlation was not statistically significant (P = 0.24). The number of studies with "positive" and "negative" correlations (5 vs. 1) did not significantly differ (P = 0.22).

ESS and SF-36 Mental Component Summary

The same six studies reported data on changes in ESS and the SF-36 Mental Component Summary score (Figure 16D).[12, 101, 154, 160, 165, 166] Across studies there was a moderate "negative" correlation (correlation coefficient = 0.77), consistent with larger improvements (reductions) in ESS correlating with larger worsening (decreases) in the SF-36 score. However, the correlation was not statistically significant (P = 0.10). The number of studies with "positive" and "negative" correlations (3 vs. 3) were the same (P = 1.00).

Change in ESS and Functional Status
ESS and FOSQ

Five studies reported data on changes in ESS and the FOSQ score (Figure 16E).[104, 154, 161, 162, 165] Across studies there was a moderate "positive" correlation (correlation coefficient = −0.40), consistent with larger improvements (reductions) in ESS correlating with larger improvements (increases) in the FOSQ score. However, the correlation was not statistically significant (P = 0.52). The number of studies with "positive" and "negative" correlations (4 vs. 1) did not significantly differ (P = 0.38).

Figure 16. Correlation between changes in sleepiness scores and health outcomes
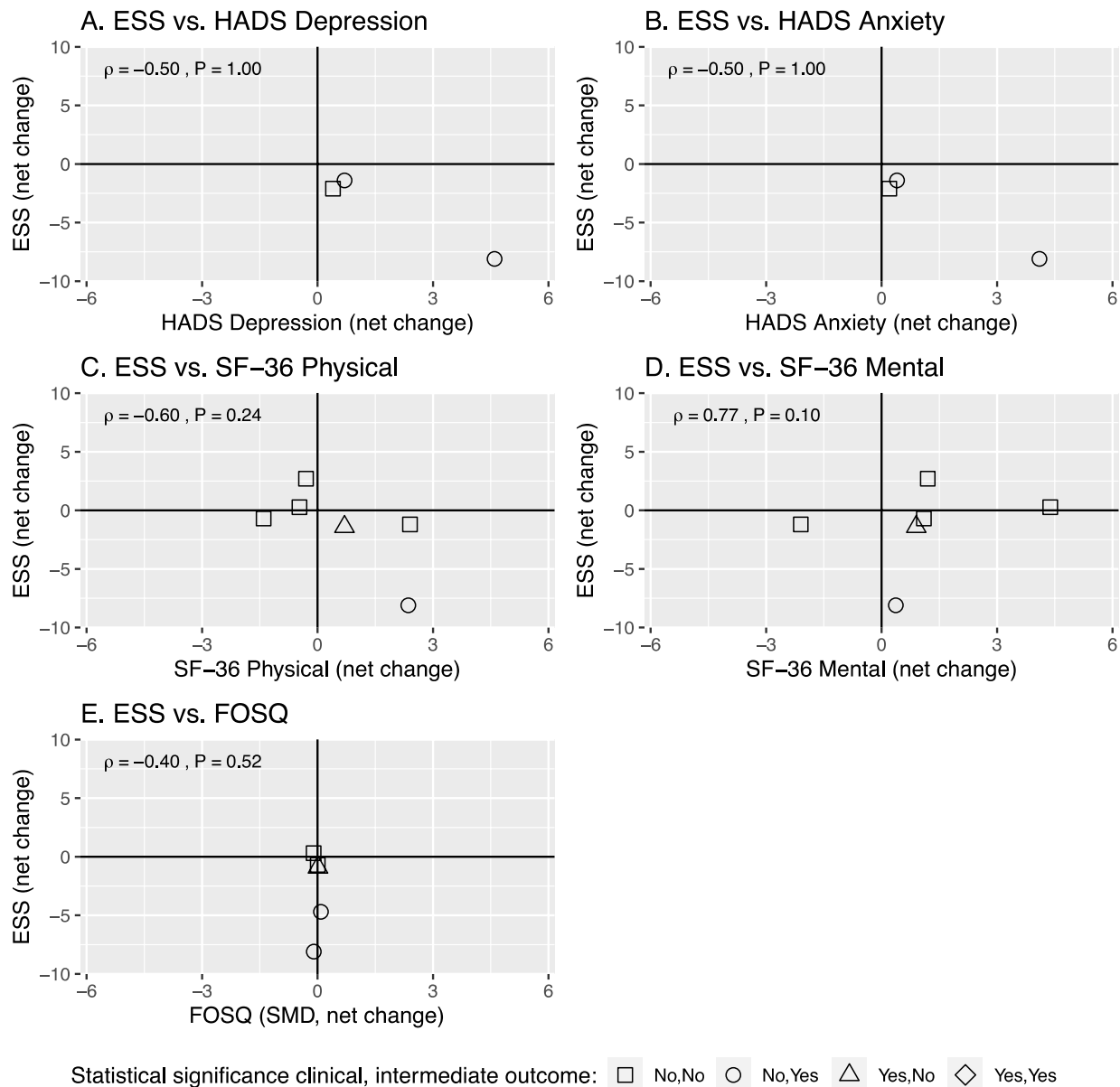
Correlation figures between net changes (between groups) of sleepiness scores (ESS) versus (continuous) health outcomes. Northwest and southeast quadrants represent improvements (reductions) in ESS correlating with improvements in health outcome measures.

Summary of Evidence Regarding Correlation of Changes in Intermediate or Surrogate Measures With Health Outcomes

Overall, the evidence base neither supports nor refutes whether changes in commonly used measures (AHI, ODI, ESS) are valid intermediate or surrogate measures for long-term health outcomes. We did not assess the validity of single measurements of breathing or sleepiness measures (e.g., measured pretreatment) as predictors of outcomes or treatment effect.

No studies that compared different groups of patients being treated for OSA evaluated within-study participant-level correlation analyses between changes in any potential intermediate or surrogate measure and any long-term health outcome of interest. Thus, no conclusions can be made regarding whether any measures may be actual intermediate or surrogate outcomes for health outcomes.

Only three trials (two of which, SAVE and PREDICT, compared CPAP and no CPAP) reported both changes in potential intermediate or surrogate measures (AHI and ESS) and effect sizes of clinical event outcomes (composite CV events and others). The two CPAP versus no CPAP trials found clinically and statistically significant improvements in ESS (reduction >2)[169] but not concomitant reductions in risk of composite CV events. Consistent with numerous RCTs that evaluated AHI,[11] incidentally, the SAVE trial found that CPAP reduced AHI by a large degree (−25.3), but yielded nonsignificant effects on composite CV events, mortality, and other clinical events. However, none of these findings provide sufficient information as to whether or not AHI or ESS are valid surrogate or intermediate measures for health outcomes.

Meta-regression correlation analyses failed to indicate possible correlations between breathing measures (AHI, etc.) on the one side and psychometric outcome measures (SF-36, etc.) on the other. All such correlations were highly nonsignificant (and even more so had we accounted for multiple testing). Too few studies reported changes in sleepiness measures (ESS) and clinical event outcomes (composite CV events, accidents) to allow correlation analysis. Any correlations we might have found would have been hypothesis-generating only, due to the possibility of ecological fallacy (erroneous conclusions based on applying group-level correlations to individuals within the groups).

Ideal Study Design to Establish Validity of Mediator (Intermediate) and Surrogate Measures

To assist with the discussion below, we first provide definitions of terms. A surrogate outcome is a factor or variable that is correlated with another outcome. By assessing how a surrogate outcome is affected by treatment, we can infer how the treatment affects a different outcome. Surrogate outcomes are commonly measured as substitutes for clinical endpoints (that are typically more difficult or time consuming to measure). Notably, there is no assumption that the surrogate outcome is in the causal pathway for the primary outcome; a change in the surrogate measure does not cause a change in the primary outcome. For example, reduction of fever may be a valid surrogate measure for the effect of an antibiotic to treat pneumonia but is not plausibly in the pathway for how antibiotics cure pneumonia.

A mediator (also called intermediate factor) is a factor or variable in a causal pathway between treatment and outcome through which the effect of the treatment on outcome is, partly

127

or fully, exhibited.[170, 171] If the effect of a treatment on a measure (such as blood pressure) directly impacts the outcome (e.g., stroke), then the measure can be considered to be a mediator. Notably, the effect on the mediator need not be the only way in which an intervention impacts the primary outcome. For example, while cholesterol levels may (or may not) be a mediator for statin's effect on risk of myocardial infarction, the statins effects on anti-inflammatory milieu may also play a role. Understanding mediators that exist in a causal pathway can explain how a given treatment affects the outcome.

We note that no eligible studies conducted mediation or surrogacy analyses.

Mediation Analyses

The effects of CPAP on health outcomes may be mediated through a variety of mechanisms involving multiple pathways to explain any observed treatment effects. As an example, the (putative) major pathway of clinical relevance is illustrated in Figure 17. The overall (total) effect of CPAP on a particular outcome (e.g., stroke) is decomposed into a direct effect and an indirect effect which is mediated through the effect of CPAP on mediating outcomes (e.g., AHI).

Demonstrating the validity of such mechanisms of action is important because it can determine the pathways through which CPAP leads to improved health outcomes, point to pathophysiological mechanisms, and indicate new targets for intervention.

Various study design and analytical approaches exist to answer questions of whether intermediate outcomes are mediators of effect. One simple approach involves estimating the association between CPAP and the outcome both with and without controlling for AHI (e.g., by means of a regression model).[172, 173] The magnitude of the difference in the CPAP effect when controlling versus not controlling for AHI indicates the fraction of the effect mediated by (or operated through) AHI. For example, if the unadjusted OR for all-cause mortality with CPAP is 0.87 and the OR after adjusting for change in AHI is 0.93, then it can be estimated that 54 percent of the effect of CPAP is mediated by change in AHI: $(1-0.93)/(1-0.87) = 0.54$.

In recent years, methodological advances have cast mediation analysis within the potential outcomes framework.[170, 171, 174, 175] Within this framework, the following effects are estimated:

1. Natural direct effect (NDE): the difference in outcome (e.g., mortality) if all patients were assigned to the test intervention (e.g., CPAP) versus if all patients were assigned to the control intervention (e.g., no CPAP), with the putative mediator outcome (e.g., AHI) held constant for each patient and equal to the outcome value that this patient would have if they were assigned to the control intervention, and
2. Natural indirect effect (NIE): the change in mortality risk if a patient received the intervention and the mediator outcome level changed from its value under the control intervention to its value under the test intervention.

Accordingly, the proportion mediated through AHI is $(NDE \times (NIE - 1))/(NDE \times NIE - 1)$.

Estimation of the above effects requires either randomized trials or well-designed observational studies that carefully apply the assumptions required for the implementation of the potential outcomes framework. However, as we noted, the available evidence was not adequate to answer the mediation questions of interest. In particular, no studies performed a formal mediation analysis by estimating the effects mentioned above.

Surrogacy Analyses

A related question to mediation is that of surrogacy: whether AHI (or other related measures) can be considered a valid endpoint to be measured instead of clinically important outcomes.

AHI, for example, would be a valid surrogate outcome in CPAP trials if and only if CPAP has an effect on both AHI and the health outcome. It would not be necessary for AHI to be a mediator of the health outcome, as long as the change in AHI tracks with, and therefore predicts, the effect of CPAP on the health outcome. It is important to note that while an intermediate measure may be a valid surrogate measure for some conditions, specific populations, and outcomes, this would not imply that it is a valid surrogate measure in all instances. Thus, if AHI is found to be a valid surrogate outcome for CHF in patients with very severe OSA (e.g., AHI ≥45) being treated with CPAP, this would *not* imply that it is necessarily valid as a surrogate measure for MI in patients with mild OSA (e.g., AHI 10 with sleepiness symptoms) being treated with MAD.

Statistically, surrogacy can be evaluated using principal stratification as well as the meta-analytic approach that we implemented in this review. Principal stratification is an approach to evaluate surrogacy that relies of effect estimation for four strata (called "principal strata") defined by the combination of high and low values of the surrogate measure (or other stratification) under treatment assignment to both the intervention and no intervention (or other comparison) groups.[176] In the example relevant to this review, AHI values are stratified (e.g. high AHI, low AHI) under CPAP and no CPAP treatment assignments. The direct and indirect effects of CPAP on the outcome can be estimated by stratifying on the mediator group (low or high AHI values). However, for principal stratification to be efficiently used, the associations of CPAP with both the outcome and AHI need to be reported in the same sample of participants (along with sufficient data on standard errors or other uncertainty measures), a practice that was not common across the studies that we identified.

Figure 17. Mediation graph
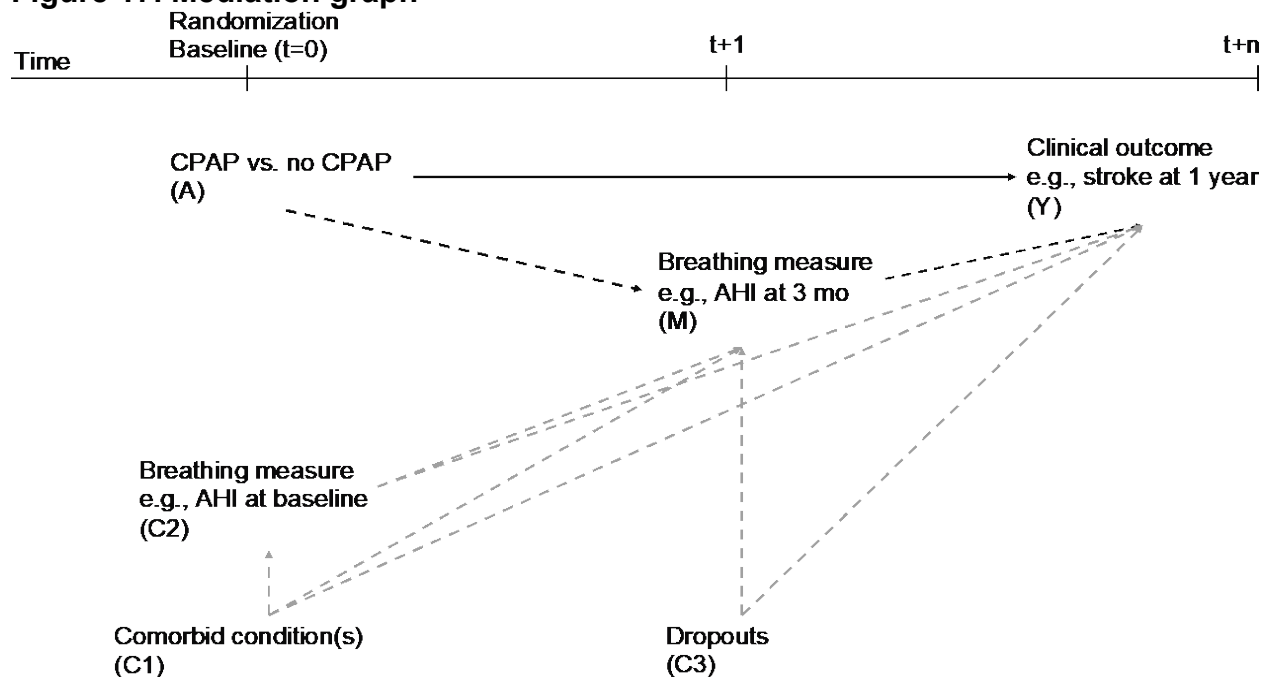
Directed acyclic graph (DAG) of mediation of a breathing measure on a health outcome. The direct effect of treatment is represented by the solid black line and arrow from *A* to *Y* without passing through a mediator *M*. Indirect effects are represented by the black dashed lines and arrows from *A* to *M* and then to *Y*. The gray dashed lines represent potential confounders *C1 to C3*.

Abbreviations: AHI = apnea-hypopnea index, CPAP = continuous positive airway pressure.

Discussion

Findings in Relation to the Decisional Dilemma(s)

The primary clinical question addressed by this systematic review about the effectiveness of continuous positive airway pressure (CPAP) devices on long-term clinically important outcomes in patients with obstructive sleep apnea (OSA) remains largely unanswered. Much of the evidence base is sparse. The randomized controlled trials (RCTs) did not provide statistically significant and clinically meaningful results that could establish a causal relationship. The nonrandomized comparative studies (NRCSs) provided some significant associations between CPAP use and decreased risk of clinical outcomes, particularly for all-cause mortality, but are only suggestive of a possible benefit of CPAP use. Despite at least fair methodological quality of many of the studies, the estimates of comparative effectiveness were largely imprecise. Thus, there is at best low strength of evidence (SoE), mostly of no evidence of a long-term effect of CPAP on clinical outcomes. If one also considers the adjusted NRCS evidence, there is, at best, low SoE that CPAP use may be associated with a decreased risk of all-cause mortality. The low SoE suggests that we have limited confidence that the estimate of effect lies close to the true effect for this outcome. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect."[29]

Definitions of Breathing Measures Findings

Across all studies included in this systematic review, the criteria used to define sleep study measures (apnea, hypopnea, and oxygen desaturation) were highly inconsistent, even among studies stating that definitions are based on the same standard criteria. A major part of the problem is inadequate reporting of specific criteria used. Most studies (60%) did not fully and explicitly report the definitions of sleep study measures used. Even among those that did report definitions, it was often difficult to discern the actual criteria used. One issue is that, while many studies cited specific American Academy of Sleep Medicine (AASM) criteria (based on update year), those that cited the same specific AASM criteria did not universally apply the same actual criteria. For example, among the eight studies that cited the ASSM 1999 criteria, three explicitly defined apnea as 100 percent airflow cessation and two used a 90 percent threshold (3 did not report an apnea definition). Among the six studies that cited the AASM 1997 criteria, three each used either a 3 percent or a 4 percent threshold to define oxygen desaturation. However, it was generally unclear whether authors were citing the AASM (or other criteria) as the source for the set of criteria used or to cite a specific criterion (e.g., only hypopnea), which further complicated interpretation of how apnea-hypopnea index (AHI) or other sleep measures were defined and how patients were, thus, selected for inclusion in the study. We could not discern the extent to which the variability was due to flexibility (or lack of clarity) of the often-changing AASM (and other) criteria, to possible misinterpretations of the AASM criteria, to deliberate alterations in the standard criteria (e.g., to accommodate local practices), or simply to incomplete reporting of study methodology. Nevertheless, for most studies, it would be very difficult for an outside researcher or clinician to replicate how AHI or oxygen desaturation index (ODI), among other various parameters, were defined and/or to determine which patients would be eligible for study inclusion.

CPAP Versus No CPAP Findings

All conclusions regarding the relative effect of CPAP versus no CPAP on clinically important outcomes are at best of low SoE. We have limited confidence that the summary estimates (and their confidence intervals) are close to the true effect or that the findings are stable. Because of this uncertainty, future, good quality evidence could alter these conclusions.

The currently available RCTs do not demonstrate that CPAP treatment affects all-cause mortality or various cardiovascular (CV) outcomes, clinically important changes in psychosocial measures, or other clinically important outcomes. However, these conclusions do not definitively determine or imply that CPAP has been proven to be ineffective to reduce adverse health outcomes. It is unclear whether the failure to find an effect of CPAP treatment on long-term health outcomes is related to a lack of power, the result of insufficient followup duration, or due to an actual lack of effect of CPAP. Notably, only two of the eligible RCTs were adequately powered for a composite CV outcome, and none were powered for other CV or mortality outcomes.

Regarding all-cause mortality, while most of the RCTs appeared to favor CPAP prescription in direction, all were statistically nonsignificant. The summary effect size, by meta-analysis, was somewhat imprecise failing to support an effect of CPAP. Despite inclusion of mostly high-risk patients (related to CV disease) with at least moderately severe OSA (e.g., AHI ≥15), the RCTs had low event rates. These event rates may have been related to relatively short duration of followup (2 to about 5 years), the frequent exclusion of patients with excessive daytime sleepiness, and the levels of CPAP adherence (38% to 64% of participants). Within RCTs, no significant differences were found between intention-to-treat (ITT) analyses of all participants and analyses of adherent users. The numbers need to treat (NNT) or harm (NNH) suggested by RCTs ranged from an NNT of 47 to an NNH of 70. In the SAVE (Sleep Apnea cardioVascular Endpoints) study, the largest RCT reporting all-cause mortality, the NNT was 426, with a 95 percent confidence interval ranging from one prevented death per 65 treated individuals to one additional death per 93 treated individuals. However, since determination of NNT to prevent one death is dependent on a specific, specified underlying risk of death, estimates of NNT are not generalizable beyond each individual study.

In contrast, the NRCSs all found statistically significant associations between CPAP (prescription or use) and reduced risk of all-cause death, with mostly stronger effect sizes than the RCTs. The NRCSs were generally longer in duration than the RCTs (3 to 11 years) with mostly higher event rates. They also mostly included all patients with at least moderately severe OSA, regardless of sleepiness symptoms. However, the NRCSs were mostly at high risk of bias related to conducting simple, often incomplete, regression analyses. Even for the two NRCSs that conducted propensity score matching, it was not clear the studies could adequately adjust for the inherent, underlying differences among patients who choose to start and be adherent to CPAP and those who do not. The extent to which reporting bias affects the summary findings, particularly among studies without *a priori* protocols, is unclear.

The exploratory combinations of RCTs and NRCSs suggest that CPAP may reduce the risk of, or may be associated with a reduced risk of, all-cause mortality, but did not change conclusions about a lack of evidence of an effect on other outcomes. Even though we restricted our analyses to NRCSs that were adjusted for potential confounders (with in some cases propensity score-matched analyses), the NRCSs reported associations between CPAP use and outcomes, not causal evidence of effects. Such associations may support RCT evidence, but do not provide the basis for evidence of effect *per se*.

An important concern about the NRCSs is that their comparator groups were non-users of CPAP, who either chose not to use CPAP or whose clinicians did not prescribe CPAP. Thus, the groups of CPAP users (especially adherent CPAP users) and non-users may be fundamentally different in ways that adjustment for measurable confounders cannot fully account for. Furthermore, the associated differences in precision of estimates of association between CPAP use and outcomes may explain the difference in statistical significance between all-cause mortality and other CV outcomes (including CV death). The single NRCS that analyzed revascularization found a strong, significant association with lower risk of surgery and the studies reporting composite CV outcomes mostly found associations favoring CPAP use (even though, cumulatively, they did not provide adequate evidence to support a difference in outcomes between CPAP and no CPAP).

While the NRCSs lend support for a possible association between CPAP use and all-cause mortality, the RCTs and NRCSs do not support that CPAP reduces the specific causes of death that might be ascribable to OSA (e.g., CV events, accidents). While there is no clear explanation for this possible discrepancy, the eligible NRCSs reported on few CV outcomes (only CV death, coronary artery revascularization, and composite CV outcomes). Notably, more than 10-times as many adults with OSA were included in studies reporting all-cause mortality than for other CV outcomes. It is possible that the current evidence base does not reveal differences in risk of specific outcomes with CPAP use that actually exist. Furthermore, these (not statistically significant) differences in risk of specific outcomes may be additive or multiplicative such that cumulatively they have a discernable effect on risk of all-cause mortality. Alternatively, the conclusions regarding specific CV outcomes may be accurate (i.e., no effect) and the low SoE finding of an effect on all-cause mortality may be spurious.

Of note, it has been postulated that adherence as measured by hours of utilization per night or as CPAP termination rates may impact CPAP outcome results. In a recent large insurance cohort study, the termination rates were 23, 37 and 48 percent at 1, 2, and 3 years, respectively.[177] In this Technology Assessment, among RCTs that reported data, adherence rates ranged from 35 to 64 percent, without clear explanation for the variation across studies. For most categorical outcomes (including all-cause mortality, composite CV events, other CV outcomes, incident hypertension, incident diabetes), effect sizes appear to be somewhat stronger in the CPAP adherent user analyses than the ITT or "as-prescribed" analyses (of all who were prescribed CPAP), but almost universally the differences between analyses were not statistically significant (both within and across studies) and the differences were generally small. It is unclear whether the lack of significant differences between adherent user and intention-to-treat analyses is due to a lack of power to indicate a difference in effect or whether the use of CPAP has any real measurable effect on the occurrence of CV events. However, such comparisons between ITT (or, in NRCSs, "as-prescribed" analyses) and adherent user analyses, are complicated by the fact that patients who are adherent have different characteristics than those who are nonadherent (or stop using CPAP). Adherence is not a random event. Poor adherence may be due to remediable or non-remediable reasons. In the large insurance cohort, early termination was more likely in those who were younger (under 41 years) and older (over 81 years) and if there was at least one comorbidity, especially diabetes.[177] How well CPAP works for immediate symptom relief (e.g., sleepiness, snoring) may be a more important determinant of adherence than possible effects on long-term health outcomes. This may explain why some analyses suggested that patients with worse baseline AHI or Epworth Sleepiness Scale (ESS) scores were more adherent.[11] Conversely, patients may not adhere with CPAP use due to factors such as claustrophobia, mask

discomfort, convenience, vanity, and complexity of management of multiple chronic conditions.[11] It is also unclear to what extent diligence with adherence may correlate with other health practices such as tobacco use, attempts at weight loss, exercise, other patient characteristics (such as age, psychosocial factors), device features (such as-use of heated humidification), and others.[177, 178] It is likely that the NRCSs and adherent user analyses were not able to fully control for the inherent differences between CPAP users and non-users/non-adherent individuals. Overall, we believe it is likely that studies comparing CPAP use (adherence) with nonuse (nonadherence) may be biased toward increased effectiveness of CPAP.

Notably, the within-study comparisons based on adherence were all *post hoc* analyses, and the between-study comparisons should be considered only hypothesis-generating. If there is, in fact, no difference in effect between adherent and nonadherent CPAP users, this may suggest that either any benefit seen with CPAP use is not due to CPAP itself, but is due to some other behavior or action by CPAP users (maybe such as increased communication with the sleep clinic) or that even the "low dose" of CPAP achieved by nonadherent users is effective. However, such explanations are just conjecture and would need to be explored critically.

Regarding non-CV outcomes, the RCTs provide low SoE that CPAP improves measures of depression and anxiety symptoms, executive cognitive function, and quality of life, but, importantly, these changes are small and not clinically significant. Four trials evaluated the link between CPAP adherence and the Mental Component Score of the Short Form (SF) 36 but had conflicting findings. Furthermore, no studies evaluated health outcomes related to these measures (e.g., diagnoses of depression, anxiety, or dementia) and none of these measures have been adequately validated in the OSA population. Additionally, there is low SoE that CPAP has no effect on risk of traffic accidents and incident diabetes. Relatively few NRCSs with adjustment for confounding reported on non-CV health outcomes; thus, most conclusions were unchanged when also considering the NRCSs.

For numerous health outcomes of interest, there is insufficient evidence, primarily because only a single study evaluated a given outcome or because effect estimates were highly imprecise (such that neither large benefits nor large harms could be excluded). These included risks of transient ischemic attacks, angina, coronary artery (or other major artery) revascularization, congestive heart failure, atrial fibrillation, hypertension (incident or resolution), sexual function, and days of work missed.

Few studies evaluated potential subgroup differences and, across studies for any given outcome, effect size estimates were either imprecise or were similar in magnitude with each other (or both). Thus, the comparative study literature base (RCTs and adjusted NRCSs, together) does not provide adequate evidence to suggest which patients may benefit most (or least) from CPAP treatment. In particular, very few studies compared patients with different levels of "severity" of OSA (based on, for example, baseline apnea-hypopnea index [AHI], oxygen desaturation index [ODI], symptomatology [e.g., by Epworth Sleepiness Score (ESS)], or by variable criteria to diagnose OSA [e.g., by AHI alone or also including comorbidities and symptoms]). One NRCS (ONSLEEP) reported adjusted hazard ratios (HR) among different subgroups by AHI,[130] but failed to report an appropriate analysis to support possible heterogeneity of treatment effect. The HR for all-cause mortality was about half for participants with AHI ≥30 compared with those with AHI 5 to 29 (implying a stronger effect in those with higher AHI levels), but no interaction term was included to determine whether the HRs in the two subgroups were statistically different. One RCT (APPLES) compared various measures of cognitive function in subgroups of participants with mild (AHI 10 to 15), moderate (AHI 15.1 to

30), and severe (AHI >30) OSA,[97, 98] but differences (CPAP vs. no CPAP) were similar among subgroups. Another NRCS (Bjornsdottir 2015) reported similar net differences in quality of life as measured by the Short Form 12 (SF-12) in subgroups based on sleepiness symptoms (ESS ≥10 vs. <10).[120]

As was discussed in more detail above (*Definitions of Breathing Measures*), studies were highly variable in specific criteria used both to define breathing events (apneas, hypopneas, oxygen desaturations, and arousals) and to diagnose OSA (or to define eligibility for CPAP use). Moreover, the majority of studies were insufficiently clear regarding specific criteria used. Several studies also included a mix of criteria. Thus, it is impossible to look across studies and discern any patterns regarding degree of clinical effect size of CPAP based on definitions of breathing measures or criteria to prescribe CPAP. However, even if such criteria were more clearly described across studies, it would still be the case that studies were generally statistically similar in their findings; thus, any differences in effect across studies would be difficult to evaluate. Thus, even for assessments of better-defined factors, such as duration of followup or particular CPAP device, no conclusions regarding heterogeneity of treatment effect were feasible across studies.

CPAP Versus Other Active Treatments Findings

Long-term health outcome data from RCTs or NRCSs with adjustment for confounding are relatively sparse. Based on only two relatively small RCTs, there is low SoE of no statistically or clinically significant differences between use of CPAP and mandibular advancement devices in changes in depression or anxiety symptom scores, and insufficient evidence (based on a single study each) regarding changes in quality of life or sexual function measures. Similarly, based on three small studies, no differences were found in OSA-related functional status (by the Functional Outcomes of Sleep Questionnaire [FOSQ], which may have limited validity), with low SoE. Further conclusions regarding other health outcomes or potential heterogeneity of treatment effect are not feasible, given the limited evidence base.

Intermediate and Surrogate Measures Findings

The large majority of screened RCTs did not meet eligibility criteria for this systematic review primarily because they did not report on long-term clinically important health outcomes. Instead, most reported on only (generally shorter-term) intermediate outcomes, including AHI, ODI, or ESS. We, thus, sought to evaluate the evidence regarding whether changes in these (and related) intermediate measures during the course of longitudinal studies are valid surrogate or mediator outcomes for the long-term health outcomes. To qualify, the studies had to, at a minimum, report sufficient data to analyze whether changes in intermediate measures correlated with the health outcomes. We found no study that, itself, evaluated whether change in any intermediate outcome may be a valid surrogate or mediator measure. We did not assess the validity of single measurements of breathing or sleepiness measures (e.g., measured pretreatment) as predictors of outcomes or treatment effect. In the Results section *Ideal Study Design to Establish Validity of Mediator (Intermediate) and Surrogate Measures* we describe what a future study would need to do to appropriately analyze their data.

Although we did not systematically review noncomparative analyses of possible intermediate or surrogate measures and health outcomes (e.g., single group analyses), we note that in the studies eligible for Key Question 1, AHI, ODI, and other breathing measures improved by expected large values with CPAP use. However, these within-group effects on CPAP did not

clearly translate to effects on health outcomes. For example, in the SAVE trial (although it is not clear how changes in AHI were measured), AHI was reported to decrease from 29.0 prior to CPAP to an average of 3.7 events per hour (when CPAP was being used) over the mean 3.7 years of followup on CPAP (among those randomized to the device).

Despite the large improvement in AHI in SAVE, CV events and other clinically important health outcomes did not improve and, in fact, the HR for the study's primary outcome (composite CV events) nonsignificantly favored the no CPAP group (HR = 1.10, P = 0.34). Similarly, studies reporting clinically significant reductions in ESS did not report significant reductions in clinical event risk. However, as noted, these findings do not provide direct and definitive evidence as to whether AHI (or ESS) are valid (or invalid) surrogate or intermediate measures for health outcomes.

There is insufficient evidence to support alternative explanations. It may be that the lack of apparent correlation relates to an issue of power (there are adequately powered analyses of large effects on AHI and ESS, but inadequately powered, relatively smaller effects on health outcome risk reduction). There may also be a threshold effect such that a minimum (large) improvement in AHI or ESS is necessary to correspond with clinical event risk reduction. If only patients with the largest reductions in AHI or ESS have concomitant risk reduction, then those patients with smaller reductions may dilute the average effect sizes. It may be that changes in AHI (and ESS) are not predictive of risk of clinical events. This may imply that CPAP is effective to reduce episodes of apnea and hypopnea and may be effective to improve symptoms (as measured by ESS), but these effects do not impact health outcomes.

Relatively few studies reported on both changes in (potential) intermediate/surrogate measures (e.g., AHI, ESS) and on relative effects between groups (i.e., net differences in continuous outcome measures or relative effect estimates of categorical outcomes) for the same sets of long-term health outcomes. Correlation analysis by metaregression (i.e., correlation of measures based on mean estimates across studies) is intrinsically of low power (numerous datapoints, i.e., studies, are needed to achieve statistical significance). Thus, unsurprisingly, the analyses across studies failed to find statistically significant correlations between intermediate / surrogate measures and health outcomes. Furthermore, there was no evidence to determine the relative "strength" of associations between AHI (or other breathing measures) and ESS (or other symptom measures) and the health outcomes of interest. Thus, it remains unclear whether changes in AHI (or other intermediate/surrogate outcomes) can be considered to be surrogate or mediator measures for long-term health outcomes in studies of treatment effectiveness. Until such validation has been conducted, it may be inappropriate to assume that changes (e.g., improvements) in intermediate or surrogate outcomes would correlate with long-term health outcomes.

Other Recent Systematic Reviews

The AASM recently commissioned a broad systematic review of CPAP versus no CPAP or other PAP treatments[179] to support their clinical practice guideline.[180] They included a broader range of studies and outcomes, including unadjusted observational studies, short-term outcomes, and various intermediate and symptom measures. In brief, they found that CPAP reduces AHI and improves sleepiness symptoms. In addition, the review found that CPAP improves (mostly short-term) sleep-related, but not overall, quality of life, reduces (mostly short-term) blood pressure (in patients with hypertension). But CPAP did not reduce (mostly short-term) blood glucose (in patients with or without type 2 diabetes), improve neurocognitive function, or

improve (short-term) anxiety or depression scores by clinically significant degrees. Their findings were inconclusive regarding all-cause mortality and composite CV events. RCTs found no effect and observational studies found variable, but, in aggregate, statistically significant effects. The difference in effects between randomized and observational studies was ascribed to average (not patient level) differences in CPAP adherence, but the review authors reported that neither the RCTs nor the observational studies considered the impact of unadjusted confounding from adherence to non-CPAP therapies for healthy lifestyle habits.[179] CPAP did not affect crash rates in driving simulation studies, but actual motor vehicle crashes were less common after start of CPAP than before.

Older systematic reviews, including a 2006 Cochrane review[10] of 36 RCTs (CPAP vs no CPAP or oral devices) and a 2011 Agency for Health Care Research and Quality (AHRQ) review[11] of 56 RCTs (CPAP vs. no CPAP or oral devices) conducted by the current review's authors, included mostly studies with short-term follow-up. Both reviews found statistically significant effects of CPAP on AHI, ESS, and blood pressure (although with some inconsistency regarding effect on blood pressure); lack of consistent statistically significant effect on quality of life or neurocognitive and psychological tests; and insufficient evidence on clinical events and blood glucose.

Strengths and Limitations

Our conclusions about the effect of CPAP on health outcomes are restricted to the findings about specific outcomes from the eligible studies; namely, studies conducting direct comparisons between intervention options (RCTs and adjusted NRCSs only, including analyses of adherent and nonadherent CPAP users) that reported long-term clinical events and certain test scores (such as quality of life). We did not evaluate sleepiness, other symptoms, or intermediate outcomes (including AHI, blood pressure, or glucose).

CPAP (Strengths and Limitations)

The evidence base examined in this systematic review is limited in its ability to establish the long-term effects of CPAP treatment on health outcomes in patients with OSA. As noted, for most health outcomes, an insufficient number of studies and/or an insufficient number of study participants have been evaluated to allow for precise estimates of effects (or lack of effect). Few RCTs were powered for health outcomes of interest, including only three powered for composite CV outcomes (one of which ended up being underpowered) and one for neurocognitive outcomes.

The emphasis on composite CV outcomes in most trials that reported long-term clinical events hampers a clear understanding of the effect of CPAP on specific well-characterized cardio/cerebrovascular events of interest to patients and clinicians. Better standardized application of outcomes (particularly "composite CV outcomes") would improve interpretation of study findings and allow for more coherent evidence summary (e.g., meta-analysis).

This concern is not restricted to OSA studies. A recent systematic review of CV outcome definitions for the Standardized Endpoints in Perioperative Medicine (StEP) initiative found at least five different definitions for major adverse CV events (MACE).[181] Notably, none of the major trials in the StEP review used the same definition. In a Delphi process, the highest-rated definition included cardiac death, myocardial infarction, and nonfatal cardiac arrest (notably, omitting revascularization, unstable angina, heart failure, cerebrovascular events, and arrhythmias, which were all commonly included in CPAP study composite CV outcomes).[182]

Similarly, the European Medicines Agency, in their 2016 assessment of CV safety profile of medicinal products, recommends prioritizing MACE for CV outcomes studies, but with a different definition (CV death, nonfatal myocardial infarction, and non-fatal stroke).[183]

An inherent limitation of the literature is the great variability in, and the often poor descriptions of, how breathing and sleep measures were defined in studies and, thus, exactly how OSA was diagnosed (or the decision to treat with CPAP was made). Separate from whether AHI is a valid measure of OSA severity or whether change in AHI is a valid intermediate or surrogate measure of any health outcome, the variability in how AHI is defined leads to a number of issues related to assessing the evidence base. First, and possibly most important, the lack of clarity about which patients were enrolled in the studies limits our ability to accurately determine for which patients the studies are most applicable. This limitation applies mostly to patients with relatively low AHI (e.g., 10 to 20), for whom small differences in thresholds (or definitions of) apneas, hypopneas, and oxygen desaturation can have large effects on their estimated AHI (or ODI) value, and thus their potential eligibility for CPAP treatment. Second, the lack of clarity hampers assessments about which patients may clinically benefit most (or least) from CPAP (assuming any patients do). Even had there been more (statistical) heterogeneity across studies in the effects of CPAP treatment, we would have been unable to meaningfully evaluate whether any of the heterogeneity could have been explained by differences across studies in AHI (or ODI) definitions. For example, studies that found different effects in subgroups based on AHI categories (e.g., 10-15, 15-30, ≥30) would be difficult to interpret if we cannot be confident about which patients would fall in each category because of variable definitions of how AHI was defined. Not only does this hamper our understanding of the evidence, but it results in a failure to allow the clinical studies to help determine which particular criteria for breathing measure and OSA definitions are most clinically important. Sleep study measures should be defined to help capture those patients who would most benefit from treatment. In theory, clinical trials should provide the best evidence to help refine the definitions of the measures.

The high variability as to how AHI was defined calls into question the clarity and effectiveness of the standardized criteria (such as those by AASM). One would expect that different research groups applying the same criteria (e.g., the AASM 2012 update) would consistently use exactly the same definitions for apnea, hypopnea, and oxygen desaturation. However, this was clearly not the case. Although, we acknowledge that it is not clear to us whether AASM is allowing too much leeway for each sleep center to define criteria as they see fit or if polysomnographic technologists and sleep physicians are misinterpreting or misapplying the criteria.

Few studies reported on treatment effects in subgroups of patients and none of the studies reported appropriate interaction analyses to determine whether any differences seen between subgroups were significantly different from each other. This is a common failing of studies reporting subgroup effects. In large part, the lack of subgroup analyses was likely due to the relatively small numbers of participants included in individual trials. The SAVE trial is the largest RCT to date comparing CPAP to no CPAP with long-term health outcomes; it was of modest size, with 2687 participants followed for about 4 years. Even with restriction to patients with underlying CVD or cerebrovascular disease (with the goal of enriching the sample for CV outcomes), it was designed to be powered only for a composite CV outcome (CV death, nonfatal MI, nonfatal stroke, and any hospitalization for unstable angina, heart failure, or transient ischemic attack). Even then, the study failed to find a statistically significant effect at the effect size found by the study (HR = 1.10). With relatively small studies, that were thus underpowered

to find statistically significant effects for most incident outcomes, it is unlikely that the studies could have uncovered statistically significant, and maybe even clinically important, differences in effect across subgroups of patients. Therefore, while studies mostly found no evidence of average differences in effect (or at best, low SoE of a possible small average benefit), the evidence fails to address whether any particular sets of patients may clearly benefit from CPAP. While on average, within studies, patients did not have a reduction in CV death (for example), it is unknown whether particular patients in the studies (e.g., those with morbid obesity or those with AHI >60) did, in fact, benefit from CPAP.

Most of the RCTs were deemed to be at moderate risk of bias. The most common reason for downgrading the quality of the studies was a lack of blinding of participants or their clinicians. Only two, relatively small, relatively short-term, RCTs (APPLES [Apnea Positive Pressure Long-term Efficacy Study] and BestAIR used sham CPAP and thus blinded participants.[97, 101] While it is apparent that implementing sham CPAP for years in a large group of study participants would be difficult, the lack of blinding raises the possibility of differences in whether patients treated without CPAP were managed differently than patients not given CPAP, and the degree to which this may have impacted long-term health outcomes. Other methodological concerns among the RCTs were related to loss to followup, some crossovers, and other related issues. Nevertheless, overall the risk of bias among the RCTs was not a major concern. We were unable to discern differences in reported effect sizes related to our quality assessments of the trials.

While we restricted inclusion of NRCSs regarding effect of CPAP to those with multivariable analyses, there remain concerns that patients selected to be treated with CPAP (or self-selected to use or comply with CPAP) are inherently different from patients not treated with (and/or not adherent with) CPAP to such a degree that even well-adjusted analyses are biased toward CPAP users (who might, for example, be more concerned with their health, well-being, or risk of long-term outcomes). Only three NRCSs (Bjornsdottir 2015, Lisan 2019, ONSLEEP) used propensity score analyses (that adjust for the factors that predict for choice to use CPAP); although, even propensity score matching may not account for unmeasured differences between groups. Nevertheless, for most outcomes effect sizes from NRCSs tended to be consistent with those from RCTs.

Intermediate and Surrogate Measures (Strengths and Limitations)

The evidence regarding the validity of breathing and sleepiness symptom measures as intermediate or surrogate measures (whether surrogate or mediator measures) for health outcomes is insufficient to make conclusions. Importantly, no comparative studies with long-term health outcomes evaluated potential correlations of intermediate/surrogate and health outcomes within the studies. Our attempt to glean potential evidence for correlations from across-study meta-regression was unsuccessful largely due to too-small numbers of studies that reported both changes in intermediate or surrogate measures and effect sizes. Studies tended to report AHI (or ODI) and ESS only at study baseline as descriptors of the included populations, but not in followup.

Applicability

The largest RCT (SAVE) was, as noted, restricted to participants with prior CV or cerebrovascular disease who had relatively "severe" OSA (ODI ≥12, reported to be equivalent to AHI ≥30) and were between 40 and 75 years old (mean 61).[12] About 80 percent of participants

were male and about two-thirds were Asian (primarily from China; although the study was conducted in Australia, Brazil, India, New Zealand, Spain, and China). Patients with extreme excessive daytime sleepiness (ESS >15) were excluded. The study also used a study-specific device to determine eligibility for CPAP treatment. Since the SAVE trial was relatively large compared to the other RCTs, outcomes reported by SAVE (most CV-related outcomes, accidents, incident diabetes, and depression and anxiety), our conclusions could be considered most applicable to individuals who would have been eligible for the SAVE trial. Consistent with SAVE, other RCTs generally included patients at increased risk of CV outcomes or death. Most other RCTs also excluded patients with excessive daytime sleepiness (ESS ≥10). Another concern related to the applicability of the RCTs is that individuals (and their clinicians) who would be willing to be randomized to CPAP or no CPAP for long-term studies may be inherently different from typical patients seen in practice. Nonetheless, for most outcomes, we did not find substantive differences in the findings of RCTs and the generally more applicable NRCSs (which used more inclusive eligibility criteria). However, many of the NRCSs provided poor descriptions of patient eligibility and/or CPAP treatment, hampering the assessment of the applicability of these studies.

In terms of potential applicability to the Medicare population, only three studies, one RCT (PREDICT) and two NRCSs (Crawford-Achour 2015 and Ou 2015) specifically focused only older adults, but had very small numbers of patients, n = 231, 124, and 126 respectively. An additional NRCS (Jennum 2015) reported on a subgroup of older adults (n = 6719), with similar findings compared with other age groups.

Implications

Our systematic review of direct comparative studies mostly found, with low SoE, no evidence of statistically of clinically significant effects of CPAP on long-term health outcomes of interest. The overall lack of RCT and adjusted NRCS evidence supporting an effect of CPAP on long-term health outcomes stands in contrast to the substantive effect of CPAP on sleep measures such as AHI and symptoms, as evaluated by ESS.[11] We were also unable to make meaningful conclusions regarding the validity of breathing or symptom measures as intermediate or surrogate measures for health outcomes.

The most important limitations were imprecise (frequently highly imprecise) estimates of effect and sparse evidence. Most of the important clinical events of interest occur relatively infrequently (even among selected populations) and may take many years to manifest. For example, the all-cause mortality rate in the largest RCT (SAVE) was only 3 percent after an average of almost 4 years of followup.

The RCTs and NRCSs do not provide evidence of clinically meaningful effects on psychosocial measures. Improvements in the severity of symptoms or associated comorbidities with CPAP use, such as mental health, cognitive function, and quality of life, while in some cases statistically significant, were small and below thresholds for clinically meaningful differences in effect. Regarding quality of life and functional status, the evaluated measures have mostly not been validated in adults with OSA and the quality of life measures might not qualify as health-related quality of life measures.

Thus, overall, the current RCT evidence base provides conclusions based on low SoE regarding the potential clinical benefits of CPAP for patients with OSA over the long-term (e.g., after at least 1 year of treatment), particularly related to CV events, but also other outcomes such as psychosocial measures. For the most part, the associations reported by adjusted NRCSs are

consistent with the direction and magnitude of the effects of CPAP in the RCTs. Although, for all-cause mortality, the NRCSs found stronger associations than the RCTs, which led to an overall conclusion of possible effectiveness of CPAP in longer term followup—albeit with a low strength of evidence.

Future RCTs and NRCSs that more effectively control for the inherent differences between those who are adherent and those who are not are needed to confirm or refute the conclusions. Alternatively, the treatment may have limitations in its utility because aspects of treatment choice and user adherence may not be remediable.

The generally low SoE regarding the use of CPAP to prevent long-term health outcomes (for most outcomes) is in contrast with high SoE of the effect of CPAP to improve AHI and other sleep and symptom measures.[10, 11] Given the relative ease of conducting short-term studies of these putative intermediate/surrogate measures, it is not surprising that the large majority of RCTs have focused on these outcomes. Most researchers, at least implicitly, assume that AHI is the "best" intermediate outcome to measure. Despite biological plausibility, it is arguably incorrect to simply assume that changes in AHI are associated with, or mediate, improvements in long-term health outcomes. Other measures, that may have better biological plausibility, such as measures of cumulative exposure to reduced inhaled air during a night of sleep (e.g., time spent with oxygen desaturation below a threshold) are rarely evaluated. From prior comparative effectiveness research conducted for the Agency for Healthcare Research and Quality (AHRQ),[11] it is well-understood that CPAP is effective to improve (reduce) AHI—to a greater extent in patients with higher AHI levels without CPAP—but there is not adequate evidence to support the contention that changes in AHI translate to improvements in health outcomes. Until well-conducted studies test the validity of the range of sleep (and symptom) measures to predict important health outcomes, it remains difficult to interpret trials of putative intermediate outcomes and to determine whether their findings (e.g., of reduced AHI) are of any value to patients.

Although existing systematic reviews have reported that CPAP improves sleepiness symptoms, the evidence regarding other health outcomes is more tentative and additional studies are needed before we have a clear understanding of the potential effects of CPAP on long-term outcomes for patients with OSA. Furthermore, there is inadequate evidence to support whether any particular group of patients may benefit to a greater or lesser degree from CPAP treatment to affect health outcomes. Future high-quality studies could provide more definitive results.

Future Research

The Centers for Medicare & Medicaid Services (CMS) commissioned this Technology Assessment from AHRQ to evaluate the evidence on improvement of long-term health outcomes of CPAP treatment in patients with OSA, as well as the validity of criteria used as surrogate outcomes (e.g., AHI), all in aid of coverage decision making. In addition, the report has identified gaps that should direct future research.

This Technology Assessment is notable for its inability to identify evidence that would warrant drawing anything but low SoE conclusions about the efficacy of CPAP. It is also notable for the limitations regarding surrogate markers for OSA (e.g., lack of validation) and the reliance of the sleep apnea research on these parameters.

Despite decades of research that have made the implicit assumption that improvements in intermediate measures, such as AHI, ODI, and ESS, are associated with improvements in long-term health outcomes, we found no well-designed and implemented RCTs that have verified this

hypothesis. We also note the lack of consistency across studies (and sometimes within studies) of definitions and criteria used to assess sleep study measures and to diagnose OSA. It is critical that both verified and appropriate criteria are applied and that these criteria are applied consistently and accurately across patients. The reader of a study should be able to identify the sleep study devices that were used, where sleep studies were done, how sleep and breathing measures were measured, what thresholds and other criteria were applied to each measure, and how study eligibility criteria (including such factors as AHI) were implemented.

Various investigators over the years have raised questions about the value of the evidence base supporting CPAP. Most notably, in 2008, the American College of Cardiology and the American Heart Association noted the poor quality of data in this area and cited poor study design.[184] In 2010, the SAVE investigators cited these deficiencies as part of their rationale for conducting a large, long-term, placebo-controlled, randomized study with hard clinical endpoints.[185] However, despite the attempt, the SAVE study did not identify a CV benefit from CPAP. It was underpowered for most clinical outcomes and relied on a composite CV endpoint. In addition, despite the lack of CV benefit, AHI improved, raising questions about the linkage between measurement parameters and outcomes.

There are no other large RCTs with hard clinical endpoints of sufficient duration, and with proper execution and analysis. Adherence, and whether it is patient or device related, has not been rigorously assessed.[177, 178] The largest studies were NRCSs. These were often undertaken because CPAP efficacy, often based on AHI-type parameters, was presumed and placebo treatment was thought to be inappropriate and perhaps unethical. However, as this review demonstrates, there remains clinical equipoise regarding the effect of CPAP on clinical outcomes. These types of studies are not, however, a substitute for RCTs. By their design these nonrandomized studies are subject to bias. Moreover, presumptive treatment with ineffective therapy is no benefit to patients and is perhaps unethical.

Other data sets have other challenges. Over 150 CPAP devices have been cleared through the Food and Drug Administration (FDA) 510k process in which a device with a prior approval/clearance can serve as a predicate device. Many of these clearances relied on engineering performance or short studies with unvalidated sleep apnea parameters. It is notable that these same devices could be used for noninferiority studies for other sleep apnea technologies.

To identify ongoing clinical trials that might be germane to the subject of this Technology Assessment, we searched government data bases for other planned studies with a minimum duration of 12 months. There were no relevant studies listed in the National Institutes of Health (NIH) RePORTER, which lists NIH-funded studies. In ClinicalTrials.gov, there were six RCT listings. Three of these listed mortality or CV endpoints (NCT02080156, NCT04513483, and NCT03936751), but the trials were being conducted in very select patient populations (patients with coronary atherosclerosis treated with percutaneous procedures or bypass grafting, patients with paroxysmal atrial fibrillation, and patients with hypertensive acute cardiogenic pulmonary edema, respectively). One included only patients under age 65. The ClinicalTrials.gov record for the largest of these trials (NCT02080156) has not been updated since 2015. The other three RCTs (NCT02886156, NCT02553902, and NCT02420184) were small and had other types of primary endpoints: Montreal Cognitive Assessment score, AHI, and estimated glomerular filtration rate (eGFR). One trial (NCT02553902) had two active treatment arms. The results for NCT02420184 were published after the review deadline: CPAP in patients with chronic kidney disease (stage 3 or 4) did not improve eGFR.[186] The available information on the methodology—

including patient population, size, control group—and endpoints suggests that these studies may be unable to address outstanding questions about the management of Medicare beneficiaries with OSA. Final assessments, however, can be made only after the trials are completed and published in peer-reviewed journals.

Conclusions

The studies are highly inconsistent in how they define breathing measures (such as AHI and ODI) and, by extension, how they define OSA, despite statements of following standard sleep study criteria. Reporting of specific diagnosis and sleep study criteria is generally poor.

RCTs have not found significant effects of CPAP on long-term health outcomes for adults with OSA. RCTs of long-term outcomes do not provide evidence that CPAP has an effect on all-cause mortality, stroke, acute myocardial infarction, composite CV outcomes, accidents, incident diabetes (all with low SoE), and on clinically significant improvements in (unvalidated) measures of depression and anxiety symptoms, cognitive function, or quality of life (all with low SoE). The effect of CPAP on other long-term health outcomes from RCTs is unclear due to insufficient evidence related to sparse studies and/or highly imprecise estimates. Inclusion of information from adjusted NRCSs mostly does not alter these conclusions, except that the NRCSs support an association between CPAP use and lower risk of all-cause death with longer term followup. Underpowered studies have contributed to some of this evidentiary weakness. RCTs have mostly been powered for composite CV events, however the definitions of this outcome is highly variable across studies. Clinical equipoise remains. Future high quality, long-term studies that are adequately powered for clinical endpoints are needed. .

A small number of studies comparing CPAP with mandibular advancement devices provided low SoE that (unvalidated) measures of depression and anxiety symptoms are unaffected by either OSA treatment. There is insufficient or no evidence regarding the relative effect of CPAP versus mandibular advancement devices on other outcomes. There is no evidence regarding comparisons of CPAP with other (nonsurgical, nonpharmacologic) interventions. There is low strength of evidence that functional status remains similar between patients prescribed either fixed CPAP or autoCPAP. Other long-term health outcomes have not been reported.

Across studies, there is insufficient evidence to evaluate which, if any, patients have greater benefits with any given treatment.

No studies have evaluated whether breathing measures or sleepiness symptom measures may be valid surrogate or mediator measures for long-term health outcomes. Across studies, there was insufficient evidence to support whether changes in intermediate or surrogate measures are correlated with health outcomes.

Future well-conducted, well-reported studies are needed to allow more definitive conclusions regarding the clinical effect of CPAP for adults with OSA, to determine who might most benefit from long-term CPAP treatment, and to evaluate the validity of intermediate measures as potential surrogate or mediator measures of long-term health outcomes.

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, https://aasm.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. Centers for Medicare & Medicaid Services. Decision Memo for Continuous Positive Airway Pressure (CPAP) Therapy for Obstructive Sleep Apnea (OSA) (CAG-00093R2). https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=204. 2008.

9. Centers for Medicare & Medicaid Services. 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. 2001.

10. Giles TL, Lasserson TJ, Smith BH, et al. Continuous positive airways pressure for obstructive sleep apnoea in adults. Cochrane Database Syst Rev. 2006 Jul 19(3):Cd001106. doi: 10.1002/14651858.CD001106.pub3. PMID: 16855960.

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. Berkman ND, Lohr KN, Ansari M, et al. AHRQ Methods for Effective Health Care: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.

16. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009 Oct;62(10):1006-12. doi: 10.1016/j.jclinepi.2009.06.005. PMID: 19631508.

17. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017 Sep 21;358:j4008. doi: 10.1136/bmj.j4008. PMID: 28935701.

18. 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.

19. 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.

20. 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.

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. 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.

24. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919. PMID: 27733354.

25. Yao XI, Wang X, Speicher PJ, et al. Reporting and Guidelines in Propensity Score Analysis: A Systematic Review of Cancer and Cancer Surgical Studies. J Natl Cancer Inst. 2017 Aug 1;109(8). doi: 10.1093/jnci/djw323. PMID: 28376195.

26. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014 Apr 29;(4):Mr000034. doi: 10.1002/14651858.MR000034.pub2. PMID: 24782322.

27. Burzykowski T, Molenberghs G, Buyse M. The evaluation of surrogate endpoints: Springer Science & Business Media; 2006.

28. Spearman C. The proof and measurement of association between two things. Int J Epidemiol. 2010 Oct;39(5):1137-50. doi: 10.1093/ije/dyq191. PMID: 21051364.

29. Berkman ND, Lohr KN, Ansari MT, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. J Clin Epidemiol. 2015 Nov;68(11):1312-24. doi: 10.1016/j.jclinepi.2014.11.023. PMID: 25721570.

30. Punjabi NM. COUNTERPOINT: Is the Apnea-Hypopnea Index the Best Way to Quantify the Severity of Sleep-Disordered Breathing? No. Chest. 2016 Jan;149(1):16-9. doi: 10.1378/chest.14-2261. PMID: 26182150.

31. Rapoport DM. POINT: Is the Apnea-Hypopnea Index the Best Way to Quantify the Severity of Sleep-Disordered Breathing? Yes. Chest. 2016 Jan;149(1):14-6. doi: 10.1378/chest.15-1319. PMID: 26181884.

32. Pitson DJ, Stradling JR. Autonomic markers of arousal during sleep in patients undergoing investigation for obstructive sleep apnoea, their relationship to EEG arousals, respiratory events and subjective sleepiness. J Sleep Res. 1998 Mar;7(1):53-9. doi: 10.1046/j.1365-2869.1998.00092.x. PMID: 9613428.

33. Bennett LS, Langford BA, Stradling JR, et al. Sleep fragmentation indices as predictors of daytime sleepiness and nCPAP response in obstructive sleep apnea. Am J Respir Crit Care Med. 1998 Sep;158(3):778-86. doi: 10.1164/ajrccm.158.3.9711033. PMID: 9731004.

34. Amatoury J, Azarbarzin A, Younes M, et al. Arousal Intensity is a Distinct Pathophysiological Trait in Obstructive

Sleep Apnea. Sleep. 2016 Dec 1;39(12):2091-100. doi: 10.5665/sleep.6304. PMID: 27784404.

35. Azarbarzin A, Ostrowski M, Hanly P, et al. Relationship between arousal intensity and heart rate response to arousal. Sleep. 2014 Apr 1;37(4):645-53. doi: 10.5665/sleep.3560. PMID: 24899756.

36. Mazzotti DR, Lim DC, Sutherland K, et al. Opportunities for utilizing polysomnography signals to characterize obstructive sleep apnea subtypes and severity. Physiol Meas. 2018 Sep 13;39(9):09tr1. doi: 10.1088/1361-6579/aad5fe. PMID: 30047487.

37. Azarbarzin A, Sands SA, Stone KL, et al. The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the Osteoporotic Fractures in Men Study and the Sleep Heart Health Study. Eur Heart J. 2019 Apr 7;40(14):1149-57. doi: 10.1093/eurheartj/ehy624. PMID: 30376054.

38. Shahar E. Apnea-hypopnea index: time to wake up. Nat Sci Sleep. 2014;6:51-6. doi: 10.2147/nss.S61853. PMID: 24741336.

39. Wimms A, Woehrle H, Ketheeswaran S, et al. Obstructive Sleep Apnea in Women: Specific Issues and Interventions. Biomed Res Int. 2016;2016:1764837. doi: 10.1155/2016/1764837. PMID: 27699167.

40. O'Connor C, Thornley KS, Hanly PJ. Gender differences in the polysomnographic features of obstructive sleep apnea. Am J Respir Crit Care Med. 2000 May;161(5):1465-72. doi: 10.1164/ajrccm.161.5.9904121. PMID: 10806140.

41. Muraja-Murro A, Nurkkala J, Tiihonen P, et al. Total duration of apnea and hypopnea events and average desaturation show significant variation in patients with a similar apnea-hypopnea index. J Med Eng

Technol. 2012 Nov;36(8):393-8. doi: 10.3109/03091902.2012.712201. PMID: 22953737.

42. Kulkas A, Tiihonen P, Eskola K, et al. Novel parameters for evaluating severity of sleep disordered breathing and for supporting diagnosis of sleep apnea-hypopnea syndrome. J Med Eng Technol. 2013 Feb;37(2):135-43. doi: 10.3109/03091902.2012.754509. PMID: 23360196.

43. Tam S, Woodson BT, Rotenberg B. Outcome measurements in obstructive sleep apnea: beyond the apnea-hypopnea index. Laryngoscope. 2014 Jan;124(1):337-43. doi: 10.1002/lary.24275. PMID: 23794206.

44. Balakrishnan K, James KT, Weaver EM. Composite severity indices reflect sleep apnea disease burden more comprehensively than the apnea-hypopnea index. Otolaryngol Head Neck Surg. 2013 Feb;148(2):324-30. doi: 10.1177/0194599812464468. PMID: 23077154.

45. Piccirillo JF, Gates GA, White DL, et al. Obstructive sleep apnea treatment outcomes pilot study. Otolaryngol Head Neck Surg. 1998 Jun;118(6):833-44. doi: 10.1016/s0194-5998(98)70277-3. PMID: 9627245.

46. Manser RL, Rochford P, Naughton MT, et al. Measurement variability in sleep disorders medicine: the Victorian experience. Intern Med J. 2002 Aug;32(8):386-93. doi: 10.1046/j.1445-5994.2002.00256.x. PMID: 12162395.

47. Ho V, Crainiceanu CM, Punjabi NM, et al. Calibration Model for Apnea-Hypopnea Indices: Impact of Alternative Criteria for Hypopneas. Sleep. 2015 Dec 1;38(12):1887-92. doi: 10.5665/sleep.5234. PMID: 26564122.

48. Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology. 2008 May;108(5):812-21. doi: 10.1097/ALN.0b013e31816d83e4. PMID: 18431116.

49. Chung F, Abdullah HR, Liao P. STOP-Bang Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea. Chest. 2016 Mar;149(3):631-8. doi: 10.1378/chest.15-0903. PMID: 26378880.

50. Farney RJ, Walker BS, Farney RM, et al. The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index. J Clin Sleep Med. 2011 Oct 15;7(5):459-65b. doi: 10.5664/jcsm.1306. PMID: 22003340.

51. Silva GE, Vana KD, Goodwin JL, et al. Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. J Clin Sleep Med. 2011 Oct 15;7(5):467-72. doi: 10.5664/jcsm.1308. PMID: 22003341.

52. Nicholl DD, Ahmed SB, Loewen AH, et al. Diagnostic value of screening instruments for identifying obstructive sleep apnea in kidney failure. J Clin Sleep Med. 2013 Jan 15;9(1):31-8. doi: 10.5664/jcsm.2334. PMID: 23319902.

53. Ramachandran SK, Josephs LA. A meta-analysis of clinical screening tests for obstructive sleep apnea. Anesthesiology. 2009 Apr;110(4):928-39. doi: 10.1097/ALN.0b013e31819c47b6. PMID: 19293704.

54. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991 Dec;14(6):540-5. doi: 10.1093/sleep/14.6.540. PMID: 1798888.

55. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. Chest. 1993

Jan;103(1):30-6. doi: 10.1378/chest.103.1.30. PMID: 8417909.

56. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep. 1992 Aug;15(4):376-81. doi: 10.1093/sleep/15.4.376. PMID: 1519015.

57. Kendzerska TB, Smith PM, Brignardello-Petersen R, et al. Evaluation of the measurement properties of the Epworth sleepiness scale: a systematic review. Sleep Med Rev. 2014 Aug;18(4):321-31. doi: 10.1016/j.smrv.2013.08.002. PMID: 24135493.

58. Pilcher JJ, Switzer FS, 3rd, Munc A, et al. Psychometric properties of the Epworth Sleepiness Scale: A factor analysis and item-response theory approach. Chronobiol Int. 2018 Apr;35(4):533-45. doi: 10.1080/07420528.2017.1420075. PMID: 29341789.

59. Walker NA, Sunderram J, Zhang P, et al. Clinical utility of the Epworth sleepiness scale. Sleep Breath. 2020 Jan 14. doi: 10.1007/s11325-020-02015-2. PMID: 31938991.

60. Jonas DE, Amick HR, Feltner C, et al. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Screening for Obstructive Sleep Apnea in Adults: An Evidence Review for the US Preventive Services Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2017.

61. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun;30(6):473-83. PMID: 1593914.

62. Flemons WW, Reimer MA. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. Am J Respir Crit Care Med. 1998 Aug;158(2):494-503. doi: 10.1164/ajrccm.158.2.9712036. PMID: 9700127.

63. Weaver TE, Laizner AM, Evans LK, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep. 1997 Oct;20(10):835-43. PMID: 9415942.

64. Silva GE, Goodwin JL, Vana KD, et al. Obstructive Sleep Apnea and Quality of Life: Comparison of the SAQLI, FOSQ, and SF-36 Questionnaires. Southwest J Pulm Crit Care. 2016;13(3):137-49. doi: 10.13175/swjpcc082-16. PMID: 27738560.

65. Feeny DH, Eckstrom E, Whitlock EP, et al. AHRQ Methods for Effective Health Care. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.

66. U.S. Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. https://www.fda.gov/media/77832/download . 2009.

67. Patrick DL, Erickson P. Health status and health policy: Quality of life in health care evaluation and resource allocation. New York: Oxford University Press; 1993.

68. Chasens ER, Ratcliffe SJ, Weaver TE. Development of the FOSQ-10: a short version of the Functional Outcomes of Sleep Questionnaire. Sleep. 2009 Jul;32(7):915-9. doi: 10.1093/sleep/32.7.915. PMID: 19639754.

69. Billings ME, Rosen CL, Auckley D, et al. Psychometric performance and responsiveness of the functional outcomes of sleep questionnaire and sleep apnea quality of life instrument in a randomized

trial: the HomePAP study. Sleep. 2014 Dec 1;37(12):2017-24. doi: 10.5665/sleep.4262. PMID: 25325491.

70. Hoffstein V. Review of oral appliances for treatment of sleep-disordered breathing. Sleep Breath. 2007 Mar;11(1):1-22. doi: 10.1007/s11325-006-0084-8. PMID: 17136406.

71. Kushida CA, Morgenthaler TI, Littner MR, et al. Practice parameters for the treatment of snoring and Obstructive Sleep Apnea with oral appliances: an update for 2005. Sleep. 2006 Feb;29(2):240-3. doi: 10.1093/sleep/29.2.240. PMID: 16494092.

72. Elshaug AG, Moss JR, Southcott AM, et al. Redefining success in airway surgery for obstructive sleep apnea: a meta analysis and synthesis of the evidence. Sleep. 2007 Apr;30(4):461-7. doi: 10.1093/sleep/30.4.461. PMID: 17520790.

73. Fujita S, Conway W, Zorick F, et al. Surgical correction of anatomic abnormalities in obstructive sleep apnea syndrome: uvulopalatopharyngoplasty. Otolaryngology--Head and Neck Surgery. 1981;89(6):923-34.

74. Prinsell JR. Maxillomandibular advancement surgery for obstructive sleep apnea syndrome. J Am Dent Assoc. 2002 Nov;133(11):1489-97; quiz 539-40. doi: 10.14219/jada.archive.2002.0079. PMID: 12462692.

75. Berry RB, Kryger MH, Massie CA. A novel nasal expiratory positive airway pressure (EPAP) device for the treatment of obstructive sleep apnea: a randomized controlled trial. Sleep. 2011 Apr 1;34(4):479-85. doi: 10.1093/sleep/34.4.479. PMID: 21461326.

76. Colrain IM, Black J, Siegel LC, et al. A multicenter evaluation of oral pressure therapy for the treatment of obstructive sleep apnea. Sleep Med. 2013 Sep;14(9):830-7.

doi: 10.1016/j.sleep.2013.05.009. PMID: 23871259.

77. Gupta RJ, Kademani D, Liu SY. Upper Airway (Hypoglossal Nerve) Stimulation for Treatment of Obstructive Sleep Apnea. Atlas Oral Maxillofac Surg Clin North Am. 2019 Mar;27(1):53-8. doi: 10.1016/j.cxom.2018.11.004. PMID: 30717924.

78. Strollo PJ, Jr., Soose RJ, Maurer JT, et al. Upper-airway stimulation for obstructive sleep apnea. N Engl J Med. 2014 Jan 9;370(2):139-49. doi: 10.1056/NEJMoa1308659. PMID: 24401051.

79. Mahadevia AK, Onal E, Lopata M. Effects of expiratory positive airway pressure on sleep-induced respiratory abnormalities in patients with hypersomnia-sleep apnea syndrome. Am Rev Respir Dis. 1983 Oct;128(4):708-11. doi: 10.1164/arrd.1983.128.4.708. PMID: 6354028.

80. Dedhia RC, Strollo PJ, Soose RJ. Upper Airway Stimulation for Obstructive Sleep Apnea: Past, Present, and Future. Sleep. 2015 Jun 1;38(6):899-906. doi: 10.5665/sleep.4736. PMID: 25409109.

81. U.S. Food and Drug Administration. 510(k) Clearance. https://www.fda.gov/medical-devices/device-approvals-denials-and-clearances/510k-clearances. 2018. Accessed Nov.27, 2020. 2018.

82. U.S. Food and Drug Administration. Devices@FDA. https://www.accessdata.fda.gov/scripts/cdrh/devicesatfda/. Accessed Mar. 1, 2020.

83. Johnson KG, Johnson DC. Treatment of sleep-disordered breathing with positive airway pressure devices: technology update. Med Devices (Auckl). 2015;8:425-37. doi: 10.2147/mder.S70062. PMID: 26604837.

84. Xie C, Zhu R, Tian Y, et al. Association of obstructive sleep apnoea with the risk of vascular outcomes and all-cause mortality: a meta-analysis. BMJ Open. 2017 Dec 22;7(12):e013983. doi: 10.1136/bmjopen-2016-013983. PMID: 29275335.

85. Zhang X, Fan J, Guo Y, et al. Association between obstructive sleep apnoea syndrome and the risk of cardiovascular diseases: an updated systematic review and dose-response meta-analysis. Sleep Med. 2020 Jul;71:39-46. doi: 10.1016/j.sleep.2020.03.011. PMID: 32485597.

86. Xia W, Huang Y, Peng B, et al. Relationship between obstructive sleep apnoea syndrome and essential hypertension: a dose-response meta-analysis. Sleep Med. 2018 Jul;47:11-8. doi: 10.1016/j.sleep.2018.03.016. PMID: 29880142.

87. Bubu OM, Andrade AG, Umasabor-Bubu OQ, et al. Obstructive sleep apnea, cognition and Alzheimer's disease: A systematic review integrating three decades of multidisciplinary research. Sleep Med Rev. 2020 Apr;50:101250. doi: 10.1016/j.smrv.2019.101250. PMID: 31881487.

88. Khalil M, Power N, Graham E, et al. The association between sleep and diabetes outcomes - A systematic review. Diabetes Res Clin Pract. 2020 Mar;161:108035. doi: 10.1016/j.diabres.2020.108035. PMID: 32006640.

89. Garbarino S, Bardwell WA, Guglielmi O, et al. Association of Anxiety and Depression in Obstructive Sleep Apnea Patients: A Systematic Review and Meta-Analysis. Behav Sleep Med. 2020 Jan-Feb;18(1):35-57. doi: 10.1080/15402002.2018.1545649. PMID: 30453780.

90. Pauletto P, Réus JC, Bolan M, et al. Association between obstructive sleep apnea and health-related quality of life in untreated adults: a systematic review. Sleep Breath. 2021 Mar 11. doi: 10.1007/s11325-021-02323-1. PMID: 33709191.

91. Liu L, Kang R, Zhao S, et al. Sexual Dysfunction in Patients with Obstructive Sleep Apnea: A Systematic Review and Meta-Analysis. J Sex Med. 2015 Oct;12(10):1992-2003. doi: 10.1111/jsm.12983. PMID: 26395783.

92. Kellesarian SV, Malignaggi VR, Feng C, et al. Association between obstructive sleep apnea and erectile dysfunction: a systematic review and meta-analysis. Int J Impot Res. 2018 Jun;30(3):129-40. doi: 10.1038/s41443-018-0017-7. PMID: 29795528.

93. Aarab G, Nikolopoulou M, Ahlberg J, et al. Oral appliance therapy versus nasal continuous positive airway pressure in obstructive sleep apnea: a randomized, placebo-controlled trial on psychological distress. Clin Oral Investig. 2017 Sep;21(7):2371-8. doi: 10.1007/s00784-016-2045-3. PMID: 28083705.

94. Nikolopoulou M, Byraki A, Ahlberg J, et al. Oral appliance therapy versus nasal continuous positive airway pressure in obstructive sleep apnoea syndrome: a randomised, placebo-controlled trial on self-reported symptoms of common sleep disorders and sleep-related problems. J Oral Rehabil. 2017 Jun;44(6):452-60. doi: 10.1111/joor.12505. PMID: 28294380.

95. Batool-Anwar S, Omobomi O, Quan SF. The effect of CPAP on HRQOL as measured by the Quality of Well-Being Self Administered Questionaire (QWB-SA). Southwest journal of pulmonary & critical care. 2020;20(1):29. PMID: 32190413.

96. Batool-Anwar S, Goodwin JL, Kushida CA, et al. Impact of continuous positive airway pressure (CPAP) on quality of life in patients with obstructive sleep apnea (OSA). Journal of sleep research. 2016;25(6):731-8. PMID: 27242272.

97. Kushida CA, Nichols DA, Holmes TH, et al. Effects of continuous positive airway pressure on neurocognitive function in obstructive sleep apnea patients: The Apnea Positive Pressure Long-term Efficacy Study (APPLES). Sleep. 2012 Dec 1;35(12):1593-602. doi: 10.5665/sleep.2226. PMID: 23204602.

98. Kushida CA, Nichols DA, Quan SF, et al. The Apnea Positive Pressure Long-term Efficacy Study (APPLES): rationale, design, methods, and procedures. J Clin Sleep Med. 2006 Jul 15;2(3):288-300. PMID: 17561541.

99. Barbé F, Durán-Cantolla J, Sánchez-de-la-Torre M, et al. Effect of continuous positive airway pressure on the incidence of hypertension and cardiovascular events in nonsleepy patients with obstructive sleep apnea: a randomized controlled trial. Jama. 2012 May 23;307(20):2161-8. doi: 10.1001/jama.2012.4366. PMID: 22618923.

100. Reid ML, Gleason KJ, Bakker JP, et al. The role of sham continuous positive airway pressure as a placebo in controlled trials: Best Apnea Interventions for Research Trial. Sleep. 2019 Aug 1;42(8). doi: 10.1093/sleep/zsz099. PMID: 31116848.

101. Zhao YY, Wang R, Gleason KJ, et al. Effect of Continuous Positive Airway Pressure Treatment on Health-Related Quality of Life and Sleepiness in High Cardiovascular Risk Individuals With Sleep Apnea: Best Apnea Interventions for Research (BestAIR) Trial. Sleep. 2017 Apr 1;40(4). doi: 10.1093/sleep/zsx040. PMID: 28419387.

102. Huang Z, Liu Z, Luo Q, et al. Long-term effects of continuous positive airway pressure on blood pressure and prognosis in hypertensive patients with coronary heart disease and obstructive sleep apnea: a randomized controlled trial. Am J Hypertens. 2015 Mar;28(3):300-6. doi: 10.1093/ajh/hpu147. PMID: 25125635.

103. Sánchez-de-la-Torre M, Sánchez-de-la-Torre A, Bertran S, et al. Effect of obstructive sleep apnoea and its treatment with continuous positive airway pressure on the prevalence of cardiovascular events in patients with acute coronary syndrome (ISAACC study): a randomised controlled trial. Lancet Respir Med. 2020 Apr;8(4):359-67. doi: 10.1016/s2213-2600(19)30271-1. PMID: 31839558.

104. Monasterio C, Vidal S, Duran J, et al. Effectiveness of continuous positive airway pressure in mild sleep apnea-hypopnea syndrome. Am J Respir Crit Care Med. 2001 Sep 15;164(6):939-43. doi: 10.1164/ajrccm.164.6.2008010. PMID: 11587974.

105. Craig SE, Kohler M, Nicoll D, et al. Continuous positive airway pressure improves sleepiness but not calculated vascular risk in patients with minimally symptomatic obstructive sleep apnoea: the MOSAIC randomised controlled trial. Thorax. 2012 Dec;67(12):1090-6. doi: 10.1136/thoraxjnl-2012-202178. PMID: 23111478.

106. Tang I, Turnbull CD, Sen D, et al. Effect of CPAP on cardiovascular events in minimally symptomatic OSA: long-term follow-up of the MOSAIC randomised controlled trial. BMJ Open Respir Res. 2020 Sep;7(1). doi: 10.1136/bmjresp-2020-000742. PMID: 32928788.

107. Turnbull CD, Craig SE, Kohler M, et al. Cardiovascular event rates in the MOSAIC trial: 2-year follow-up data. Thorax. 2014

Oct;69(10):950. doi: 10.1136/thoraxjnl-2013-204826. PMID: 24508706.

108. Pelletier-Fleury N, Meslier N, Gagnadoux F, et al. Economic arguments for the immediate management of moderate-to-severe obstructive sleep apnoea syndrome. European Respiratory Journal. 2004;23(1):53-60. PMID: 14738231.

109. McMillan A, Bratton DJ, Faria R, et al. Continuous positive airway pressure in older people with obstructive sleep apnoea syndrome (PREDICT): a 12-month, multicentre, randomised trial. Lancet Respir Med. 2014 Oct;2(10):804-12. doi: 10.1016/s2213-2600(14)70172-9. PMID: 25172769.

110. Balcan B, Thunström E, Strollo PJ, Jr., et al. Continuous Positive Airway Pressure Treatment and Depression in Adults with Coronary Artery Disease and Nonsleepy Obstructive Sleep Apnea. A Secondary Analysis of the RICCADSA Trial. Ann Am Thorac Soc. 2019 Jan;16(1):62-70. doi: 10.1513/AnnalsATS.201803-174OC. PMID: 30130421.

111. Celik Y, Thunström E, Strollo Jr PJ, et al. Continuous positive airway pressure treatment and anxiety in adults with coronary artery disease and nonsleepy obstructive sleep apnea in the RICCADSA trial. Sleep Medicine. 2021;77:96-103. PMID: 33341644.

112. Peker Y, Glantz H, Eulenburg C, et al. Effect of Positive Airway Pressure on Cardiovascular Outcomes in Coronary Artery Disease Patients with Nonsleepy Obstructive Sleep Apnea. The RICCADSA Randomized Controlled Trial. Am J Respir Crit Care Med. 2016 Sep 1;194(5):613-20. doi: 10.1164/rccm.201601-0088OC. PMID: 26914592.

113. Wallström S, Balcan B, Thunström E, et al. CPAP and Health-Related Quality of Life in Adults With Coronary Artery Disease and Nonsleepy Obstructive Sleep Apnea in the RICCADSA Trial. J Clin Sleep Med. 2019 Sep 15;15(9):1311-20. doi: 10.5664/jcsm.7926. PMID: 31538602.

114. Antic NA, Heeley E, Anderson CS, et al. The Sleep Apnea cardioVascular Endpoints (SAVE) Trial: Rationale, Ethics, Design, and Progress. Sleep. 2015 Aug 1;38(8):1247-57. doi: 10.5665/sleep.4902. PMID: 25669180.

115. Heeley E, Billot L, Anderson CS, et al. Statistical analysis plan for the Sleep Apnea cardioVascular Endpoints study: An international randomised controlled trial to determine whether continuous positive airways pressure treatment for obstructive sleep apnea in patients with CV disease prevents secondary cardiovascular events. Int J Stroke. 2016 Jan;11(1):148-50. doi: 10.1177/1747493015607504. PMID: 26763030.

116. Quan W, Zheng D, Douglas McEvoy R, et al. High Risk Characteristics for Recurrent Cardiovascular Events among Patients with Obstructive Sleep Apnoea in the SAVE Study. EClinicalMedicine. 2018 Aug-Sep;2-3:59-65. doi: 10.1016/j.eclinm.2018.09.002. PMID: 31193555.

117. Zheng D, Xu Y, You S, et al. Effects of continuous positive airway pressure on depression and anxiety symptoms in patients with obstructive sleep apnoea: results from the sleep apnoea cardiovascular Endpoint randomised trial and meta-analysis. EClinicalMedicine. 2019 May-Jun;11:89-96. doi: 10.1016/j.eclinm.2019.05.012. PMID: 31312807.

118. Shaw JE, Punjabi NM, Naughton MT, et al. The Effect of Treatment of Obstructive Sleep Apnea on Glycemic Control in Type 2 Diabetes. Am J Respir Crit Care Med. 2016 Aug 15;194(4):486-92. doi:

10.1164/rccm.201511-2260OC. PMID: 26926656.

119. Wu SQ, Liao QC, Xu XX, et al. Effect of CPAP therapy on C-reactive protein and cognitive impairment in patients with obstructive sleep apnea hypopnea syndrome. Sleep Breath. 2016 Dec;20(4):1185-92. doi: 10.1007/s11325-016-1331-2. PMID: 26993342.

120. Bjornsdottir E, Keenan BT, Eysteinsdottir B, et al. Quality of life among untreated sleep apnea patients compared with the general population and changes after treatment with positive airway pressure. J Sleep Res. 2015 Jun;24(3):328-38. doi: 10.1111/jsr.12262. PMID: 25431105.

121. Botros N, Concato J, Mohsenin V, et al. Obstructive sleep apnea as a risk factor for type 2 diabetes. Am J Med. 2009 Dec;122(12):1122-7. doi: 10.1016/j.amjmed.2009.04.026. PMID: 19958890.

122. Budweiser S, Luigart R, Jörres RA, et al. Long-term changes of sexual function in men with obstructive sleep apnea after initiation of continuous positive airway pressure. J Sex Med. 2013 Feb;10(2):524-31. doi: 10.1111/j.1743-6109.2012.02968.x. PMID: 23088487.

123. Campos-Rodriguez F, Pena-Grinan N, Reyes-Nunez N, et al. Mortality in obstructive sleep apnea-hypopnea patients treated with positive airway pressure. Chest. 2005;128(2):624-33. PMID: 16100147.

124. Crawford-Achour E, Dauphinot V, Martin MS, et al. Protective Effect of Long-Term CPAP Therapy on Cognitive Performance in Elderly Patients with Severe OSA: The PROOF Study. J Clin Sleep Med. 2015 Apr 15;11(5):519-24. doi: 10.5664/jcsm.4694. PMID: 25700873.

125. Jennum P, Tønnesen P, Ibsen R, et al. All-cause mortality from obstructive sleep apnea in male and female patients with and without continuous positive airway pressure treatment: a registry study with 10 years of follow-up. Nat Sci Sleep. 2015;7:43-50. doi: 10.2147/nss.S75166. PMID: 25914563.

126. Lisan Q, Van Sloten T, Marques Vidal P, et al. Association of Positive Airway Pressure Prescription With Mortality in Patients With Obesity and Severe Obstructive Sleep Apnea: The Sleep Heart Health Study. JAMA Otolaryngol Head Neck Surg. 2019 Jun 1;145(6):509-15. doi: 10.1001/jamaoto.2019.0281. PMID: 30973594.

127. López-Padilla D, Alonso-Moralejo R, Martínez-García M, et al. Continuous positive airway pressure and survival of very elderly persons with moderate to severe obstructive sleep apnea. Sleep Med. 2016 Mar;19:23-9. doi: 10.1016/j.sleep.2015.10.015. PMID: 27198943.

128. Myllylä M, Hammais A, Stepanov M, et al. Nonfatal and fatal cardiovascular disease events in CPAP compliant obstructive sleep apnea patients. Sleep Breath. 2019 Dec;23(4):1209-17. doi: 10.1007/s11325-019-01808-4. PMID: 30848437.

129. Navarro-Soriano C, Martínez-García MA, Torres G, et al. Long-term Effect of CPAP Treatment on Cardiovascular Events in Patients With Resistant Hypertension and Sleep Apnea. Data From the HIPARCO-2 Study. Arch Bronconeumol. 2021 Mar;57(3):165-71. doi: 10.1016/j.arbres.2019.12.006. PMID: 32029279.

130. Nakamura H, Kanemura T, Takara C, et al. A retrospective analysis of 4000 patients with obstructive sleep apnea in Okinawa, Japan. Sleep and Biological Rhythms. 2009;7(2).

131. Nakamura K, Nakamura H, Tohyama K, et al. Survival benefit of continuous positive airway pressure in Japanese patients with obstructive sleep apnea: a propensity-score matching analysis. J Clin Sleep Med. 2021 Feb 1;17(2):211-8. doi: 10.5664/jcsm.8842. PMID: 33006309.

132. Ou Q, Chen YC, Zhuo SQ, et al. Continuous Positive Airway Pressure Treatment Reduces Mortality in Elderly Patients with Moderate to Severe Obstructive Severe Sleep Apnea: A Cohort Study. PLoS One. 2015;10(6):e0127775. doi: 10.1371/journal.pone.0127775. PMID: 26068440.

133. Schipper MH, Jellema K, Thomassen BJW, et al. Stroke and other cardiovascular events in patients with obstructive sleep apnea and the effect of continuous positive airway pressure. J Neurol. 2017 Jun;264(6):1247-53. doi: 10.1007/s00415-017-8522-z. PMID: 28550476.

134. Sheth U, Monson RS, Prasad B, et al. Association of continuous positive airway pressure compliance with complications in patients with type 2 diabetes and obstructive sleep apnea. J Clin Sleep Med. 2021 Mar 17. doi: 10.5664/jcsm.9248. PMID: 33729910.

135. Wu X, Lv S, Yu X, et al. Treatment of OSA reduces the risk of repeat revascularization after percutaneous coronary intervention. Chest. 2015 Mar;147(3):708-18. doi: 10.1378/chest.14-1634. PMID: 25412159.

136. Chang ET, Chen W, Wang LY, et al. Continuous positive pressure therapy usage and incident stroke in patients with obstructive sleep apnea: A nationwide population-based cohort study. Clin Respir J. 2020 Sep;14(9):822-8. doi: 10.1111/crj.13214. PMID: 32421898.

137. de Batlle J, Bertran S, Turino C, et al. Continuous Positive Airway Pressure (CPAP) Treatment Reduces Mortality at the Population Level. American Journal of Respiratory and Critical Care Medicine. 2018;197:A4399.

138. Jara SM, Phipps AI, Maynard C, et al. Long-term survival in veterans with sleep apnea. SLEEP. 2018;41.

139. Lemay KR, Tulloch HE, Pipe AL, et al. Establishing the Minimal Clinically Important Difference for the Hospital Anxiety and Depression Scale in Patients With Cardiovascular Disease. J Cardiopulm Rehabil Prev. 2019 Nov;39(6):E6-e11. doi: 10.1097/hcr.0000000000000379. PMID: 30489438.

140. Folstein MF, Robins LN, Helzer JE. The Mini-Mental State Examination. Arch Gen Psychiatry. 1983 Jul;40(7):812. doi: 10.1001/archpsyc.1983.01790060110016. PMID: 6860082.

141. Andrews JS, Desai U, Kirson NY, et al. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimers Dement (N Y). 2019;5:354-63. doi: 10.1016/j.trci.2019.06.005. PMID: 31417957.

142. Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res. 2009 Jan;43(4):411-31. doi: 10.1016/j.jpsychires.2008.04.014. PMID: 18579155.

143. Ricker JH, Axelrod BN, Houtler BD. Clinical validation of the oral trail making test. Neuropsychiatry, Neuropsychology, & Behavioral Neurology. 1996.

144. Fendler TJ, Spertus JA, Gosch KL, et al. Incidence and predictors of cognitive decline in patients with left ventricular assist devices. Circ Cardiovasc Qual Outcomes.

2015 May;8(3):285-91. doi: 10.1161/circoutcomes.115.001856. PMID: 25925372.

145. Shah CH, Brown JD. Reliability and Validity of the Short-Form 12 Item Version 2 (SF-12v2) Health-Related Quality of Life Survey and Disutilities Associated with Relevant Conditions in the U.S. Older Adult Population. J Clin Med. 2020 Feb 29;9(3). doi: 10.3390/jcm9030661. PMID: 32121371.

146. Wyrwich KW, Tierney WM, Babu AN, et al. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. Health Serv Res. 2005 Apr;40(2):577-91. doi: 10.1111/j.1475-6773.2005.00373.x. PMID: 15762908.

147. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011 Dec;20(10):1727-36. doi: 10.1007/s11136-011-9903-x. PMID: 21479777.

148. Lacasse Y, Godbout C, Sériès F. Independent validation of the Sleep Apnoea Quality of Life Index. Thorax. 2002 Jun;57(6):483-8. doi: 10.1136/thorax.57.6.483. PMID: 12037222.

149. Weaver EM, Crosby RD, Bron M, et al. Using Multiple Anchor-based And Distribution-based Estimates To Determine The Minimal Important Difference (MID) For The FOSQ-10. Sleep. 2018;41:A227.

150. Ferrer M, Vilagut G, Monasterio C, et al. [Measurement of the perceived impact of sleep problems: the Spanish version of the functional outcomes sleep questionnaire and the Epworth sleepiness scale]. Med Clin (Barc). 1999 Sep 11;113(7):250-5. PMID: 10544379.

151. Jara SM, Hopp ML, Weaver EM. Association of Continuous Positive Airway Pressure Treatment With Sexual Quality of Life in Patients With Sleep Apnea: Follow-up Study of a Randomized Clinical Trial. JAMA Otolaryngol Head Neck Surg. 2018 Jul 1;144(7):587-93. doi: 10.1001/jamaoto.2018.0485. PMID: 29800001.

152. Sweere Y, Kerkhof GA, De Weerd AW, et al. The validity of the Dutch Sleep Disorders Questionnaire (SDQ). J Psychosom Res. 1998 Dec;45(6):549-55. doi: 10.1016/s0022-3999(98)00030-0. PMID: 9859856.

153. Rosen RC, Riley A, Wagner G, et al. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. Urology. 1997 Jun;49(6):822-30. doi: 10.1016/s0090-4295(97)00238-0. PMID: 9187685.

154. de Vries GE, Hoekema A, Vermeulen KM, et al. Clinical- and Cost-Effectiveness of a Mandibular Advancement Device Versus Continuous Positive Airway Pressure in Moderate Obstructive Sleep Apnea. J Clin Sleep Med. 2019 Oct 15;15(10):1477-85. doi: 10.5664/jcsm.7980. PMID: 31596213.

155. Berg LM, Ankjell TKS, Sun YQ, et al. Health-Related Quality of Life and Sleep Quality after 12 Months of Treatment in Nonsevere Obstructive Sleep Apnea: A Randomized Clinical Trial with Continuous Positive Airway Pressure and Mandibular Advancement Splints. Int J Otolaryngol. 2020;2020:2856460. doi: 10.1155/2020/2856460. PMID: 32665778.

156. Uniken Venema JA, Doff MH, Joffe-Sokolova D, et al. Long-term obstructive sleep apnea therapy: a 10-year follow-up of mandibular advancement device and continuous positive airway pressure. Journal of Clinical Sleep Medicine. 2020;16(3):353-9. PMID: 31992403

157. Ware JE, Jr., Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998 Nov;51(11):945-52. doi: 10.1016/s0895-4356(98)00085-7. PMID: 9817111.

158. Russell A, Ball J, Spallek M. ALSWH Data Dictionary Supplement Section 2 Core Survey Dataset 2.3 Health-related Quality of Life Variables SF-36 – Standardised Component Scores. http://www.alswh.org.au/InfoData/DataDictionarySupplement/DDSSection2SF36.pdf/. 2009.

159. Doff MH, Hoekema A, Wijkstra PJ, et al. Oral appliance versus continuous positive airway pressure in obstructive sleep apnea syndrome: a 2-year follow-up. Sleep. 2013 Sep 1;36(9):1289-96. doi: 10.5665/sleep.2948. PMID: 23997361.

160. Meurice JC, Cornette A, Philip-Joet F, et al. Evaluation of autoCPAP devices in home treatment of sleep apnea/hypopnea syndrome. Sleep Med. 2007 Nov;8(7-8):695-703. doi: 10.1016/j.sleep.2007.03.019. PMID: 17638595.

161. Bloch KE, Huber F, Furian M, et al. Autoadjusted versus fixed CPAP for obstructive sleep apnoea: a multicentre, randomised equivalence trial. Thorax. 2018 Feb;73(2):174-84. doi: 10.1136/thoraxjnl-2016-209699. PMID: 28982804.

162. Kushida CA, Berry RB, Blau A, et al. Positive airway pressure initiation: a randomized controlled trial to assess the impact of therapy mode and titration process on efficacy, adherence, and outcomes. Sleep. 2011 Aug 1;34(8):1083-92. doi: 10.5665/sleep.1166. PMID: 21804670.

163. Doff MH, Finnema KJ, Hoekema A, et al. Long-term oral appliance therapy in obstructive sleep apnea syndrome: a controlled study on dental side effects. Clin Oral Investig. 2013 Mar;17(2):475-82. doi: 10.1007/s00784-012-0737-x. PMID: 22562077.

164. Doff MH, Veldhuis SK, Hoekema A, et al. Long-term oral appliance therapy in obstructive sleep apnea syndrome: a controlled study on temporomandibular side effects. Clin Oral Investig. 2012 Jun;16(3):689-97. doi: 10.1007/s00784-011-0555-6. PMID: 21538074.

165. Corral J, Sánchez-Quiroga M, Carmona-Bernal C, et al. Conventional Polysomnography Is Not Necessary for the Management of Most Patients with Suspected Obstructive Sleep Apnea. Noninferiority, Randomized Controlled Trial. Am J Respir Crit Care Med. 2017 Nov 1;196(9):1181-90. doi: 10.1164/rccm.201612-2497OC. PMID: 28636405.

166. Gagnadoux F, Nguyen XL, Le Vaillant M, et al. Comparison of titrable thermoplastic versus custom-made mandibular advancement device for the treatment of obstructive sleep apnoea. Respir Med. 2017 Oct;131:35-42. doi: 10.1016/j.rmed.2017.08.004. PMID: 28947040.

167. de Ruiter MHT, Benoist LBL, de Vries N, et al. Durability of treatment effects of the Sleep Position Trainer versus oral appliance therapy in positional OSA: 12-month follow-up of a randomized controlled trial. Sleep Breath. 2018 May;22(2):441-50. doi: 10.1007/s11325-017-1568-4. PMID: 28913630.

168. Fernández-Julián E, Pérez-Carbonell T, Marco R, et al. Impact of an oral appliance on obstructive sleep apnea severity, quality of life, and biomarkers. Laryngoscope. 2018 Jul;128(7):1720-6. doi: 10.1002/lary.26913. PMID: 29152745.

169. Patel S, Kon SSC, Nolan CM, et al. The Epworth Sleepiness Scale: Minimum Clinically Important Difference in Obstructive Sleep Apnea. Am J Respir Crit Care Med. 2018 Apr 1;197(7):961-3. doi: 10.1164/rccm.201704-0672LE. PMID: 28961021.

170. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010 Dec;15(4):309-34. doi: 10.1037/a0020761. PMID: 20954780.

171. VanderWeele T. Explanation in causal inference: methods for mediation and interaction: Oxford University Press; 2015.

172. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986 Dec;51(6):1173-82. doi: 10.1037//0022-3514.51.6.1173. PMID: 3806354.

173. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol. 2007;58:593-614. doi: 10.1146/annurev.psych.58.110405.085542. PMID: 16968208.

174. Imai K, Jo B, Stuart EA. Using Potential Outcomes to Understand Causal Mediation Analysis: Comment on. Multivariate Behav Res. 2011 Sep;46(5):861-73. doi: 10.1080/00273171.2011.606743. PMID: 23788819.

175. 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.

176. 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.

177. Pépin JL, Bailly S, Rinder P, et al. CPAP Therapy Termination Rates by OSA Phenotype: A French Nationwide Database Analysis. J Clin Med. 2021 Mar 1;10(5). doi: 10.3390/jcm10050936. PMID: 33804319.

178. Sawyer AM, Gooneratne NS, Marcus CL, et al. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. Sleep Med Rev. 2011 Dec;15(6):343-56. doi: 10.1016/j.smrv.2011.01.003. PMID: 21652236.

179. Patil SP, Ayappa IA, Caples SM, et al. Treatment of Adult Obstructive Sleep Apnea With Positive Airway Pressure: An American Academy of Sleep Medicine Systematic Review, Meta-Analysis, and GRADE Assessment. J Clin Sleep Med. 2019 Feb 15;15(2):301-34. doi: 10.5664/jcsm.7638. PMID: 30736888.

180. Patil SP, Ayappa IA, Caples SM, et al. Treatment of Adult Obstructive Sleep Apnea with Positive Airway Pressure: An American Academy of Sleep Medicine Clinical Practice Guideline. J Clin Sleep Med. 2019 Feb 15;15(2):335-43. doi: 10.5664/jcsm.7640. PMID: 30736887.

181. Beattie WS, Lalu M, Bocock M, et al. Systematic review and consensus definitions for the Standardized Endpoints in Perioperative Medicine (StEP) initiative: cardiovascular outcomes. Br J Anaesth. 2021 Jan;126(1):56-66. doi: 10.1016/j.bja.2020.09.023. PMID: 33092804.

182. Devereaux PJ, Yang H, Yusuf S, et al. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. Lancet. 2008 May 31;371(9627):1839-47. doi: 10.1016/s0140-6736(08)60601-7. PMID: 18479744.

183. Committee for Medicinal Products for Human Use (CHMP). European Medicines Agency. Reflection paper on assessment of cardiovascular safety profile of medicinal products. 2016 https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-assessment-cardiovascular-safety-profile-medicinal-products_en.pdf.

184. Somers VK, White DP, Amin R, et al. Sleep apnea and cardiovascular disease: an American Heart Association/American College of Cardiology Foundation Scientific Statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council on Cardiovascular Nursing. J Am Coll Cardiol. 2008 Aug 19;52(8):686-717. doi: 10.1016/j.jacc.2008.05.002. PMID: 18702977.

185. McEvoy RD, Anderson CS, Antic NA, et al. The sleep apnea cardiovascular endpoints (SAVE) trial: Rationale and start-up phase. J Thorac Dis. 2010 Sep;2(3):138-43. doi: 10.3978/j.issn.2072-1439.2010.02.03.5. PMID: 22263035.

186. Rimke AN, Ahmed SB, Turin TC, et al. Effect of CPAP Therapy on Kidney Function in Patients With Chronic Kidney Disease: A Pilot Randomized Controlled Trial. Chest. 2021 May;159(5):2008-19. doi: 10.1016/j.chest.2020.11.052. PMID: 33316238.

Abbreviations and Acronyms

AASM	American Academy of Sleep Medicine
AFib	atrial fibrillation
AHI	apnea-hypopnea index (apnea and hypopnea events per hour)
AHRQ	Agency for Healthcare Research and Quality
AI	apnea index (events per hour)
AMI	acute myocardial infarction
AMSTAR 2	A Measurement Tool to Assess Systematic Reviews version 2
APPLES	Apnea Positive Pressure Long-term Efficacy Study
BPAP	bilevel positive airway pressure (device)
BMI	body mass index
CAD	coronary artery disease
CeVD	cerebrovascular disease
CHF	congestive heart failure
CI	confidence interval
CMS	Centers for Medicare & Medicaid Services
CPAP	continuous positive airway pressure (device)
CQ	Contextual Question
CV	cardiovascular
CVD	cardiovascular disease
DAG	directed acyclic graph
DM	diabetes mellitus
EHC	AHRQ's Effective Health Care (Program)
ESS	Epworth Sleepiness Scale
EPAP	(nasal) expiratory positive airway pressure (device)
EPC	Evidence-based Practice Center
FDA	Food and Drug Administration
FOSQ	Functional Outcomes of Sleep Questionnaire
HADS	Hospital Anxiety and Depression Scale
HR	hazard ratio
HTN	hypertension
ICD	International Statistical Classification of Diseases and Related Health Problems
ICSD	International Classification of Sleep Disorders
IQR	interquartile range
ITT	intention-to-treat
IIEF-15	International Index of Erectile Function
KQ	Key Question
MAD	mandibular advancement device
MCID	minimal clinically important difference
MI	myocardial infarction
MMSE	Mini-Mental State Examination
MOSAIC	Multicentre Obstructive Sleep Apnoea Interventional Cardiovascular (trial)
NDE	natural direct effect
NIE	natural indirect effect
NRCS	nonrandomized comparative study

ODI	oxygen desaturation index (events per hour)
ONSLEEP	Okinawa Nakamura Sleep (registry)
OR	odds ratio
OSA	obstructive sleep apnea
PREDICT	trial with undefined acronym
PRISMA	Preferred Items for Reporting in Systematic Reviews and Meta-Analyses
QoL	quality of life
RCT	randomized controlled trial
RDI	respiratory disturbance index (events per hour)
REM	rapid eye movement
RERA	respiratory effort related arousals (events per hour)
RICCADSA	Randomized Intervention with Continuous Positive Airway Pressure in CAD and OSA (trial)
RoB	risk of bias
ROBINS-I	Risk Of Bias In Non-randomised Studies - of Interventions (tool)
SAQLI	Sleep Apnea Quality of Life Index
SAVE	Sleep Apnea cardioVascular Endpoints (trial)
SCL-90-R	Symptom Checklist-90-Revised
SDQ	Sleep Disorders Questionnaire
SDS	Self-rating Depression Scale
SE	standard error
SF-36/12	Short Form 36/12 health survey
SHHS	Sleep Heart Health Study
SMD	standardized mean difference
SNORE-25	Symptoms of Nocturnal Obstruction and Related Events
SoE	strength of evidence
SR	systematic review
SRDR	Systematic Review Data Repository
TIA	transient ischemic attack
TMT-B	Trail Making Test B